Math. Systems Theory 20, 53–82 (1987)

Mathematical
Systems Theory

Differential Geometry of a Parametric Family of Invertible Linear Systems—Riemannian Metric, Dual Affine Connections, and Divergence

Shun-ichi Amari

Department of Mathematical Engineering and Instrumentation Physics,
University of Tokyo, Tokyo, 113 Japan

Abstract. A parametric model of systems is regarded as a geometric manifold imbedded in the enveloping manifold consisting of all the linear systems. The present paper aims at establishing a new geometrical method and framework for analyzing properties of manifolds of systems. A Riemannian metric and a pair of dual affine connections are introduced to a system manifold. They are called α-connections. The duality of connections is a new concept in differential geometry. The manifold of all the linear systems is α-flat so that it admits natural and invariant α-divergence measures. Such geometric structures are useful for treating the problems of approximation, identification, and stochastic realization of systems. By using the α-divergences, we solve the problem of approximating a given system by one included in a model. For a sequence of α-flat nesting models such as AR models and MA models, it is shown that the approximation errors are decomposed additively corresponding to each dimension of the model.

1. Introduction

A parametric family of systems is widely used in systems theory to specify systems and to study their characteristics. Such a family usually forms a finite-dimensional manifold imbedded in the set of all the linear systems. It is not only interesting in its own right but also useful for investigating geometric properties of such a parametric family. For example, consider the problem of approximating a given linear system by one belonging to a family. If the problem is to approximate a point in a Euclidean space by one belonging to its smooth submanifold, the orthogonal projection gives the best solution. Here, the Euclidean distance measure, together with the notion of orthogonality and straightness, plays an important role. It is therefore interesting and useful to study whether or not there exist in a manifold of systems a natural and invariant distance or divergence measure, a notion of orthogonality, and a notion of straightness or curvature. Moreover, provided such geometric structures exist, it is interesting to know if they are useful for the approximation problem, identification problem, robust problem, and many other problems in systems theory.

The purpose of the present paper is firstly to study invariant differential geometric structures inherent in a manifold of systems. We show that a Riemannian metric given by the Fisher information, $\pm\alpha$-affine connections which are dually coupled to each other, and α-divergence measures are introduced naturally in such a manifold. We then consider the problems of approximation, identification, and stochastic realization of systems in this framework. The approximation problem is explicitly solved by using the notion of α-projection. Some parametric families are α-flat, e.g., an AR model is 1-flat and an MA model is -1-flat, and the orthogonal decomposition theorem of approximation errors holds in such a family of nesting α-flat models. These results show the usefulness of the present geometric framework. The notion of dual affine connections or $\pm\alpha$-connections is a new geometrical concept introduced in a theory of statistical inference [1], [4], [27], and has been proved to be useful for solving statistical problems [13], [24].

It is only recently that systems theorists have had interests in the geometric properties of a family of linear systems, as Brockett [8] pointed out in his interesting paper. See also Kalman [21] and Hazewinkel [17].

Algebraic geometry and differential geometry have been used to study properties of manifolds of systems by Brockett [8], Brockett and Krishnaprasad [9], Segal [29], Byrnes [11], Hazewinkel [18], Hermann and Martin [19], Tannenbaum [31], Hannan [15], Hannan and Deistler [16], and others. Recently, a Morse-theoretic study was given by Delchamps [13]. All of these studies are mainly concerned with the global topological or qualitative properties of a manifold of systems. However, local quantitative properties such as metric and curvature are no less important, because they are directly connected with the problem of identification, approximation, and realization of systems as well as their robustness. The present paper aims at giving a differential-geometric framework for constructing such a theory.

Since the present paper aims at proposing a new mathematical framework for system theory, our discussions are limited to a very simple case of scalar-input scalar-output stable and invertible (i.e., of minimal phase) systems. However, it is not difficult to apply our method to more general systems. It should again be noted that the new geometric structure was first introduced in a manifold of statistical models [12], [1], [4], [27], and has already been proved to play an essential role in the theory of statistical inference [4], [5].

2. Manifold L of Linear Systems and α-models

Let us consider stationary and stable discrete-time linear systems with a scalar input and a scalar output. We denote by

$$H(z)=\sum_{t=0}^{\infty} h_t z^{-t}$$

the transfer function of a system, where z is the time shift operator and $\mathbf{h}=(h_0, h_1, h_2, \ldots)$ is the impulse response. The present paper treats only minimal phase or invertible rational systems, although our method is applicable to a more general case. Therefore, it is assumed that $H(z)^{-1}$ is also a transfer function of a stable system. In other words, we assume that both $H(z)$ and $H(z)^{-1}$ are analytic in the domain $|z|\geq 1$ of a complex variable z. This implies that $H(z)$ is an outer function in the inner-outer factorization of functions in the Hardy class (see Duren [14]).

When a unit white Gaussian noise $\{\varepsilon_t\}$, $t=0, 1, 2, \ldots$, is applied to a system, its outputs $\{x_t\}$,

$$x_t=\sum_{i=0}^{\infty} h_i \varepsilon_{t-i},$$

form a regular stationary zero-mean Gaussian process. Its spectral density $S(\omega)$ is given by the square of the gain of the system as

$$S(\omega)=|H(e^{i\omega})|^2, \tag{2.1}$$

which satisfies the condition

$$\int_{-\pi}^{\pi} \log S(\omega)\, d\omega > -\infty. \tag{2.2}$$

Conversely, for a given spectral density $S(\omega)$ satisfying this condition, there exists a unique system which satisfies (2.1). In particular, we have interests in systems whose spectral densities are continuous and $0<S(\omega)<\infty$.

Remark. Let us consider two systems which have the same amplitude function but differ only in the phase factor (or the inner function). Then, they produce stochastic processes of the same spectral density $S(\omega)$. It should be noted that a zero-mean Gaussian process is determined only by their second moments or the spectral density. Since any systems having the same amplitude function produce the same stochastic process for a white Gaussian input, we treat here only minimal phase systems, which are in one-to-one correspondence with spectral densities. However, if we use a non-Gaussian white noise $\{\varepsilon_i\}$ as an input stochastic process, the stochastic properties of the output process $\{x_t\}$ are responsible not only for the amplitude function but also for the phase factor of the system. Hence, we can analyze the geometric properties of a manifold of general systems by the same method as proposed in the present paper. In this general case, we need to use not only the second-order moments $S(\omega)$ but also higher-order moments of $\{x_t\}$.

Let L be the Banach space consisting of those systems whose spectral densities are continuous with $0<S(\omega)<\infty$, where the norm of a point $S(\omega)$ is given by

$$\|S(\omega)\|=\max_{0<\omega<2\pi} |\log S(\omega)|. \tag{2.3}$$

It includes all the minimal phase invertible rational systems. We next consider an extended manifold L_0,

$$L_0=\left\{S(\omega)\mid \int [\log S(\omega)]^2\,d\omega<\infty\right\},$$

which includes L (but has a different topology). This L_0 is a Hilbert manifold (see Lang [25] and Klingenberg [22]), which implies, roughly speaking, that every point L has a neighborhood homeomorphic to a Hilbert space $\mathscr{H}$. Indeed, we can define a homeomorphism φ: $N_S\to\mathscr{H}$ by

$$\varphi S'=\log(S'/S),\qquad \varphi^{-1}A=S\exp\{A(\omega)\},$$

where $S'\in N_S$, $A(\omega)\in\mathscr{H}$, N_S is a neighborhood of S, and the inner product in $\mathscr{H}$ of A and B is given by

$$\langle A, B\rangle=\int A(\omega)B(\omega)\,d\omega.$$

It is convenient to introduce an infinite-dimensional coordinate system $\mathbf{c}^{(0)}=(c_0^{(0)}, c_1^{(0)}, c_2^{(0)},\dots)$ in L_0 by

$$\log S(\omega)=\sum_{t=0}^{\infty} c_t^{(0)}e_t(\omega), \tag{2.4}$$

$$c_t^{(0)}=(2\pi)^{-1}\int [\log S(\omega)]e_t(\omega)\,d\omega, \tag{2.5}$$

where

$$e_0(\omega)=1,\qquad e_t(\omega)=\sqrt{2}\cos\omega t,\qquad t=1,2,\dots$$

We call it the 0-coordinate system.

There exist more popular parameters specifying an $S(\omega)\in L$. For example, the Fourier coefficients of a spectral density

$$c_t=(2\pi)^{-1}\int S(\omega)e_t(\omega)\,d\omega,$$

$$S(\omega)=\sum c_t e_t(\omega)$$

are the autocovariances of the related time series $\{x_t\}$,

$$c_t=E[x_s x_{s+t}],\qquad t=0,1,2,\dots,$$

where E denotes the expectation. The set $(c_0, c_1, c_2,\dots)$ can be used to specify an $S(\omega)$. Another one is the Fourier coefficients of the inverse of a spectral density,

$$\tilde{c}_t=(2\pi)^{-1}\int [1/S(\omega)]e_t(\omega)\,d\omega,$$

$$S(\omega)=\left[\sum \tilde{c}_t e_t(\omega)\right]^{-1}.$$

The sequence $\tilde{\mathbf{c}}=(\tilde{c}_0, \tilde{c}_1,\dots)$ is known as the inverse autocovariances [6]. These two parameter sets have nice properties dually coupled to each other, as will be shown later. In order to study properties related to such parameters, we introduce a nesting family of α-models, where α is a scalar parameter, by generalizing AR models, MA models, etc. To this end, we define a parametric model of systems. We often treat a model of linear systems or associated time series whose members are specified by a finite number of parameters. Let $\mathbf{v}=(v_0, v_1, \ldots, v_p)$ be a $(p+1)$-dimensional real vector parameter, and let M_p be a set of linear systems in L whose spectral densities are smoothly specified by $\mathbf{v}$ as $S(\omega, \mathbf{v})$. Such an $M_p=\{S(\omega, \mathbf{v})\}$ is called a model, and, in general, forms a $(p+1)$-dimensional submanifold in the system manifold L_0, where $\mathbf{v}$ defines a coordinate system of M_p. The set of all the invertible linear systems in L whose McMillan degree is n is an example of a model, and its dimension number is $2n+1$. It is also called an ARMA model.

An MA model $(\mathrm{MA})_p$ of degree p consists of systems whose transfer functions are written as

$$H(z)=\sum_{t=0}^{p} h_t z^{-t}$$

or, equivalently,

$$S(\omega; \mathbf{h})=\left|\sum_{t=0}^{p} h_t e^{i\omega t}\right|^2,$$

where $\mathbf{h}=(h_0, \ldots, h_p)$ is a coordinate system of $(\mathrm{MA})_p$ and is called the MA parameters. An AR model $(\mathrm{AR})_p$ of degree p consists of systems of the form

$$H(z)=1\Big/\left(\sum_{t=0}^{p} a_t z^{-t}\right)$$

or

$$S(\omega; \mathbf{a})=\left|\sum_{t=0}^{p} a_t e^{i\omega t}\right|^{-2}$$

with a coordinate system $\mathbf{a}=(a_0, a_1, \ldots, a_p)$ called the AR parameters. A Bloomfield exponential model B_P is specified in the spectral form as

$$S(\omega; \mathbf{b})=\exp\left(\sum_{t=0}^{p} b_t e_t(\omega)\right),$$

where $\mathbf{b}=(b_0, \ldots, b_p)$ is a coordinate system [7].

By generalizing the above three models, we define an α-model $M_p^{(\alpha)}$. To this end, we define the α-representation $R^{(\alpha)}(\omega)$ of a spectral density $S(\omega)$ by

$$R^{(\alpha)}(\omega)=\begin{cases}-(1/\alpha)S(\omega)^{-\alpha}, & \alpha \neq 0,\\ \log S(\omega), & \alpha=0.\end{cases} \tag{2.6}$$

Obviously the -1-representation is $S(\omega)$ itself and the 1-representation is $-[1/S(\omega)]$. It should be remarked that the limit $\alpha \to 0$ of $-(1/\alpha)[S(\omega)^{-\alpha}-1]$ is equal to $\log S(\omega)$, so that the α-representation is continuous in α, if we add a constant. An α-model $M_p^{(\alpha)}$ of degree p consists of those systems whose α-representations of the spectral densities are given by the following finite sums,

$$R^{(\alpha)}(\omega)=\sum_{t=0}^{p} c_t^{(\alpha)} e_t(\omega), \tag{2.7}$$

so that $\mathbf{c}^{(\alpha)}=(c_0^{(\alpha)}, c_1^{(\alpha)}, \ldots, c_p^{(\alpha)})$ is a coordinate system of $M_p^{(\alpha)}$. An AR model is a 1-model ($\alpha=1$), an MA model is a -1-model ($\alpha=-1$), and a Bloomfield exponential model is a 0-model.

A sequence of models $M_0, M_1, M_2, \ldots$ is said to be nesting, when they satisfy the inclusion relation

$$M_0 \subset M_1 \subset M_2 \subset \cdots.$$

A sequence of α-models is nesting. For any finite p, $M_p^{(\alpha)}$ is a submanifold of L_0 and is included in L. However, $L_\alpha = M_\infty^{(\alpha)}$ which is composed of infinite sums in (2.7), is also included in L, but is not necessarily a submanifold of L_0, because the natural topology of L_α ($\alpha \neq 0$) is not equal to that of L_0. A natural coordinate system of L_α, which is called the α-coordinate system, is $\mathbf{c}^{(\alpha)}=(c_0^{(\alpha)}, c_1^{(\alpha)}, \ldots)$,

$$c_t^{(\alpha)}=(2\pi)^{-1}\int R^{(\alpha)}(\omega)e_t(\omega)\,d\omega,$$

and

$$R^{(\alpha)}(\omega)=\sum c_t^{(\alpha)} e_t(\omega) \tag{2.8}$$

holds. Since any $S(\omega)\in L$ has α-coordinates $\mathbf{c}^{(\alpha)}$, we treat $\mathbf{c}(\alpha)$ as if it is a coordinate system of L. However, it should be noted that $\mathbf{c}^{(\alpha)}$ is valid in $M_p^{(\alpha)}$ as a submanifold of L_0 or in L_α having a different topology. The -1-coordinates $\mathbf{c}^{(-1)}$ give the autocovariances, and the 1-coordinates $\mathbf{c}^{(1)}$ give the negative of the inverse autocovariances. The 0-coordinates $\mathbf{c}^{(0)}$ are the same as those previously defined. The coordinate transformations are given by

$$c_t^{(\alpha)}=-(2\pi)^{-1}\int \exp[-\sum c_s^{(0)} e_s(\omega)]e_t(\omega)]e_t(\omega)\,d\omega,$$

$$c_t^{(0)}=-(2\pi)^{-1}\int \log[-\sum c_s^{(\alpha)} e_s(\omega)]e_t(\omega)\,d\omega.$$

3. Riemannian Metric in L

We introduce a Riemannian metric in L or more precisely in L_0 or its subspace $M_p^{(\alpha)}$, which defines a distance between two adjacent systems. It is invariant under concatenation with a known system, as is shown in Section 8.1. Let $\mathbf{u}=(u^i)$, $i=0, 1, 2, \ldots$, be any coordinate system of L so that any spectral density $S(\omega)\in L$ is uniquely and smoothly specified by $\mathbf{u}$ as $S(\omega; \mathbf{u})$. When we fix a system $S(\omega; \mathbf{u}_0)$ and consider those systems $S(\omega; \mathbf{u})$ which are very close to $\mathbf{u}_0$, i.e., $|\mathbf{u}-\mathbf{u}_0|$ is very small, we may use a "linear approximation" of L at $\mathbf{u}_0$. Mathematically, this is

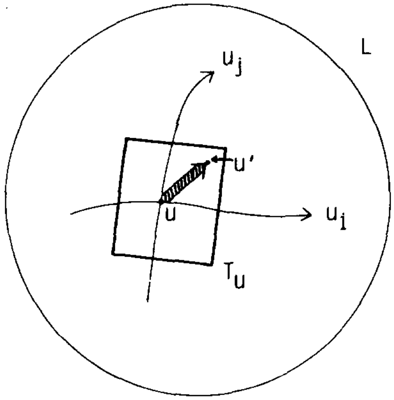

Fig. 1. Tangent space T_u.

to consider the tangent space $T\mathbf{u}_0$ of L at $\mathbf{u}_0$. Let $\mathbf{E}_i$ $(i=0,1,2,\ldots)$ be a vector which is tangent to the ith coordinate axis u^i. Then, the tangent space T_u of L at any $\mathbf{u}$ is a vector space spanned by the vectors $\mathbf{E}_i$ $(i=0,1,2,\ldots)$ (see Fig. 1). The basis $\mathbf{E}_i$ is called the natural basis associated with the coordinate system $\mathbf{u}$. Mathematicians use the symbol ∂_i or $\partial/\partial u^i$ instead of $\mathbf{E}_i$ for the natural basis in order to emphasize its abstract character as a differential operator. Anyway, $\mathbf{E}_i$ or ∂_i represents the direction in which the value of the ith coordinate increases while the other coordinates are fixed. Any vectors $\mathbf{A}\in T_u$ are written as a linear combination of the basis vectors,

$$\mathbf{A}=A^i\mathbf{E}_i,$$

where A^i are the components of $\mathbf{A}$. Here, and throughout the paper, the Einstein summation convention is assumed: the summation is taken for those indices which are repeated twice in a term, once as a superscript and once as a subscript (such as i in the above), so that $A^i\mathbf{E}_i$ automatically implies $\sum A^i\mathbf{E}_i$.

Let us consider two adjacent systems whose spectral densities are $S(\omega,\mathbf{u})$ and $S'=S(\omega,\mathbf{u}+d\mathbf{u})$, where (du^i) are infinitesimally small. Then, the difference between two points S and S' in L is identified with an infinitesimal tangent vector

$$d\mathbf{u}=du^i\mathbf{E}_i, \tag{3.1}$$

which belongs to T_u. If we use the 0-representation or the logarithmic scale $\log S(\omega,\mathbf{u})$ to represent the spectral density, we may write

$$d\mathbf{u}=\overrightarrow{SS'}=\log S(\omega,\mathbf{u}+d\mathbf{u})-\log S(\omega,\mathbf{u})=du^i\partial_i\log S(\omega,\mathbf{u}),$$

where ∂_i denotes the partial derivative $\partial/\partial u^i$. Comparing this with (3.1), we may represent the vector $\mathbf{E}_i\in T_u$ by the following function in ω,

$$\mathbf{E}_i=E_i(\omega,\mathbf{u})=\partial_i\log S(\omega,\mathbf{u}). \tag{3.2}$$

When we use the 0-coordinate system $\mathbf{u}=\mathbf{c}^{(0)}$, then $E_i(\omega,\mathbf{u})=e_i(\omega)$.

A vector $\mathbf{A}=A^i\mathbf{E}_i\in T_u$ is then represented by a function $A(\omega)=A^i\partial_i\log S(\omega,u)$ in $\mathscr{H}$. Let us introduce an inner product $\langle\mathbf{A},\mathbf{B}\rangle$ of two vectors $\mathbf{A}$

and $\mathbf{B}$ in the linear space T_u by using their representations $A(\omega)$, $B(\omega)$ as

$$\langle \mathbf{A}, \mathbf{B} \rangle = (2\pi)^{-1} \int A(\omega) B(\omega)\, d\omega. \tag{3.3}$$

Then, the inner product of two basis vectors $\mathbf{E}_i$ and $\mathbf{E}_j$ is given by

$$g_{ij}(\mathbf{u}) = \langle \mathbf{E}_i, \mathbf{E}_j \rangle = (2\pi)^{-1} \int E_i(\omega, \mathbf{u}) E_j(\omega, \mathbf{u})\, d\omega. \tag{3.4}$$

The square of the distance ds between two neighboring systems $S(\omega, \mathbf{u})$ and $S(\omega, \mathbf{u} + d\mathbf{u})$ is given by the quadratic form

$$ds^2 = \langle d\mathbf{u}, d\mathbf{u} \rangle = g_{ij}(\mathbf{u})\, du^i\, du^j. \tag{3.5}$$

It is clear from the definition that the above distance does not depend on the coordinate system $\mathbf{u}$ which is adopted temporarily. The same value ds^2 is assigned when another coordinate system $\mathbf{u}'$ is adopted. In fact, the square of the Riemannian distance between $S(\omega)$ and $S(\omega) + \delta S(\omega)$ in L is written as

$$ds^2 = (2\pi)^{-1} \int \{S(\omega)\}^{-2} \{\delta S(\omega)\}^2\, d\omega,$$

which is free of any coordinate systems. When the α-coordinate system $\mathbf{c}^{(\alpha)}$ is used, we have

$$g_{ij} = (2\pi)^{-1} \int \{S(\omega, \mathbf{c})\}^{2\alpha} e_i(\omega) e_j(\omega)\, d\omega, \tag{3.6}$$

in this coordinate system, because of

$$e_i(\omega) = \partial_i^{(\alpha)} R^{(\alpha)} = S^{-\alpha} \partial_i \log S, \tag{3.7}$$

where $\partial_i^{(\alpha)} = \partial / \partial c_i^{(\alpha)}$.

A manifold is said to be Riemannian when an inner product is defined in its tangent space T_u at every $\mathbf{u}$. Since each T_u is a Euclidean space, a Riemannian manifold can be approximated locally by a tangent Euclidean space.

Let M_p be a model in L, $M_p = \{S(\omega, \mathbf{v})\}$ where $\mathbf{v} = (v^a)$ $(a = 0, 1, \ldots, p)$ is a coordinate system of M_p and the suffixes a, b, c, etc., run from 0 to p. The tangent space $T_v(M_p)$ of M_p at $\mathbf{v}$ is spanned by $(p+1)$ vectors $\mathbf{E}_a (a = 0, 1, \ldots, p)$, where $\mathbf{E}_a$ is the natural basis associated with the coordinate system $\mathbf{v}$, i.e., the tangent vector of the ath coordinate axis v^a. It is a linear subspace of the tangent space T_u of L at $\mathbf{u}$, where

$$\mathbf{u} = \mathbf{u}(\mathbf{v}) \tag{3.8}$$

are the coordinates in L of the system $S(\omega, \mathbf{v})$. Equation (3.8) is the parametric representation of M_p in L. Since $\mathbf{E}_a$ is also a vector in T_u, it can be represented by a linear combination of the basis vectors $\mathbf{E}_i$ in T_u as

$$\mathbf{E}_a = B_a^i \mathbf{E}_i,$$

where

$$B_a^i(\mathbf{v}) = \partial u^i(\mathbf{v}) / \partial v^a \tag{3.9}$$

are the components of $\mathbf{E}_a$ with respect to the basis $\mathbf{E}_i$. This can easily be obtained from the differentials of (3.8),

$$du^i = (\partial u^i/\partial v^a)\, dv^a,$$

and the representation of the vector $\overrightarrow{SS'}$,

$$\overrightarrow{SS'} = du^i \mathbf{E}_i = dv^a \mathbf{E}_a,$$

where $S = S(\omega; \mathbf{v})$ and $S' = S(\omega; \mathbf{v} + d\mathbf{v})$. An inner product is defined in $T_v(M_p)$ by restricting vectors in $T_v(M_p)$. For example, the inner product of two basis vectors $\mathbf{E}_a$ and $\mathbf{E}_b$ is given by

$$g_{ab}(\mathbf{v}) = \langle \mathbf{E}_a, \mathbf{E}_b \rangle = (2\pi)^{-1} \int E_a(\omega, \mathbf{v}) E_b(\omega, \mathbf{v})\, d\omega, \tag{3.10}$$

where

$$E_a(\omega, \mathbf{v}) = (\partial/\partial v^a) \log S(\omega, \mathbf{v}), \tag{3.11}$$

or directly in the component form by

$$g_{ab}(\mathbf{v}) = B_a^i(\mathbf{v}) B_b^j(\mathbf{v}) g_{ij}(\mathbf{u}), \tag{3.12}$$

where $\mathbf{u} = \mathbf{u}(\mathbf{v})$. The $(p+1) \times (p+1)$ matrix $g = (g_{ab})$ is called the Riemannian metric tensor of a model M_p.

The metric tensor $g_{ij}(\mathbf{u})$ of L or the metric tensor $g_{ab}(\mathbf{v})$ in M_p is twice the Fisher information matrix which plays a fundamental role in statistics [28], [4]. The Riemannian distance ds between two adjacent systems $S(\omega; \mathbf{v})$ and $S(\omega; \mathbf{v} + d\mathbf{v})$ is given by

$$ds^2 = \langle d\mathbf{v}, d\mathbf{v} \rangle = g_{ab}\, dv^a\, dv^b,$$

and it indeed represents how the two systems are different in their behaviors. This can be understood from the following Cramér-Rao theorem, which is related to the system identification problem.

Theorem 1. *Let $\hat{\mathbf{v}}_T$ be an unbiased estimator of the system parameter $\mathbf{v}$ in a model $M_p = \{S(\omega, \mathbf{v})\}$ based on the observation of T outputs*

$$\mathbf{x}_T = (x_1, x_2, \ldots, x_T),$$

where the input of the system is a white Gaussian noise. Then the covariance matrix of the estimation error $\hat{\mathbf{v}}_T - \mathbf{v}$ is bounded from below by

$$E[(\hat{v}_T^a - v^a)(\hat{v}_T^b - v^b)] \geq 2T^{-1} g^{ab}, \tag{3.13}$$

where $(2g^{ab})$ is the inverse of the Fisher information matrix $g_{ab}/2$ and the matrix inequality $g \geq h$ implies that $g - h$ is a positive semidefinite matrix. Moreover, when T is large, there exists an estimator $\hat{\mathbf{v}}_T$ (for example, the maximum likelihood estimator) such that the equality holds asymptotically in (3.13).

This shows that (g^{ab}) represents the bound of estimation error and (g_{ab}) represents the amount of information utilizable in estimating the parameter $\mathbf{v}$.

The distance introduced in the system manifold L or in a model M_p is based on the Fisher information matrix so that ds between two systems S and S' represents how well S can be distinguished from S' based on their outputs.

4. α-connections of the System Manifold

At each point $\mathbf{u}$ of the system manifold L, the tangent space T_u is regarded as local linearization of L at around $\mathbf{u}$. The collection of these T_u ($\mathbf{u} \in L$) forms a fiber bundle called the tangent bundle. The two tangent spaces T_u and $T_{u'}$ approximate L in small neighborhoods of $\mathbf{u}$ and $\mathbf{u}'$, respectively. In order to study more global properties of L where local linear approximation is insufficient, we need to connect various T_u's such that they together cover a larger domain of L. However, since T_u and $T_{u'}$ are two different vector spaces, they cannot be directly connected without any criterion for comparing them. In order to connect two T_u and $T_{u'}$ at two adjacent $\mathbf{u}$ and $\mathbf{u}' = \mathbf{u} + d\mathbf{u}$, mathematicians define a linear correspondence between them, which reduces to the identity map as $\mathbf{u}'$ approaches $\mathbf{u}$. Such a correspondence is called an affine connection.

Let $\mathbf{E}_i(\mathbf{u})$ and $\mathbf{E}_i(\mathbf{u} + d\mathbf{u})$ be two basis vectors of T_u and $T_{u'}$, where $\mathbf{u}$ and $\mathbf{u}' = \mathbf{u} + d\mathbf{u}$ are infinitesimally close. The vector $\mathbf{E}_i(\mathbf{u} + d\mathbf{u}) \in T_{u'}$ should be mapped to a vector in T_u which is close to $\mathbf{E}_i(\mathbf{u})$ by the correspondence between T_u and $T_{u'}$. Hence, we may write that $\mathbf{E}_i(\mathbf{u} + d\mathbf{u})$ in $T_{u'}$ corresponds to a vector

$$\tilde{\mathbf{E}}_i = \mathbf{E}_i(\mathbf{u}) + \mathbf{A}_{ji}\, du^j$$

in T_u which reduces to $\mathbf{E}_i$ as $d\mathbf{u}$ tends to 0 (Fig. 2). This correspondence is determined by $(p+1)^2$ vectors $\mathbf{A}_{ji}$, since the correspondence of the basis vectors

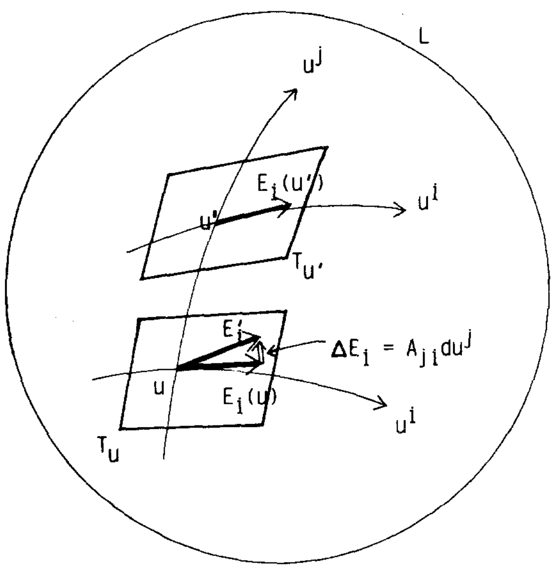

Fig. 2. Affine connection.

uniquely determine an affine correspondence between T_u and $T_{u'}$. The vectors $\mathbf{A}_{ji}$ represent how the corresponding $\tilde{\mathbf{E}}_i$ in T_u of the basis vector $\mathbf{E}_i(\mathbf{u}')$ changes as the point $\mathbf{u}'$ moves from $\mathbf{u}$ in the direction of the jth coordinate axis u^j. We may denote it by

$$\nabla_j \mathbf{E}_i = \lim_{du^j \to 0} (\tilde{\mathbf{E}}_i - \mathbf{E}_i)/du^j = \mathbf{A}_{ji}$$

and call it the covariant derivative of the basis vector field $\mathbf{E}_i(\mathbf{u})$ in the direction of $\mathbf{E}_j$. The vectors $\mathbf{A}_{ji}$ are given by their components which are obtained from the inner products

$$\Gamma_{jik}(\mathbf{u}) = \langle \nabla_j \mathbf{E}_i, \mathbf{E}_k \rangle. \tag{4.1}$$

This is a quantity having three indices and is called the components of the affine connection. Hence, an affine connection is introduced in L by defining $\mathbf{A}_{ji}$, or equivalently by the covariant derivative ∇, or the quantity $\Gamma_{jik}(\mathbf{u})$.

Let $\mathbf{C} = \mathbf{C}(\mathbf{u})$ be a vector field defined over L, i.e., a vector $\mathbf{C}(\mathbf{u})$ is specified at each T_u. Once an affine connection is introduced, we can calculate its covariant derivative $\nabla_{\mathbf{B}} \mathbf{C}$ in the direction of another vector field $\mathbf{B} = \mathbf{B}(\mathbf{u})$, which designates how $\mathbf{C}(\mathbf{u})$ changes as the point $\mathbf{u}$ moves in the direction of $\mathbf{B}$. In fact, by the use of linearity and the Leibnitz law for the differentiation operator, we define

$$\begin{aligned} \nabla_B \mathbf{C} &= B^j \nabla_j (C^i \mathbf{E}_i) \\ &= B^j (\partial_j C^i) \mathbf{E}_i + B^j C^i \nabla_j \mathbf{E}_i. \end{aligned} \tag{4.2}$$

Here, not only the change in the components $C^i(\mathbf{u})$ but also the change in the basis vectors $E_i(\mathbf{u})$ is taken into account, where $\mathbf{B} = B^j \mathbf{E}_j$, $\mathbf{C} = C^i \mathbf{E}_i$.

What connection is naturally and invariantly introduced in L? It has been shown that a one-parameter family of connections, called the α-connections, is naturally and uniquely introduced in statistical manifolds [12], [1], [4]. By using this fact, we also introduce α-connections in our system manifold L, where α is a scalar parameter. We define the α-covariant derivative $\nabla^{(\alpha)}$ of $\mathbf{E}_i$ in the direction of $\mathbf{E}_j$ by

$$\nabla_j^{(\alpha)} \mathbf{E}_i = \partial_j E_i(\omega, \mathbf{u}) - \alpha E_i(\omega, \mathbf{u}) E_j(\omega, \mathbf{u}), \tag{4.3}$$

where we use the function representation (3.2) of vectors.

The components of the α-connection are then given by

$$\Gamma_{ijk}^{(\alpha)}(\mathbf{u}) = (2\pi)^{-1} \int \{\partial_i E_j(\omega, \mathbf{u}) - \alpha E_i E_j\} E_k(\omega, \mathbf{u})\, d\omega. \tag{4.4}$$

By using the α-representation, they are rewritten as

$$\Gamma_{ijk}^{(\alpha)}(\mathbf{u}) = (2\pi)^{-1} \int \{S(\omega, \mathbf{u})\}^{2\alpha} \partial_i \partial_j R^{(\alpha)} \partial_k R^{(\alpha)}\, d\omega. \tag{4.5}$$

We can prove that the α-connections introduced here are the same as those introduced in statistical manifolds (see Amari [2], [3]), if we regard the outputs $\{x_t\}$ as a random process specified by $\mathbf{u}$. The α-connections have become indispensable tools in the asymptotic theory of statistical inference [1], [4], [5]. They also seem to play a fundamental role in the theory of systems.

The α-connections can also be introduced to any finite-dimensional model M_p with a coordinate system $\mathbf{v}=(v^a)$. The tangent space $T_v(M_p)$ at $\mathbf{v}$ is spanned by $p+1$ basis bectors

$$\mathbf{E}_a = B_a^i \mathbf{E}_i. \tag{4.6}$$

By the use of the α-connection, we can define how curved a submanifold M_p in L is by calculating the changes in the tangent directions $T_v(M_p)$ of M_p as the point $\mathbf{v}$ moves in M_p. The changes are given by the α-covariant derivatives $\nabla_a^{(\alpha)}\mathbf{E}_b$ of the basis vectors $\mathbf{E}_b$ in the tangent directions $\mathbf{E}_a$,

$$\begin{aligned}\mathbf{K}_{ab}^{(\alpha)} &= \nabla_a^{(\alpha)}\mathbf{E}_b \\ &= (\partial_a B_b^i)\mathbf{E}_i + B_b^i B_a^j \nabla_j^{(\alpha)}\mathbf{E}_i.\end{aligned}$$

The quantity $\mathbf{K}_{ab}^{(\alpha)}$ represents how the basis vectors of $T_v(M_p)$ change in the sense of the α-connection as the point $\mathbf{v}$ moves in M_p. The vector $\mathbf{K}_{ab}^{(\alpha)}$ belongs to the entire tangent space T_u of L at $\mathbf{u}=\mathbf{u}(\mathbf{v})$. It can be decomposed into two components. One is the tangential component, i.e., the component belonging to $T_v(M_p)$, and is given from

$$\langle \mathbf{K}_{ab}^{(\alpha)}, \mathbf{E}_c\rangle = (\partial_a B_b^i) B_c^j g_{ij} + B_b^i B_a^j B_c^k \Gamma_{ijk}^{(\alpha)}.$$

By defining

$$\Gamma_{abc}^{(\alpha)} = \langle \mathbf{K}_{ab}^{(\alpha)}, \mathbf{E}_c\rangle, \tag{4.7}$$

the tangential component is written as

$$\tilde{\nabla}_a^{(\alpha)}\mathbf{E}_b = \Gamma_{abc}^{(\alpha)}\mathbf{E}_d g^{cd},$$

where g^{cd} is the inverse of the matrix g_{cd}. This is the projection of $\nabla_a^{(\alpha)}\mathbf{E}_b$ to $T_v(M_p)$. This defines the α-covariant derivative $\tilde{\nabla}_a^{(\alpha)}$ of $\mathbf{E}_b$ in the direction of $\mathbf{E}_a$ in the manifold M_p induced from the enveloping L, because it defines an affine correspondence of $T_v(M_p)$ and $T_{v+dv}(M_p)$. The other component, the orthogonal component (i.e., the component orthogonal to $T_v(M_p)$), of $\mathbf{K}_{ab}^{(\alpha)}$ is given by

$$\mathbf{H}_{ab}^{(\alpha)} = \mathbf{K}_{ab}^{(\alpha)} - \tilde{\nabla}_a^{(\alpha)}\mathbf{E}_b. \tag{4.8}$$

It represents the changes in the tangential directions of the manifold M_p in L as $\mathbf{v}$ moves. We call it the α-curvature of M_p. It is called by various names such as relative curvature, imbedding curvature, Euler-Schouten curvature, etc. When it vanishes identically, the submanifold M_p is said to be completely α-flat, i.e., flat in L in the sense of the α-connection, because the directions of $T_v(M_p)$ remain fixed as $\mathbf{v}$ moves, although $\mathbf{E}_a$ may change within $T_v(M_p)$. If M_p is completely α-flat, $\tilde{\nabla}^{(\alpha)}$ coincides with $\nabla^{(\alpha)}$ when it is operated on vector fields of M_p.

Let

$$C = \{S(\omega, t)\}$$

be a one-parameter family, parametrized by a scalar t, of spectral densities. It forms a curve in L, where t is a scalar parameter or a coordinate specifying points in the curve. When a coordinate system $\mathbf{u}$ is used in L, the parametric form of a curve is written as

$$\mathbf{u} = \mathbf{u}(t).$$

The tangent vector $\mathbf{E}_t$ of the curve is given by

$$\mathbf{E}_t = \dot{u}^i \mathbf{E}_i, \qquad \dot{u}^i = du^i/dt$$

or directly by

$$E_t = (d/dt) \log S(\omega, t) \tag{4.9}$$

in the function representation. A curve is said to be α-geodesic when its tangent direction does not change on the curve in the sense of the α-connection, i.e. when it satisfies the geodesic equation

$$\nabla_t^{(\alpha)} \mathbf{E}_t = 0 \tag{4.10}$$

for an adequate parametrization t. The above geodesic equation is written in the coordinate form as

$$g_{ji} \ddot{u}^j(t) + \Gamma_{jki}^{(\alpha)} \dot{u}^j \dot{u}^k = 0$$

or

$$(d^2/dt^2) \log S(\omega, t) - \alpha\{(d/dt) \log S(\omega, t)\}^2 = 0. \tag{4.11}$$

An α-geodesic is a "straight line" in L in the sense of the α-connection. Since (4.11) is rewritten as

$$d^2/dt^2 R^{(\alpha)}(\omega, t) = 0,$$

an α-geodesic is given in the α-representation by a linear form in t,

$$R^{(\alpha)}(\omega, t) = R_1(\omega) + tR_2(\omega). \tag{4.12}$$

5. Theory of Dual Connections

Before studying the geometric properties of L, we recapitulate the theory of dual connections studied by Nagaoka and Amari [27]; see also Amari [4]. Let us consider a Riemannian manifold M in which two (torsion-free) affine connections Γ and Γ^* or the corresponding covariant derivatives ∇ and ∇^* are defined. The connections defined by ∇ and ∇^* are said to be mutually dual when

$$\mathbf{A}\langle \mathbf{B}, \mathbf{C}\rangle = \langle \nabla_A \mathbf{B}, \mathbf{C}\rangle + \langle \mathbf{B}, \nabla_A^* \mathbf{C}\rangle \tag{5.1}$$

holds for any three vector fields $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, where $\mathbf{A}$ on the left-hand side denotes the directional derivative $A^i \partial_i$ of a scalar function $\langle \mathbf{B}, \mathbf{C}\rangle$. When $\nabla = \nabla^*$, condition (5.1) implies that ∇ is a metric connection. Since ∇ is torsion-free, it is given by the Levi-Civita parallelism from the metric. Hence, a torsion-free metric connection is self-dual in our sense.

A manifold M with an affine connection ∇ is said to be flat when the Riemann-Christoffel curvature tensor and the torsion tensor vanish. When M is flat, there exists a (local) coordinate system $\mathbf{u}$ such that

$$\nabla_i \mathbf{E}_j = 0$$

or

$$\Gamma_{ijk}(\mathbf{u}) = 0,$$

and vice versa. This implies that in a flat M a basis vector $\mathbf{E}_j(\mathbf{u}) \in T_u$ at $\mathbf{u}$ in this coordinate system corresponds to the same basis vector $\mathbf{E}_j(\mathbf{u}') \in T_{u'}$ at $\mathbf{u}'$ for any two points $\mathbf{u}$ and $\mathbf{u}'$. Moreover, the coordinate curves u^i are geodesics. Any geodesic curve $u(t)$ is represented by a linear equation $u^i(t) = ta^i + b^i$ in this coordinate system. Hence, a flat space may be regarded as a linear space with a linear coordinate system $\mathbf{u}$, which is called an affine coordinate system. But this does not mean that the manifold is Euclidean, because the connection is not necessarily metric, i.e., it is not necessarily the Levi-Civita connection, so that the metric tensor $g_{ij}(\mathbf{u})$ depends on $\mathbf{u}$.

Let ∇ and ∇^* be two dual connections of a Riemannian manifold M. It is proved that when M is ∇-flat, i.e., flat with respect to ∇, then it is also flat with respect to ∇^*. Hence, this manifold has two affine coordinate systems θ and η such that $\theta = (\theta^i)$ is ∇-affine and $\eta = (\eta_i)$ is ∇^*-affine. Such a dually flat manifold has been studied in Nagaoka and Amari [27] and recapitulated in Amari [4].

Theorem 2. *There exist two potential functions ψ and φ in a dually flat manifold M such that the metric tensor*

$$g_{ij} = \langle \mathbf{E}_i, \mathbf{E}_j \rangle$$

in a ∇-affine coordinate system θ is given by

$$g_{ij}(\theta) = \partial_i \partial_j \psi(\theta), \qquad \partial_i = \partial/\partial\theta^i, \tag{5.2}$$

and the metric tensor

$$g^{ij} = \langle \mathbf{E}^i, \mathbf{E}^j \rangle, \qquad \mathbf{E}^i = \partial/\partial\eta_i,$$

in a ∇^*-affine coordinate system η is the inverse of (g_{ij}) and is given by

$$g^{ij} = \partial^i \partial^j \varphi(\eta), \qquad \partial^i = \partial/\partial\eta_i. \tag{5.3}$$

The two natural bases $\{\mathbf{E}_i\}$ and $\{\mathbf{E}^j\}$ associated with the coordinate systems θ and η are mutually dual or reciprocal bases of T_u,

$$\langle \mathbf{E}_i, \mathbf{E}^j \rangle = \delta_i^j, \tag{5.4}$$

where δ_i^j is the Kronecker delta. The two coordinate systems are connected by the Legendre transformation

$$\theta^i = \partial^i \varphi(\eta), \qquad \eta_i = \partial_i \psi(\theta) \tag{5.5}$$

and we can choose the potential functions such that

$$\psi(\theta) + \varphi(\eta) - \theta^i \eta_i = 0 \tag{5.6}$$

holds.

By using the potential functions and the affine coordinate systems θ and η, a divergence function $D(P, P')$ between two points P and P' in M is introduced by

$$D(P, P') = \psi(\theta) + \varphi(\eta') - \theta \cdot \eta', \tag{5.7}$$

where θ is the θ-coordinates of P and η' is the η-coordinates of P', and $\theta \cdot \eta'$ is the abbreviation for $\theta^i \eta'_i$. The divergence function satisfies

$$D(P, P') \geq 0,$$

where the equality holds when and only when $P = P'$. When P' is close to P, we have

$$D(P, P') = \tfrac{1}{2} g_{ij}(\theta)(\theta^i - \theta'^i)(\theta^j - \theta'^j) + O(|\theta - \theta'|^3), \tag{5.8}$$

where θ' is the θ-coordinates of P', so that D can be regarded as an extension of the square of the Riemannian distance ds. However, it is not symmetric therefore $D(P, P') = D(P', P)$ does not necessarily hold. On the other hand, the following generalized Pythagorean theorem holds. Let us consider a triangle consisting of three points P, P', and P''. A triangle $\triangle PP'P''$ is said to be a right triangle in M, when the ∇-geodesic connecting P and P' is orthogonal at P' to the ∇^*-geodesic connecting P' and P'' (Fig. 3). (Two intersecting curves are said to be orthogonal when their tangent vectors are orthogonal at the intersection.)

Theorem 3. *For a right triangle $\triangle PP'P''$, the following Pythagorean theorem holds,*

$$D(P, P') + D(P', P'') = D(P, P''). \tag{5.9}$$

Let M' be a submanifold of M. Given a point P in M, we sometimes search for the point $\hat{P}$ in M' which is closest to P in the sense of the divergence, i.e., the point $\hat{P}$ satisfying

$$D(P, \hat{P}) = \min_{P' \in M'} D(P, P').$$

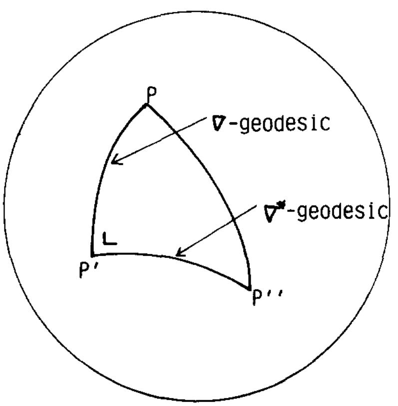

Fig. 3. Pythagorean relation.

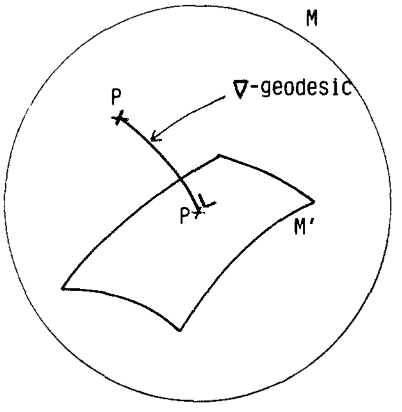

Fig. 4. The ∇-projection.

The point $\hat{P}$ is regarded as the best approximation of P in M'. This problem of approximation is solved by the ∇-projection (Fig. 4). A point $P' \in M'$ is called the ∇-projection of a point $P \in M$ on M', if the ∇-geodesic connecting P and P' is orthogonal to M' at P', i.e., orthogonal to the tangent space $T_{P'}(M')$. The following projection theorem is a direct consequence of the Pythagorean relation.

Theorem 4. *For a point $P \in M$, the point $\hat{P}$ which is closest to P in a submanifold M' is given by the ∇-projection of P on M'. Moreover, ∇-projection is unique when M' is completely ∇^*-flat in M.*

6. Geometric Structures of the System Manifold L

We now study geometric properties of the system manifold L and the nesting families of α-models along the lines in Section 5. We first prove that the α-connections introduced in L have a dual structure.

Theorem 5. *The α- and $-\alpha$-connections are mutually dual. Especially, the 0-connection is self-dual so that it is a Riemannian or Levi-Civita connection.*

Proof. It suffices to prove relation (5.1) for three basis vector fields, $\mathbf{A} = \mathbf{E}_i$, $\mathbf{B} = \mathbf{E}_j$, and $\mathbf{C} = \mathbf{E}_k$. This is proved as follows:

$$\begin{aligned}
\mathbf{E}_i\langle \mathbf{E}_j, \mathbf{E}_k\rangle &= \partial_i \langle \mathbf{E}_j, \mathbf{E}_k\rangle \\
&= (2\pi)^{-1}\partial_i \int E_j(\omega, \mathbf{u}) E_k(\omega, \mathbf{u})\, d\omega \\
&= (2\pi)^{-1} \int \{(\partial_i E_j) E_k + E_j(\partial_i E_k)\}\, d\omega \\
&= (2\pi)^{-1} \int \{(\partial_i E_j - \alpha E_i E_j) E_k + (\partial_i E_k + \alpha E_i E_k) E_j\}\, d\omega \\
&= \langle \nabla_i^{(\alpha)} \mathbf{E}_j, \mathbf{E}_k\rangle + \langle \mathbf{E}_j, \nabla_i^{(-\alpha)} \mathbf{E}_k\rangle.
\end{aligned}$$

Since the α-connections are torsion free, the 0-connection is self-dual so that it is the Levi-Civita connection derived from the metric g_{ij}. □

We next prove that the system manifold L is α-flat for all α. This is a fundamental characteristic of L.

Theorem 6. *The system manifold L is α-flat for all α, and the α-coordinate system $\mathbf{c}^{(\alpha)}$ is its affine coordinate system.*

Proof. Let $\mathbf{u}=\mathbf{c}^{(\alpha)}$ be the α-coordinate system. Then the α-representation $R^{(\alpha)}(\omega, \mathbf{u})$ of the spectral density satisfies

$$\partial_i \partial_j R^{(\alpha)} = 0,$$

because of (2.8). Hence, from (4.5), we have

$$\Gamma^{(\alpha)}_{ijk}(\mathbf{u}) = 0,$$

proving the theorem. □

Since the system manifold L is α-flat, i.e., flat with respect to the dual connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$, there exist an α-potential function ψ_α and its dual function φ_α for any α. The α-divergence $D_\alpha(S, S')$ between two systems $S(\omega)$ and $S'(\omega)$ is calculated by using the potential functions. In order to obtain an explicit form of ψ_α, let us define a function

$$H(\mathbf{u}) = (4\pi)^{-1} \int \log S(\omega, \mathbf{u})\, d\omega + \tfrac{1}{2}\log(2\pi e). \tag{6.1}$$

This is the entropy of the associated stochastic process $\{x_t\}$ whose spectral density is $S(\omega, \mathbf{u})$. The α-potential functions ψ_α and φ_α are given in terms of the entropy function.

Theorem 7. *The α-potential function is given by*

$$\psi_\alpha = \begin{cases} (2/\alpha)H + (2\alpha^2)^{-1}, & \alpha \neq 0, \\ (4\pi)^{-1} \displaystyle\int (\log S)^2\, d\omega, & \alpha = 0, \end{cases} \tag{6.2}$$

$$\varphi_\alpha = \psi_{-\alpha}. \tag{6.3}$$

Proof. Let $\theta = \mathbf{c}^{(\alpha)}$ be the α-coordinate system, and let $\partial_i = \partial/\partial\theta^i$ be the partial derivative with respect to the ith component of θ. We then have, for $\alpha \neq 0$,

$$e_i(\omega) = \partial_i R^{(\alpha)}(\omega, \theta) = S^{-1-\alpha}\partial_i S(\omega, \theta),$$

from which follows

$$\partial_i \log S = S^{\alpha} e_i(\omega),$$

$$\partial_i \partial_j \log S = \alpha S^{2\alpha} e_i(\omega) e_j(\omega).$$

By integrating this and comparing it with (3.6), we have (6.2), except for an arbitrary constant term. The case when $\alpha = 0$ is proved in a similar manner by taking into account that $\partial_i \partial_j \log S = 0$ holds in the 0-affine coordinate system. Since φ_α is the potential for the dual coordinate system $\eta = \mathbf{c}^{(-\alpha)}$, $\varphi_\alpha = \psi_{-\alpha}$ holds. The constant term in (6.2) can be determined as follows. We have from (5.6)

$$0 = \psi_\alpha(\theta) + \psi_{-\alpha}(\eta) - \theta \cdot \eta.$$

On the other hand, the following identity

$$\begin{aligned} \theta \cdot \eta &= \sum_t (2\pi)^{-2} \int R^{(\alpha)}(\omega) R^{(-\alpha)}(\omega') e_t(\omega')\, d\omega\, d\omega' \\ &= (2\pi)^{-1} \int R^{(\alpha)}(\omega) R^{(-\alpha)}(\omega)\, d\omega \\ &= (-1)/(2\pi\alpha^2) \int 1\, d\omega = -1/\alpha^2 \end{aligned} \tag{6.4}$$

holds for the α-affine coordinates θ and the $-\alpha$-affine coordinates η of one and the same system, where we used the Parceval relation

$$\sum_t e_t(\omega) e_t(\omega') = 2\pi\delta(\omega - \omega').$$

This shows that (5.6) is satisfied by (6.2). □

The entropy function H plays an important role in the system manifold. Its second derivative with respect to the α-coordinate system $\mathbf{c}^{(\alpha)}$ gives the metric tensor,

$$g_{ij}^{(\alpha)} = (2/\alpha)\partial_i \partial_j H,$$

expressed in this coordinate system. Identity (6.4) or

$$\sum c_t^{(\alpha)} c_t^{(-\alpha)} = -1/\alpha^2$$

is also important. When $\alpha = 1$, they reduce to a relation between the autocovariances and the inverse autocovariances. The α-divergence $D_\alpha(S_1, S_2)$ between two systems $S_1(\omega)$ and $S_2(\omega)$ is obtained as follows.

Theorem 8. *The α-divergence between two systems whose output spectral densities are $S_1(\omega)$ and $S_2(\omega)$ is given by*

$$D_\alpha(S_1, S_2) = \begin{cases} (2\pi\alpha^2)^{-1} \int \{(S_2/S_1)^\alpha - 1 - \alpha \log(S_2/S_1)\}\, d\omega, & \alpha \neq 0, \\ (4\pi)^{-1} \int (\log S_2 - \log S_1)^2\, d\omega, & \alpha = 0. \end{cases} \tag{6.5}$$

Proof. Let $\theta = \mathbf{c}^{(\alpha)}$ and $\eta = \mathbf{c}^{(-\alpha)}$ be the α and $-\alpha$ affine coordinate systems, respectively. Then the α-divergence is given by

$$D_\alpha(S_1, S_2) = \psi_\alpha(S_1) + \varphi_\alpha(S_2) - \theta_1 \cdot \eta_2,$$

where θ_1 is the θ-coordinate of S_1 and η_2 is the η-coordinate of S_2. From

$$\theta_1 \cdot \eta_2 = \sum_t (2\pi)^{-2} \int R_1^{(\alpha)} e_t(\omega)\, d\omega \int R_2^{(-\alpha)} e_t(\omega')\, d\omega'$$

$$= (2\pi)^{-1} \int R_1^{(\alpha)} R_2^{(-\alpha)}\, d\omega = -(2\pi\alpha^2)^{-1} \int (S_2/S_1)\, d\omega,$$

and (6.2), we have (6.5). □

The α-divergence is not symmetric in general. Instead, it satisfies the relation

$$D_\alpha(S_2, S_1) = D_{-\alpha}(S_1, S_2). \tag{6.6}$$

Hence, 0-divergence is symmetric. The -1-divergence is known as the Kullback-Leibler divergence. Hence, D_α is considered as its generalization. It is related to Renyi's α-entropy [27], [4]. For two close systems $S(\omega)$ and $S(\omega) + \delta S(\omega)$, their α-divergences are expanded as

$$D_\alpha(S, S + \delta S) = \frac{1}{4\pi} \int \left\{ \left(\frac{\delta S(\omega)}{S(\omega)} \right)^2 + \frac{\alpha - 3}{3} \left(\frac{\delta S}{S} \right)^3 + \cdots \right\} d\omega,$$

so that all the α-divergences are the same in the first approximation, giving half of the square of the Riemannian distance. Roughly speaking, (6.5) shows that the α-divergence $D_\alpha(S_1, S_2)$ for $\alpha > 0$ becomes small when S_2 approximates the valleys (zeros) of S_1 well, and $D_\alpha(S_1, S_2)$ for $\alpha < 0$ becomes small when S_2 approximates the peaks (poles) of $S_1(\omega)$ well.

7. Approximation, Identification, and Stochastic Realization

We have shown that L admits invariant α-divergence measures. The geometric structures of L are helpful and useful in solving such problems as approximations, identification, and stochastic realizations of systems, as will be shown in the present section. Given a model M_n and a system $S \in L$ which does not necessarily belong to M_n, the problem of approximation is to obtain $\hat{S}_n \in M_n$ which belongs to M_n and is closest to S under some criterion. When M_n's are a nesting family, $\hat{S}_n$ gives a series of the best approximations. Such a problem is related to the reduction of dimensionality of a system. We use the α-divergence measure to solve the approximation problem, which is called the α-approximation. When $-\alpha$-models are used, we can obtain the explicit form of the α-approximation. Moreover, the approximation errors are decomposed into a sum of those corresponding to each dimension.

The identification problem is to estimate the true unknown system S which is supposed to belong to a model M_n based on a set of observed outputs $\mathbf{x}_T = \{x_1, x_2, \ldots, x_T\}$ from the system, where T is supposed to be large. Let $\hat{S}$ be a nonparametric density estimator based on $\mathbf{x}_T$, for example, a smoothed version of the correlogram. Then the identification problem reduces to the problem of approximating $\hat{S}$ by one included in M_n. The efficiency of an estimator, in particular the higher-order efficiency, is fully studied in the problem of statistical inference [4], [5] in the framework of α-geometry. The metric and curvature play a fundamental role in the problem of estimation and testing. The results obtained in [4] and [5] are applicable to the present case of system identification, although we do not treat it in the present paper.

Let $\mathbf{c}^{(-1)}$ be the autocovariances of a system S. A system S_n is said to be a stochastic partial realization of $\mathbf{c}^{(-1)}$ of degree n, when the autocovariances of S_n coincide with those of a given $\mathbf{c}^{(-1)}$ up to $t=0, 1, \ldots, n$. It is known [10] that, when S_n belongs to the AR model $M_n^{(1)}$, its entropy is maximal among all the stochastic partial realizations. In order to study the geometric properties of the stochastic realization, we firstly generalize the problem to the α-stochastic partial realization; a problem of realizing the first $n+1$ components of $\mathbf{c}^{(\alpha)}=\mathbf{w}$ by a system. Let $Q_n^{(\alpha)}(\mathbf{w})$ be the set of all the α-realizations of a given $\mathbf{c}^{(\alpha)}=\mathbf{w}$. We obtain $Q_n^{(\alpha)}$ and show its geometric properties in the present section. The α-version of the maximum entropy principle is proved in this generalized problem.

The problem of α-approximation is easily solved within the geometrical framework: the best α-approximation to S in a model M_n is given by the α-projection of S to M_n, where the α-projection $\hat{S}_n$ implies the point in M_n such that the α-geodesic connecting S and $\hat{S}_n$ is orthogonal to M_n. Let $\mathbf{c}^{(\alpha)}$ be the α-coordinates of S, and let M_n be parametrized by $\mathbf{v}$. Let $\mathbf{c}^{(\alpha)}(\mathbf{v})$ and $\mathbf{c}^{(-\alpha)}(\mathbf{v})$ be, respectively, the α- and $-\alpha$-coordinates of a system $S(\mathbf{v})$ specified by $\mathbf{v}$ in M_p whose spectral density is $S(\omega, \mathbf{v})$. Then the α-geodesic connecting S and $S(\mathbf{v})$ is given by

$$\mathbf{c}^{(\alpha)}(t)=(1-t)\mathbf{c}^{(\alpha)}(\mathbf{v})+t\mathbf{c}^{(\alpha)}$$

in the α-coordinates. Hence, its tangent direction at $S(\mathbf{v})$ is given by

$$\dot{\mathbf{c}}^{(\alpha)}=(c_i^{(\alpha)}(\mathbf{v})-c_i)E_i^{(\alpha)}(\omega),$$

where $E_i^{(\alpha)}(\omega)$ is the natural basis corresponding to the α-coordinate system, i.e.,

$$E_i^{(\alpha)}(\omega)=[\partial/\partial c_i^{(\alpha)}]\log S=S^{-\alpha}e_i(\omega). \tag{7.1}$$

On the other hand, the tangent space of M_n is spanned by $n+1$ vectors $\partial/\partial v_i$,

$$\begin{aligned}(\partial/\partial v_i)\log S(\omega, \mathbf{v})&=[\partial c_j^{(-\alpha)}/\partial v_i][\partial/c_j^{(-\alpha)}]\log S\\&=[\partial c_j^{(-\alpha)}/\partial v_i]E_j^{(-\alpha)}(\omega), \qquad i=0, 1, \ldots, n.\end{aligned}$$

Therefore, the orthogonality condition of $\dot{c}^{(\alpha)}$ and $\partial/\partial v_i$ is written as

$$\sum_{i=0}^{n}[c_i^{(\alpha)}(\mathbf{v})-c_i][\partial c_i^{(-\alpha)}/\partial v_j]=0, \qquad j=0, \ldots, n, \tag{7.2}$$

where we used the orthogonality relation

$$\langle \mathbf{E}_i^{(-\alpha)}, \mathbf{E}_j^{(\alpha)}\rangle=\delta_{ij}.$$

This is the equation to determine the parameter $\mathbf{v}$ of the α-approximation to S in M_n.

When α-models are used we can get the explicit solution to the $-\alpha$-approximation problem. We discuss the ARMA models in Section 8.4. An α-model $M_n^{(\alpha)}$ is characterized by

$$c_{n+1}^{(\alpha)} = c_{n+2}^{(\alpha)} = \cdots = 0$$

in terms of the α-coordinate system $\mathbf{c}^{(\alpha)}$. They are linear restrictions so that $M_n^{(\alpha)}$ is completely α-flat. An α-affine coordinate system $\theta = (\theta^i)$ of $M_n^{(\alpha)}$ is given by the first $n+1$ components of $\mathbf{c}^{(\alpha)}$ as

$$\theta^i = c_i^{(\alpha)}, \qquad i = 0, 1, \ldots, n.$$

The dual affine coordinate system ($-\alpha$-coordinate system) of $M_n^{(\alpha)}$ is given by

$$\eta_i = c_i^{(-\alpha)}, \qquad i = 0, 1, \ldots, n.$$

For $\partial_i = \partial/\partial\theta^i$ and $\partial^i = \partial/\partial\eta_i$, the orthogonality condition

$$\langle \partial_i, \partial^j \rangle = \delta_i^j \tag{7.3}$$

holds. When $S(\omega)$ is a member of $M_n^{(\alpha)}$, its $\pm\alpha$-coordinates $\mathbf{c}^{(\alpha)}$ and $\mathbf{c}^{(-\alpha)}$ in the enveloping L are determined from θ and η as

$$\mathbf{c}^{(\alpha)} = (\theta, 0, 0, \ldots),$$

$$\mathbf{c}^{(-\alpha)} = (\eta, c_{n+1}^{(-\alpha)}, c_{n+2}^{(-\alpha)}, \ldots),$$

where $c_i^{(-\alpha)}$ $(i = n+1, n+2, \ldots)$ are functions of η.

Since $M_n^{(\alpha)}$ is α-flat, the $-\alpha$-approximation $\hat{S}_n$ to S by $M_n^{(\alpha)}$,

$$\min_{S' \in M_n^{(\alpha)}} D_{-\alpha}(S, S') = D_{-\alpha}(S, \hat{S}_n),$$

has a good geometric property. It is given by the $-\alpha$-projection of S to $M_n^{(\alpha)}$. Since $M_n^{(\alpha)}$ is completely α-flat, the $-\alpha$-projection $\hat{S}_n$ always exists and is unique.

Theorem 9. *The $-\alpha$-approximation $\hat{S}_n$ by $M_n^{(\alpha)}$ to a system S whose $-\alpha$-coordinates are $\mathbf{c}^{(-\alpha)}$, is given in the η-coordinates $\hat{\eta}$ by*

$$\hat{\eta}_i = c_i^{(-\alpha)}, \qquad i = 0, 1, \ldots, n.$$

The approximation error evaluated by the $-\alpha$-divergence $D_{-\alpha}(S, \hat{S}_n)$ is given by the difference of entropies,

$$D_{-\alpha}(S, \hat{S}_n) = (2/\alpha)\{H(\hat{S}_n) - H(S)\}. \tag{7.4}$$

Proof. The $-\alpha$-divergence from S to an $S_n \in M_n^{(\alpha)}$ is written as

$$D_{-\alpha}(S, S_n) = (2/\alpha)H(S_n) - (2/\alpha)H(S) - \theta \cdot \mathbf{c}^{(-\alpha)},$$

where

$$\theta \cdot \mathbf{c}^{(-\alpha)} = \sum_{i=0}^{n} \theta^i c_i^{(-\alpha)}.$$

By differentiating it with respect to the α-affine coordinates θ of S_n and putting the derivatives equal to 0, we have

$$\hat{\eta}_i = c_i^{(-\alpha)}, \qquad i = 0, \ldots, n,$$

in terms of the η-coordinates. This can be shown by a more intuitive geometrical way as follows. The $-\alpha$-geodesic connecting S and $\hat{S}_n$ in L is given by

$$\mathbf{c}^{(-\alpha)}(t) = (1-t)\hat{\mathbf{c}} + t\mathbf{c}^{(-\alpha)}$$

in terms of the $-\alpha$-coordinates in L, where $\hat{\mathbf{c}}$ is the $(-\alpha)$-coordinates of $\hat{S}_n$ in L. Hence, its tangent vector at $\hat{S}_n$ is given by

$$\dot{\mathbf{c}}^{(-\alpha)} = (\hat{c}_i - c_i^{(-\alpha)})\mathbf{E}^i.$$

The tangent space $T(M_n^\alpha)$ of $M_n^{(\alpha)}$ is spanned by $(n+1)$ vectors

$$\mathbf{E}_i = \partial/\partial\theta^i = \partial/\partial c_i^{(\alpha)}, \qquad i = 0, 1, \ldots, n.$$

Hence, from the condition that $\dot{\mathbf{c}}$ is orthogonal to $T(M_n^\alpha)$, we have

$$\langle \mathbf{E}_i, \dot{\mathbf{c}}^{(-\alpha)} \rangle = \hat{c}_i - c_i^{(-\alpha)} = 0, \qquad i = 0, 1, \ldots, n.$$

This shows that $\hat{S}_n$ is obtained by keeping the first $n+1$ components of $\mathbf{c}^{(-\alpha)}$ of S invariant and by changing only the other components such that $\hat{S}_n$ is included in $M_n^{(\alpha)}$. The entropy relation is obtained as

$$D_{-\alpha}(S, \hat{S}_n) = (2/\alpha)H(\hat{S}_n) - (2/\alpha)H(S) - \mathbf{c}^{(-\alpha)} \cdot \hat{\theta} - 1/\alpha^2,$$

because of (6.4), or

$$\mathbf{c}^{(-\alpha)} \cdot \hat{\theta} = \hat{\eta} \cdot \hat{\theta} = -1/\alpha^2. \qquad \square$$

Let $M^\alpha = \{M_n^{(\alpha)}\}$, $n = 0, 1, 2, \ldots,$ be a nested family of α-models. When a system S is given, we have a sequence $\hat{S}_0, \hat{S}_1, \hat{S}_2, \ldots$ of the $-\alpha$-approximation to S by $M_n^{(\alpha)}$. Obviously, $\hat{S}_{n+1}$ approximates S better than $\hat{S}_n$ does. It is interesting to see how the approximation errors decrease as the dimension number n of $M_n^{(\alpha)}$ increases. We can show that the approximation errors $D_{-\alpha}(S, \hat{S}_n)$ by $M_n^{(\alpha)}$ are decomposed into the sum of the approximation error $D_{-\alpha}(S, \hat{S}_{n+1})$ by $M_{n+1}^{(\alpha)}$ and the approximation error of the $\hat{S}_{n+1}$ by $M_n^{(\alpha)}$,

$$D_{-\alpha}(S, \hat{S}_n) = D_{-\alpha}(S, \hat{S}_{n+1}) + D_{-\alpha}(\hat{S}_{n+1}, \hat{S}_n)$$

(Fig. 5). This relation can be used in the problem of model selection or the problem of determining the degree n, as the AIC (Akaike's information criterion) is used. The following important theorem shows this in the form of the decomposition of the approximation error.

Theorem 10. *Let $\{\hat{S}_n\}$ $(n = 0, 1, \ldots)$ be the sequence of the $-\alpha$-approximations of S by $M_n^{(\alpha)}$. Then, the $\hat{S}_n$ is also the $-\alpha$-approximation of $\hat{S}_k$ by $M_n^{(\alpha)}$ when $k > n$. The approximation errors $D_{-\alpha}$ satisfy the additive relation*

$$D_{-\alpha}(S, \hat{S}_n) = D_{-\alpha}(S, \hat{S}_k) + D_{-\alpha}(\hat{S}_k, \hat{S}_n) \tag{7.5}$$

for $k > n$. In particular, $D_{-\alpha}(S, \hat{S}_0)$ is decomposed as

$$D_{-\alpha}(S, \hat{S}_0) = \sum_{i=1} D_{-\alpha}(\hat{S}_i, \hat{S}_{i-1}). \tag{7.6}$$

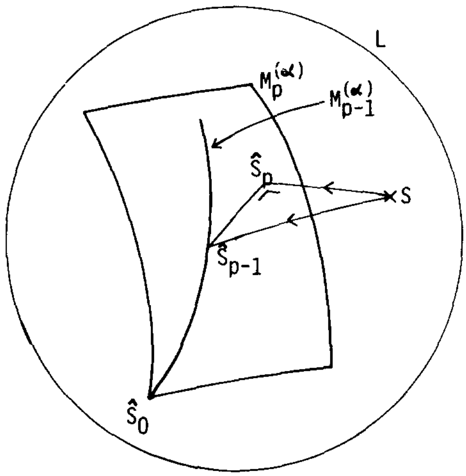

Fig. 5. Decomposition of approximation error.

Proof. $\hat{S}_k$ is an element of $M_k^{(\alpha)}$ whose first $(k+1)$ $-\alpha$-coordinates coincide with those of S. When $n \le k$, the $-\alpha$-projections of S and $\hat{S}_k$ to $M_n^{(\alpha)}$ coincide, because they have the same η-coordinates in $M_n^{(\alpha)}$. The decompositions (7.5) and (7.6) are proved from the relation

$$D_{-\alpha}(\hat{S}_k, \hat{S}_n) = (2/\alpha)\{H(\hat{S}_k) - H(\hat{S}_n)\}. \qquad \square$$

We next consider the manifold $Q_n^{(\alpha)}(\mathbf{w})$ of the α-stochastic partial realizations of degree n, i.e., the manifold consisting of all the systems whose α-coordinates $\mathbf{c}^{(\alpha)}$ are equal to a given $\mathbf{w}$ up to the nth component. It is obvious that $Q_n^{(\alpha)}(\mathbf{w})$ is given by the following set of linear equations in the α-coordinates $\mathbf{c}^{(\alpha)}$,

$$c_i^{(\alpha)} = w_i, \qquad i = 0, 1, \ldots, n.$$

Therefore, it is an infinite-dimensional α-flat manifold. The stochastic partial realization in a model M_n is obtained by solving the equation

$$c_i^{(\alpha)}(\mathbf{v}) = w_i, \tag{7.7}$$

where $\mathbf{v}$ is a vector parameter of M_n. The manifold $Q_n^{(\alpha)}(\mathbf{w})$ is explicitly given by using its realization in the $-\alpha$-model $M_n^{(-\alpha)}$. Let $\hat{S}_n$ be the α-realization of $\mathbf{w}$ in $M_n^{(-\alpha)}$. Then its α-coordinates are of the form

$$(w_0, w_1, \ldots, w_n; c_{n+1}^{(\alpha)}, \ldots),$$

where $c_{n+1}^{(\alpha)}$, etc., are determined such that $\hat{S}_n$ belongs to $M_n^{(-\alpha)}$, i.e., its $-\alpha$-coordinates are

$$(c_0^{(-\alpha)}, c_1^{(-\alpha)}, \ldots c_n^{(-\alpha)}; 0, 0, \ldots),$$

where $c_t^{(-\alpha)}$, $0 \le t \le n$, are determined from $w_0, \ldots, w_n$. The tangent space of $M_n^{(-\alpha)}$ is spanned by $n+1$ vectors $\mathbf{E}_0^{(-\alpha)}, \ldots, \mathbf{E}_1^{(-\alpha)}$, while the tangent space of $Q_n^{(\alpha)}(\mathbf{w})$ is spanned by $\mathbf{E}_{n+1}^{(\alpha)}, \mathbf{E}_{n+2}^{(\alpha)}, \ldots$. They are mutually orthogonal. This shows that $Q_n^{(\alpha)}(\mathbf{w})$ is the α-flat manifold orthogonal to $M_n^{(-\alpha)}$ at $\hat{S}_n \in M_n^{(-\alpha)}$. These $Q_n^{(\alpha)}(\mathbf{w})$ give an α-geodesic foliation of L (see Lauritzen [26] and Amari [4]).

The α-geodesic connecting $\hat{S}_n$ and any point S' in $Q_n^{(\alpha)}(\mathbf{w})$ is orthogonal to the $-\alpha$-geodesic connecting $\hat{S}_n$ and $\hat{S}_0$, which is the realization in $M_0^{(-\alpha)}$ and hence has a constant spectral density. Therefore, by the extended Pythagorean theorem, we have

$$D_\alpha(S', \hat{S}_0) = D_\alpha(\hat{S}_n, \hat{S}_0) + D_\alpha(S', \hat{S}_n).$$

This characterizes the realization in $M_n^{(-\alpha)}$: it minimizes $D_\alpha(\hat{S}_n, \hat{S}_0)$ among all the α-realizations of $\mathbf{w}$. This is the α-version of Burg's maximal entropy principle [10] because of

$$D_\alpha(\hat{S}_n, \hat{S}_0) = (2/\alpha)\{H(\hat{S}_0) - H(\hat{S}_n)\}.$$

It is known that the stochastic realization of S in an AR model ($\alpha = 1$) is characterized by the fact that the entropy $H(\hat{S})$ is maximal among all the $\alpha = -1$ stochastic realizations of given autocovariances $\mathbf{c}^{(-1)}$. This is called the maximum entropy principle [10]. The -1-projection is also used in Shore [30] and Johnson and Shore [20]. It can be generalized as follows.

Theorem 11. *When $\alpha > 0$, among all the $-\alpha$-realizations, the realization $\hat{S}_n$ in the α-model $M_n^{(\alpha)}$ has the maximum entropy. When $\alpha < 0$, it has the minimum entropy.*

8. Topics on Geometry of System Manifold L

8.1. Lie Group Structure of System Manifold L

The manifold L of systems permits an algebraic structure: the product for two elements $S_1(\omega)$ and $S_2(\omega)$ is defined by $S(\omega) = S_1(\omega)S_2(\omega)$ and also belongs to L. By this multiplication, L forms a group, where the identity element is $S_0(\omega) = 1$, and the inverse element of $S(\omega)$ is $S(\omega)^{-1}$. The product of two systems S_1 and S_2 corresponds to their concatenation or the connection in series from the system-theoretic point of view. In the present scalar-input and scalar-output case, the multiplication is commutative. Since the multiplication is analytic in L, the system manifold L forms a Lie group by this multiplication.

Let us fix an element $R(\omega)$. Then R defines a bijective mapping $\tilde{R}$ from L to itself,

$$\tilde{R}: S \mapsto RS.$$

This mapping induces a transformation $\tilde{R}_*$ from the tangent space T_S of L at S to the tangent space T_{RS} of L at RS. Let $S(\omega; \mathbf{u})$ be the parametrized form of S. Then

$$E_i(\omega; \mathbf{u}) = \partial_i \log S(\omega; \mathbf{u})$$

are the logarithmic (i.e., $\alpha = 0$) expressions of the basis vectors of T_S associated with the coordinate system $\mathbf{u}$. Since $S(\omega)$ is transformed to $R(\omega)S(\omega; \mathbf{u})$ by $\tilde{R}$, the basis vector $E_i(\omega; \mathbf{u})$ of T_S are mapped by $\tilde{R}_*$ to the same $E_i(\omega; \mathbf{u}) \in T_{RS}$, because of the logarithmic expression,

$$\partial_i\{\log R(\omega)S(\omega; \mathbf{u})\} = E_i(\omega; \mathbf{u}).$$

Since the geometric structures (the metric and the α-connections) are defined by using the functions $E_i(\omega; \mathbf{u})$, they are kept invariant by Lie group multiplication. More formally, for two vector fields A and B, their inner product and α-covariant derivatives are kept invariant as

$$\langle A, B\rangle_S = \langle R_* A, R_* B\rangle_{\tilde{R}S}, \tag{8.1}$$

$$\tilde{R}_* \nabla_A^{(\alpha)} B = \nabla_{\tilde{R}_* A}^{(\alpha)}(\tilde{R}_* B). \tag{8.2}$$

One may say that the α-geometric structures are introduced in L such that they are compatible with the Lie group structure of L. There remain many problems to be studied further from this point of view.

8.2. *Transformations of a Parametric Model*

Let $M_n = \{S(\omega, \theta)\}$ be a parametric model. Then a transformation $\tilde{R}$ maps M_n to another parametric model

$$\tilde{R}M_n = \{R(\omega)S(\omega, \theta)\}.$$

Obviously, $\tilde{R}M_n$ is a concatenation of a fixed system R and a parametric model M_n. From the previous considerations it is easy to show that M_n and $\tilde{R}M_n$ are isomorphic in the sense of α-geometry. Hence, if M_n is completely α-flat (or $-\alpha$-flat), so is $\tilde{R}M_n$. Let $\Pi_{\tilde{R}M_n}^{(\alpha)} S$ be the α-approximation or the α-projection of a system S on an induced model $\tilde{R}M_n$. It is obtained by using the α-projection $\Pi_M^{(\alpha)}$ as follows:

$$\Pi_{\tilde{R}M}^{(\alpha)} S = R\Pi_M^{(\alpha)}(R^{-1}S). \tag{8.3}$$

8.3. *Compositions of Models*

Given two models $M_n = \{S(\omega; \theta)\}$ and $M'_{n'} = \{S(\omega; \xi)\}$ parametrized by θ and ξ, respectively, we can compose a new model parametrized by (θ, ξ) by connecting the two in series, in parallel, or via a feedback loop. The composed models preserve some of the geometrical properties of the component models. We consider the serial connection or concatenation of two systems, $M_{n+n'} = S(\omega; \theta, \xi)$,

$$S(\omega; \theta, \xi) = S(\omega, \theta)S(\omega, \xi).$$

When θ is fixed at θ_0 in the above model, we have an n'-dimensional submanifold

$$M'(\theta_0) = \{S(\omega, \theta_0)S(\omega, \xi)\}$$

parametrized by ξ. This is obtained from the model $\{S(\omega, \xi)\}$ by multiplying the fixed $S(\omega, \theta_0)$, so that all of $M'(\theta_0)$ are isomorphic to the component model $S(\omega, \xi)$. Similarly, all the submanifolds

$$M(\xi_0) = \{S(\omega, \theta)S(\omega, \xi_0)\}$$

are isomorphic to $\{S(\omega, \theta)\}$.

Let us calculate the Fisher information or the metric tensor of the composite model. By putting $\partial_i = \partial/\partial\theta^i$, $\partial'_i = \partial/\partial\xi^i$, we have

$$\mathbf{E}_i = \partial_i \log S(\omega; \theta, \xi) = \partial_i \log S(\omega, \theta),$$

$$\mathbf{E}'_i = \partial'_i \log S(\omega; \theta, \xi) = \partial'_i \log S(\omega, \xi).$$

Hence, the partitioned form

$$\begin{bmatrix} g_{ij} & g'_{ij} \\ g'_{ij} & g''_{ij} \end{bmatrix}$$

of the Fisher information matrix is given by

$$g_{ij} = \langle \mathbf{E}_i, \mathbf{E}_j \rangle,$$

$$g''_{ij} = \langle \mathbf{E}'_i, \mathbf{E}'_j \rangle,$$

$$g'_{ij} = \langle \mathbf{E}_i, \mathbf{E}'_j \rangle.$$

The θ-part g_{ij} depends only on θ, and the ξ-part g''_{ij} depends only on ξ. This shows that, if we know the true value ξ_0, the estimation error g^{ij} depends only on θ_0 and does not depend on the value of ξ_0. The same situation holds for the ξ-part. However, when both θ_0 and ξ_0 are unknown, the estimation errors of θ_0 and ξ_0 are, respectively, different from g^{ij} and g''^{ij}, because there is some interference between θ and ξ. Geometrically, it is given by the angle between $M(\xi_0)$ and $M'(\theta_0)$,

$$g'_{ij} = \langle \mathbf{E}_i, \mathbf{E}'_j \rangle = \frac{1}{2\pi} \int \partial_i \log S(\omega, \theta) \partial'_j \log S(\omega, \xi)\, d\omega, \tag{8.4}$$

which does not vanish in general. The estimation error of θ is given in this case by

$$(g_{ij} - g''^{rs} g'_{ri} g'_{sj})^{-1}$$

so that

$$g_{ij} = g'_{ri} g'_{sj} g''^{rs} \tag{8.5}$$

represents the loss of information in estimating θ (in the sense of Fisher) due to the uncertainty of ξ.

The parallel or feedback connection of two models seems geometrically more complicated.

8.4. *On ARMA Models*

An ARMA model $S_{p,q}$ of degrees (p, q) consists of those systems whose transfer functions are written as

$$H(z, \mathbf{a}, \mathbf{b}) = \sum_{i=0}^{q} b_i z^{-i} \Big/ \sum_{j=0}^{p} a_j z^{-j}, \tag{8.6}$$

where $b_0 \neq 0$, $a_0 = 1$, and the denominator and the numerator have no common roots. We may use $\mathbf{a} = (a_1, \ldots, a_p)$ and $\mathbf{b} = (b_0, \ldots, b_q)$ as a coordinate system of $S_{p,q}$ which forms a $(p+q+1)$-dimensional submanifold in L.

$S_{p,q}$ is a concatenation of an MA model $\{S(\omega, b)\}$ and an AR model $\{S(\omega, \mathbf{a})\}$. Therefore, when $\mathbf{a}$ is fixed at $\mathbf{a} = \mathbf{a}_0$, the resultant submanifold $\{S(\omega, \mathbf{a}_0, \mathbf{b})\}$ is $\alpha = -1$ flat. It has locally the same α-geometric structure as the MA model $\{S(\omega, \mathbf{b})\}$. However, they have different global structures, because the submodel $\{S(\omega, \mathbf{a}_0, \mathbf{b})\}$ does not include such $\mathbf{b}$ for which the denominator and the numerator in (8.6) include the same factor and cancellation occurs for $\mathbf{a} = \mathbf{a}_0$. These $\mathbf{b}$'s are given as solutions of algebraic equations determined from $\mathbf{a}_0$, and hence they form algebraic surfaces. Hence, $\{S(\omega, \mathbf{a}_0, \mathbf{b})\}$, where $\mathbf{a}_0$ is fixed, is a $(p+1)$-dimensional submanifold which is divided into a number of disconnected components by algebraic surfaces. A similar argument holds for a submanifold $\{S(\omega; \mathbf{a}, \mathbf{b}_0)\}$ obtained by fixing $\mathbf{b} = \mathbf{b}_0$.

It was Brockett [8] (see also Segal [29]) who first studied the topological properties of a family of linear systems of McMillan degree n, which is the (n, n) ARMA model $S_{n,n}$. He pointed out that $S_{n,n}$ consists of $n+1$ disjoint components. It is an interesting problem to study the global topological properties of a model M. It is another interesting problem to study the topological properties of M imbedded in L. We give a simple example to show this.

Let us consider a two-dimensional model M whose transfer functions are given by

$$H(z; a, b) = (z^{-1} + b)/(z^{-1} + a). \tag{8.7}$$

From the stability of H and H^{-1}, the coordinates (a, b) should satisfy

$$-1 < a < 1, \qquad -1 < b < 1.$$

Moreover, the condition $a \neq b$ is necessary, because otherwise cancellation of a zero and a pole occurs. Hence, the model M is homeomorphic to an open square from which a diagonal is deleted (Fig. 6(a)). However, if we study the shape of M in the enveloping manifold L, this representation will not prove metrically good. Let $\tilde{M}$ be the set consisting of M and the unit system $H_0 = 1$ which corresponds to the case with $a = b$, i.e., H_0 is derived from (8.7) by cancellation when $a = b$. All the diagonal elements (a, a) in the square representation correspond to the same system H_0. Hence, the representation of M should be topologically as Fig. 6(b), where two open discs, which together constitute M, are

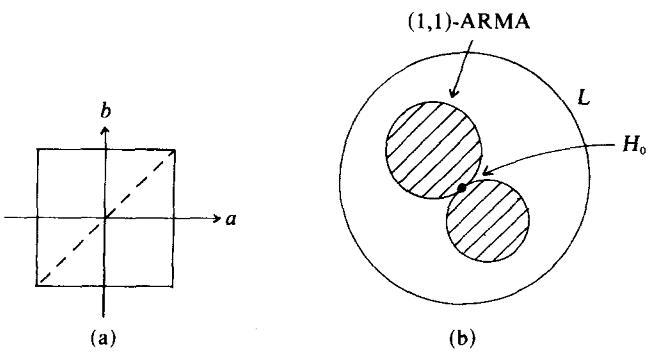

Fig. 6. Geometric shape of (1, 1)-ARMA model.

connected in L at one point H_0. In other words, the closure of M is connected in L. We should not disregard such properties that are related to the shape of M in L. This can also be confirmed from the metric property of M. As Brockett and Krishraprasad [9] pointed out, the metric tensor g_{ij} of M degenerates and becomes singular at $a=b$. This is because the length in the diagonal direction converges to 0 as a tends to b. It should be emphasized that the singular point H_0 is by itself a very natural system. What is singular is a model itself. The behavior of $S_{p,q}$ is singular at cancellation points, but the corresponding systems are not singular. Further, the coordinate system (a, b) is not so natural in a neighborhood of the singular point, because the Fisher information matrix becomes singular at that point. The ARMA model $S_{p,q}$ or the model $S_{n,n}$ of all the linear systems of McMillan degree n should be studied further from this geometrical point of view.

8.5. Generalizations

It is possible to apply the present method to the analysis of more general manifolds of linear and nonlinear systems. Let $L(\mathbf{u})$ be a system parametrized by a vector parameter $\mathbf{u}$, and let $\mathbf{x}=(x_t)$ be an infinite sequence of outputs of the system, when a white Gaussian noise is applied to the input. The α-geometrical structure is introduced in the manifold $\{L(\mathbf{u})\}$ essentially based on the probability measure $p(\mathbf{x}, \mathbf{u})$ parametrized by $\mathbf{u}$, as in the same manner as in a statistical manifold [4]. For example, when the output of a system is multiterminal and the spectral density $S(\omega, \mathbf{u})$ is a positive definite matrix, the metric and α-connection are given by

$$g_{ij}=(2\pi)^{-1}\int \operatorname{tr}(S^{-1}\partial_i S S^{-1}\partial_j S)\, d\omega,$$

$$\Gamma^{(\alpha)}_{ijk}=(2\pi)^{-1}\int \operatorname{tr}[S^{-1}(\partial_i\partial_j S)S^{-1}\partial_k S-\alpha S^{-1}\partial_i S S^{-1}\partial_j S S^{-1}\partial_k S]\, d\omega,$$

where tr is the trace of a matrix. When a system is not of minimal phase, we can use a non-Gaussian white noise to take effects of phase into account. An important point of the present method is that it is applicable to a manifold of nonlinear systems.

9. Conclusions

We have proposed a new differential geometrical method for analyzing properties of a model of systems. To this end, a Riemannian metric and dually coupled $\pm\alpha$-connections are introduced in a system manifold. We have analyzed the geometrical properties of manifolds of linear systems. A divergence measure is naturally introduced into the system manifold, and we solved the problem of approximating a system by one belonging to a model.

Acknowledgments

The author would like to thank Professor H. Kimura for his encouragment and Dr. M. Kumon for his discussions. The research was supported in part by Grant-in-Aid for Special Project Research on Environmental Sciences from the Japanese Ministry of Education, Science and Culture.

References

[1] S. Amari, Differential geometry of curved exponential families—curvatures and information loss, *Ann. Statist.*, **10** (1982), 357-387.

[2] S. Amari, A foundation of information geometry, *Electron. Comm. Japan*, **66** (1983), 1-10.

[3] S. Amari, Differential geometry of systems, Report No. 528, Institute of Mathematical Analysis, Kyoto University, 1984, pp. 235-253.

[4] S. Amari, *Differential Geometrical Methods in Statistics*, Lecture Notes in Statistics, vol. 28, Springer-Verlag, New York, 1985.

[5] S. Amari, Differential geometry of statistics—towards new developments, in *Differential Geometry in Statistical Inference*, IMS Monograph, Institute of Mathematical Statistics, Hayward, CA, 1987.

[6] R. J. Bhansali, The inverse partial correlation function of a time series and its application, *J. Multivariate Anal.*, **13** (1983), 310-327.

[7] P. Bloomfield, An exponential model for the spectrum of a scalar time series, *Biometrika*, **60** (1973), 217-226.

[8] R. W. Brockett, Some geometric questions in the theory of linear systems, *IEEE Trans. Automat. Control*, **21** (1976), 449-455.

[9] R. W. Brockett and P. S. Krishnaprasad, A scaling theory for linear systems, *IEEE Trans. Automat. Control*, **25** (1980), 197-207.

[10] J. P. Burg, Maximum entropy spectral analysis, *Proceedings of the 37th Meeting of the Society of Exploration Geophysicists*, 1967. Also in *Modern Spectrum Analysis* (D. G. Childers, ed.), IEEE Press, New York, 1978.

[11] C. I. Byrnes, Algebraic and geometric aspects of the analysis of feedback systems, in *Geometrical Methods for the Theory of Linear Systems* (C. I. Byrnes and C. F. Martin, eds), Reidel, Boston, 1980, pp. 85-124.

[12] N. N. Chentsov, *Statistical Decision and Optimal Inference*, Nauka, Moscow, 1972 (in Russian); translation in English, American Mathematical Society, Providence, RI, 1982.

[13] D. F. Delchamps, Global structure of families of multivariable linear systems with an application to identification, *Math. Systems Theory*, **18** (1985) 329-380.

[14] P. L. Duren, *Theory of H^p Spaces*, Academic Press, New York, 1970.

[15] E. J. Hannan, Estimating the dimension of a linear system, *J. Multivariate Anal.*, **11** (1981), 459-473.

[16] E. J. Hannan and M. Deistler, Some properties of the parametrization of ARMA systems with unknown order, *J. Multivariate Anal.*, **11** (1981), 474-484.

[17] M. Hazewinkel, Moduli and canonical forms for dynamical systems, II: the topological case, *Math. Systems Theory*, **10** (1977), 363-385.

[18] M. Hazewinkel, (Fine) Moduli (spaces) for linear systems: what are they and what are they good for?, in *Geometrical Methods for the Theory of Linear Systems* (C. I. Byrens and C. F. Martin, eds), Reidel, Boston, 1980, pp. 125-193.

[19] R. Hermann and C. F. Martin, Applications of algebraic geometry to systems theory: the McMillan degree and Kronecker indices of transfer functions as topological and holomorphic system invariants, *SIAM J. Control Optim.*, **16** (1978), 743-755.

[20] R. W. Johnson and E. Shore, Minimum cross entropy spectral analysis of multiple signals, *IEEE Trans. Acoust. Speech Signal Process.*, **31** (1983), 574-582.

[21] R. E. Kalman, Mathematical description of linear dynamical systems, *SIAM J. Control Optim.*, **1** (1963), 128-151.

[22] W. Klingenberg, *Riemmanian Geometry*, Walter de Gruyter, Berlin, 1982.

[23] M. Kumon and S. Amari, Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference, *Proc. Roy. Soc. London Ser. A*, **387** (1983), 429–458.

[24] M. Kumon and S. Amari, Estimation of a structural parameter in the presence of a large number of nuisance parameters, *Biometrika*, **71** (1984), 445–459.

[25] S. Lang, *Differentiable Manifolds*, Addison-Wesley, Reading, MA, 1972.

[26] S. Lauritzen, *Some Differential Geometrical Notions and Their Use in Statistical Theory*, IMS Monograph, Institute of Mathematical Statistics, Hayward, CA, 1987.

[27] H. Nagaoka and S. Amari, Differential geometry of smooth families of probability distributions, Technical Report METR 82-7, University of Tokyo, 1982.

[28] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1965.

[29] G. Segal, The topology of spaces of rational functions, *Acta Math.*, **143** (1979), 39–72.

[30] E. Shore, Minimum cross-entropy spectral analysis, *IEEE Trans. Acoust. Speech Signal Process.*, **29** (1981), 230–237.

[31] A. Tannenbaum, *Invariance and System Theory: Algebraic and Geometric Aspects*, Lecture Notes in Mathematics, vol. 845, Springer-Verlag, New York, 1981.

Received April 7, 1986, and in revised form December 31, 1986.