数理解析研究所講究録
第 528 巻 1984 年 235-254

Differential Geometry of Systems

東大 工学部 甘利俊一 (Shun-ichi Amari)

1. Introduction

When one treats a parametric family of systems such as an AR model, it is important to study not only properties of a specific system but those of a family itself. Since a parametric family of systems usually forms a geometric manifold imbedded in a larger family of systems, it is useful to study such differential-geometric structures of the manifold as Riemannian metric, affine connection, curvature, etc. The present note is a preliminary work for constructing a general differential-geometrical theory of systems. We introduce a Riemannian metric and a one-parameter family of affine connections. Characteristics of AR and MA families of systems are elucidated from this point of view.

The differential-geometrical properties of the manifold of a parametric family of probability distributions $S = \{p(x, \theta)\}$, where x is a random variable and p is a density function of x parametrized by a vector parameter θ, have fully been studied by a series of papers (Amari, 1982a, b; 1983a, b; Nagaoka and Amari, 1982, Amari, 1984). It has been proved that such a theory plays a fundamental role in studying asymptotic properties of statistical inference (Amari and Kumon, 1983; Kumon and Amari, 1983). Since a discrete-time stationary system can be regarded as a transformer of a sequence of input signals $\{\varepsilon_t\}$ into output signals $\{x_t\}$, $t = \ldots, -1, 0, 1, 2, \ldots$, when $\{\varepsilon_t\}$ is a white Gaussian noise, properties of the system are fully represented by stochastic properties of the stationary time-series $\{x_t\}$ produced by it. Hence, when we consider a parametric family of systems, their geometric properties are given by the stochastic properties of the related manifold of the parametrized stochastic processes which they produce. We can thus introduce a differential-geometric structure into a manifold of systems by using the method developed in differential geometry of statistics.

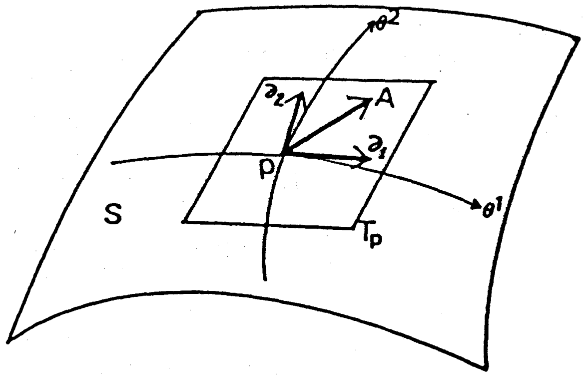
We first introduce a differential-geometrical theory of a family of probability distributions and show some fundamental properties based on Nagaoka and Amari, 1982. We then introduce a differential-geometric structure into a family of stationary Gaussian processes. This structure is then used to define the geometric structures of parametric family of systems. Some interesting properties of AR and MA models are shown by using the geometric concepts.

2. Geometrical Structure of Statistical Models

2.1. Metric and α-connection

Let $S = \{p(x, \theta)\}$ be a statistical model consisting of probability density functions $p(x, \theta)$ of random variable $x \in X$ with respect to a measure P of X such that every distribution is uniquely parametrized by an n-dimensional vector parameter $\theta = (\theta^i) = (\theta^1, \ldots, \theta^n)$. Then, under a certain regularity conditions, S is considered to be an n-dimensional

FIG. 1



manifold with a coordinate system θ. Let us denote by $\partial_i = \partial/\partial\theta^i$ the tangent vector of the i-th coordinate curve θ^i (Fig.1) at point θ. Then, n such tangent vectors $\{\partial_i\}$, i = 1, ..., n, span the tangent space T_θ at point θ of the manifold S. Any tangent vector $A \in T_\theta$ is a linear combination of the basis $\{\partial_i\}$,

$$A = A^i \partial_i ,$$

where A^i are the components of A and Einstein's summation convention is assumed throughout the paper. The tangent space T_θ is a linearized version of a small neighborhood at θ of S, and an infinitesimal vector $d\theta = d\theta^i \partial_i$ denotes the vector connecting two neighboring points θ and $\theta + d\theta$ or two neighboring distributions $p(x, \theta)$ and $p(x, \theta + d\theta)$.

Let us introduce a metric in the tangent space T_θ. It can be done by defining the inner product $g_{ij}(\theta) = \langle \partial_i, \partial_j \rangle$ of two basis vectors ∂_i and ∂_j at θ. To this end, we represent a vector $\partial_i \in T_\theta$ by a function $\partial_i \ell(x, \theta)$ in x, where $\ell(x, \theta) = \log p(x; \theta)$ and ∂_i(in $\partial_i \ell$) is the partial derivative $\partial/\partial\theta^i$. Then, it is natural to define the inner product by

$$g_{ij}(\theta) = \langle \partial_i, \partial_j \rangle = E_\theta[\partial_i \ell(x, \theta) \partial_j \ell(x, \theta)] , \qquad (2.1)$$

where E_θ denotes the expectation with respect to $p(x, \theta)$. This g_{ij} is Fisher information matrix. Two vectors A and B are orthogonal, when

$$\langle A, B \rangle = \langle A^i \partial_i, B^j \partial_j \rangle = A^i B^j g_{ij} = 0 .$$

It is sometimes necessary to compare a vector $A \in T_\theta$ of the tangent space T_θ at one point θ with a vector $B \in T_{\theta'}$ belonging to the tangent space $T_{\theta'}$ at another point θ'. This can be done by comparing the basis vectors $\{\partial_i\}$ at T_θ with the basis vectors $\{\partial_i'\}$ at $T_{\theta'}$. Since T_θ and $T_{\theta'}$ are two different vector spaces, two vectors A and B are not directly comparable, and we need some criterion to compare them. This can be done by introducing an affine connection, which maps a tangent space $T_{\theta+d\theta}$ at $\theta + d\theta$ to the tangent space T_θ at θ. The mapping should reduce to the identical map as $d\theta \to 0$. Let $m(\partial_j')$ be the map of $\partial_j' \in T_{\theta+d\theta}$ to T_θ. It is slightly different from $\partial_j \in T_\theta$. The vector

$$\nabla_{\partial_i} \partial_j = \lim_{d\theta \to 0} \frac{1}{d\theta^i} \left\{ m(\partial_j') - \partial_j \right\}$$

represents the rate with which the j-th basis vector $\partial_j \in T_\theta$ "intrinsically" changes as the point θ moves from θ to $\theta + d\theta$ (Fig.2) in the direction ∂_i.

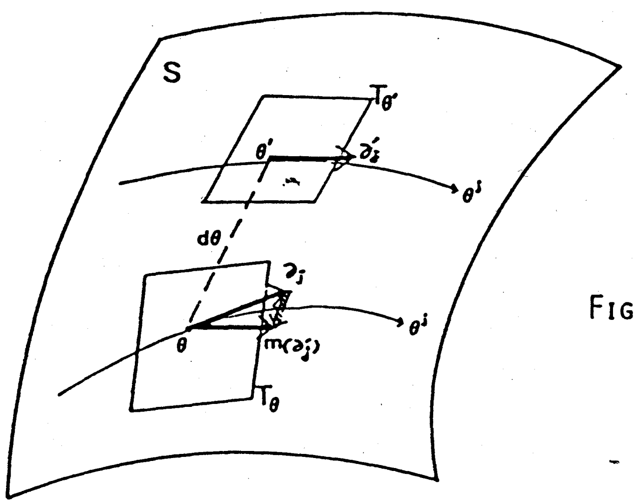


FIG. 2

We call $\nabla_{\partial_i}\partial_j$ the covariant derivative of the basis vector ∂_j in the direction ∂_i. Since it is a vector of T_θ, its components are given by

$$\Gamma_{ijk} = \langle \nabla_{\partial_i}\partial_j, \partial_k \rangle , \qquad (2.2)$$

and

$$\nabla_{\partial_i}\partial_j = \Gamma_{ij}{}^{k}\partial_k ,$$

where $\Gamma_{ijk} = \Gamma_{ij}{}^{m}g_{mk}$. We call Γ_{ijk} the components of the affine connection. An affine connection is specified by defining $\nabla_{\partial_i}\partial_j$ or Γ_{ijk}. Let $A(\theta)$ be a vector field, i.e., for every point $\theta \in S$ is defined a vector $A(\theta) \in T_\theta$. The intrinsic change of the vector $A(\theta)$ as the position θ moves is now given by the covariant derivative in the direction ∂_i of $A(\theta) = A^j(\theta)\partial_j$, defined by

$$\begin{aligned}\nabla_{\partial_i}A &= (\partial_i A^j)\partial_j + A^j(\nabla_{\partial_i}\partial_j)\\ &= (\partial_i A^j + \Gamma_{ik}{}^{j}A^k)\partial_j ,\end{aligned}$$

in which the change in the basis vectors is taken into account. The covariant derivative in the direction $B = B^i\partial_i$ is given by

$$\nabla_B A = B^i \nabla_{\partial_i} A .$$

We introduce the α-connection, where α is a real parameter, in the statistical manifold S by

$$\Gamma^{(\alpha)}_{ijk} = E_\theta[\{\partial_i\partial_j \ell(x,\theta) + \frac{1-\alpha}{2}\partial_i \ell(x,\theta)\partial_j \ell(x,\theta)\}\partial_k \ell(x,\theta)] . \qquad (2.3)$$

Especially, the 1-connection is called the exponential connection, and the -1-connection is called the mixture connection.

2.2. Duality in α-flat manifold

Once an affine connection is defined in S, we can compare two tangent vectors $A \in T_\theta$ and $A' \in T_{\theta'}$, belonging to different tangent spaces T_θ and $T_{\theta'}$ by the following parallel displacement of a vector. Let c : $\theta =$

$\theta(t)$ be a curve connecting two points θ and θ'. Let us consider a vector field $A(t) = A^i(t)\partial_i \in T_{\theta(t)}$ defined at each point $\theta(t)$ on the curve. If the vector $A(t)$ does not change along the curve, i.e., the covariant derivative of $A(t)$ in the direction $\dot{\theta}$ vanishes identically

$$\nabla_{\dot{\theta}} A(t) = \dot{A}^i(t) + \Gamma_{jk}^{\ \ i} A^k(t)\dot{\theta}^j = 0 ,$$

the field $A(t)$ is said to be a parallel vector field on c. Moreover, $A(t') \in T_{\theta(t')}$ at $\theta(t')$ is said to be a parallel displacement of $A(t) \in T_{\theta(t)}$ at $\theta(t)$. We can thus displace in parallel a vector $A \in T_\theta$ at θ to another point θ' along a curve $\theta(t)$ connecting θ and θ', by making a vector field $A(t)$ which satisfies $\nabla_{\dot{\theta}} A(t) = 0$, $\theta = \theta(0)$, $\theta' = \theta(1)$, and $A(0) = A \in T_\theta$. The vector $A' = A(1) \in T_{\theta'}$ at $\theta' = \theta(1)$ is the parallel displacement of A from θ to θ' along the curve c : $\theta = \theta(t)$. We denote it by $A' = \Pi_c A$. When the α-connection is used, we denote the α-parallel displacement operator by $\Pi_c^{(\alpha)}$. The parallel displacement of A from θ to θ' in general depends on the path c : $\theta(t)$ connecting θ and θ'. When this does not depend on paths, the manifold is said to be flat. It is known that a manifold is flat when, and only when, the Riemann-Christoffel curvature vanishes identically. A statistical manifold S is said to be α-flat, when it is flat under the α-connection.

The parallel displacement does not in general preserve the inner product, i.e., $\langle \Pi_c A, \Pi_c B \rangle = \langle A, B \rangle$ does not necessarily holds. When a manifold has two affine connections with corresponding parallel displacement operators Π_c and Π_c^*, and moreover when

$$\langle \Pi_c A, \Pi_c^* B \rangle = \langle A, B \rangle \tag{2.4}$$

holds, the two connections are said to be mutually dual. The two operators Π_c and Π_c^* are considered to be mutually adjoint. We have the following theorem in this regard.

Theorem 2.1. The α-connection and $-\alpha$-connection are mutually dual. When S is α-flat, it is also $-\alpha$-flat.

When a manifold S is α-flat, there exists a coordinate system (θ^i) such that

$$\nabla^{(\alpha)}_{\partial_i}\partial_j = 0 \qquad \text{or} \qquad \Gamma^{(\alpha)}_{ijk}(\theta) = 0$$

identically holds. In this case, a basis vector ∂_i is the same at any point θ in the sense that $\partial_i \in T_\theta$ is mapped to $\partial_i \in T_{\theta'}$ by the α-parallel displacement irrespective of the path connecting θ and θ'. Since all the coordinate curves θ^i are α-geodesics in this case, θ is called an α-affine coordinate system. A linear transformation of an α-affine coordinate system is also α-affine.

We give an example of a 1-flat (i.e., $\alpha = 1$) manifold S. The density functions of exponential family $S = \{p(x, \theta)\}$ can be written as

$$p(x, \theta) = \exp\{\theta^i x_i - \psi(\theta)\}$$

with respect to an appropriate measure, where $\theta = (\theta^i)$ is called the natural parameter. From

$$\partial_i \ell(x, \theta) = x_i - \partial_i \psi(\theta), \qquad -\partial_i\partial_j \ell(x, \theta) = \partial_i\partial_j\psi(\theta),$$

we easily have

$$g_{ij}(\theta) = \partial_i\partial_j\psi(\theta), \qquad \Gamma^{(\alpha)}_{ijk}(\theta) = \frac{1-\alpha}{2}\partial_i\partial_j\partial_k\psi.$$

Hence, 1-connection $\Gamma^{(1)}_{ijk}$ vanishes identically in the natural parameter, showing that θ gives a 1-affine coordinate system. A curve $\theta^i(t) = a^i t + b^i$, which is linear in the θ-coordinates, is a 1-geodesic.

Since an α-flat manifold is $-\alpha$-flat, there exists a $-\alpha$-flat coordinate system $\eta = (\eta_i) = (\eta_1, \ldots, \eta_n)$ in an α-flat manifold S. Let

$\partial^i = \partial/\partial\eta_i$ be the tangent vector of the coordinate curve η_i in the new coordinate system η. The vectors $\{\partial^i\}$ form a basis of the tangent space T_η or T_θ of S. When the two bases $\{\partial_i\}$ and $\{\partial^i\}$ of the tangent space T_θ satisfy

$$\langle \partial_i, \partial^j \rangle = \delta_i^j$$

at every point θ (or η), where δ_i^j is the Kronecker delta (denoting the unit matrix), the two coordinate systems θ and η are said to be mutually dual.

Theorem 2.2. When S is α-flat, there exists a pair of coordinate systems $\theta = (\theta^i)$ and $\eta = (\eta_i)$ such that i) θ is α-affine and η is $-\alpha$-affine, ii) θ and η are mutually dual, iii) there exist potential functions $\psi(\theta)$ and $\phi(\eta)$ such that the metric tensors are derived by differentiation as

$$g_{ij}(\theta) = \langle \partial_i, \partial_j \rangle = \partial_i\partial_j\psi(\theta) ,$$
$$g^{ij}(\eta) = \langle \partial^i, \partial^j \rangle = \partial^i\partial^j\phi(\eta) ,$$

where g_{ij} and g^{ij} are mutually inverse matrices so that

$$\partial_i = g_{ij}\partial^j , \qquad \partial^i = g^{ij}\partial_j$$

holds, iv) the coordinates are connected by a Legendre transformation

$$\theta^i = \partial^i\phi(\eta) , \qquad \eta_i = \partial_i\psi(\theta) \tag{2.5}$$

where the potentials satisfy the identity

$$\psi(\theta) + \phi(\eta) - \theta\cdot\eta = 0 , \tag{2.6}$$

where $\theta\cdot\eta = \theta^i\eta_i$.

In the case of exponential family S, the expectation parameter $\eta = (\eta_i)$

$$\eta_i = E_\theta[x_i] = \partial_i \psi(\theta)$$

is -1-affine, θ and η are mutually dual, and the dual potential $\phi(\eta)$ is given by the negative entropy,

$$\phi(\eta) = E[\log p] \ .$$

2. 3. α-divergence and α-projection

We can introduce the notion of α-divergence $D_\alpha(\theta, \theta')$ in an α-flat manifold S, which represents the degree of divergence from distribution $p(x, \theta)$ to $p(x, \theta')$. It is defined by

$$D_\alpha(\theta, \theta') = \psi(\theta) + \phi(\eta') - \theta\cdot\eta' \ , \tag{2.7}$$

where $\eta' = \eta(\theta')$ is the η-coordinates of the point θ', i.e., the η-coordinates of the distribution $p(x, \theta')$. The α-divergence satisfies $D_\alpha(\theta, \theta') \geq 0$ with the equality when and only when $\theta = \theta'$. The $-\alpha$-divergence satisfies $D_{-\alpha}(\theta, \theta') = D_\alpha(\theta', \theta)$. When S is exponential family, the -1-divergence is the Kullback-Leibler information,

$$D_{-1}(\theta, \theta') = I\{p(x, \theta') : p(x, \theta)\} = \int p(x, \theta)\log \frac{p(x, \theta)}{p(x, \theta')} \, dP \ .$$

When $S = \{p(x)\}$ is a function space of a non-parametric statistical model, the α-divergence is written as

$$D_\alpha\{p(x), q(x)\} = \frac{4}{1-\alpha^2} \left(1 - \int p(x)^{(1-\alpha)/2} q(x)^{(1+\alpha)/2} dP\right)$$

when $\alpha \neq \pm 1$, and is the Kullback information or its dual when $\alpha = -1$ or 1.

When θ and $\theta' = \theta + d\theta$ are infinitesimally close,

$$D_\alpha(\theta, \theta + d\theta) = \frac{1}{2} g_{ij}(\theta) d\theta^i d\theta^j \tag{2.8}$$

holds, so that it can be regarded as a generalization of a half of the square of the Riemannian distance, although neither symmetry nor the triangular inequality holds for D_α. However, the following Pythagorean theorem holds.

Theorem 2.3. Let c be an α-geodesic connecting two points θ and θ', and let c' be a $-\alpha$-geodesic connecting two points θ' and θ'' in an α-flat S. When the two curves c and c' intersect at θ' with a right angle, θ, θ' and θ'' forming a kind of a right triangle, the following Pythagorean relation holds,

$$D_\alpha(\theta, \theta') + D_\alpha(\theta', \theta'') = D_\alpha(\theta, \theta'') . \tag{2.9}$$

Let $M = \{q(x, u)\}$ be an m-dimensional submanifold imbedded in an α-flat n-dimensional manifold $S = \{p(x, \theta)\}$ by $\theta = \theta(u)$. For distribution $p(x, \theta_0) \in S$, we search for the distribution $q(x, \hat{u}) \in M$, which is the closest distribution in M from $P(x, \theta_0)$ in the sense of the α-divergence,

$$\min_{u \in M} D_\alpha\{\theta_0, \theta(u)\} = D\{\theta_0, \theta(\hat{u})\} .$$

We call the $\hat{u}(\theta_0)$ the α-approximation of $p(x, \theta_0)$ by M. It is important in many statistical problems to obtain the α-approximation, especially -1-approximation. Let c(u) be the α-geodesic connecting a point $\theta(u) \in M$ and θ_0, $c(u) : \theta = \theta(t; u)$, $\theta(u) = \theta(0, u)$, $\theta_0 = \theta(1, u)$. When the α-geodesic $c(\tilde{u})$ is orthogonal to M at $\theta(\tilde{u})$, i.e.,

$$\langle \dot{\theta}(0; \tilde{u}), \partial_a \rangle = 0,$$

where $\partial_a = \partial/\partial u^a$ are the basis vectors of $T_u(M)$, we call the $\tilde{u}$ the α-projection of θ_0 on M.

Theorem 2.4. The α-approximation $\hat{u}(\theta_0)$ of θ_0 by M is given by the α-projection $\tilde{u}(\theta_0)$ of θ_0 on M.

3. Geometrical Structure of Family of Systems

3.1. α-representation of systems

Let us consider the set of all linear regular stable invertible systems. When its input is drived by a white Gaussian noise, its output produces a stationary Gaussian process $\{x_t\}$. A system is hence characterized by the stochastic process which it produces, if the phase factor is neglected. A stationary Gaussian process is represented by the power spectrum $S(\omega)$. Hence, we first treat the set M of all the regular power spectrum functions $S(\omega)$. A spectrum $S(\omega)$ is connected with the autocovariances c_t of the process by

$$c_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\cos\omega t\, d\omega , \tag{3.1}$$

$$S(\omega) = c_0 + 2\sum_{t>0} c_t \cos\omega t , \tag{3.2}$$

where

$$c_t = E[x_r x_{r+t}]$$

for any r. A power spectrum $S(\omega)$ specifies a probability measure on the sample space $X = \{x_t\}$ of the stochastic processes. We study the geometrical structure of the manifold M of the probability measures given by $S(\omega)$. A specific parametric model, for example the AR model M_n^{AR} of order n, is treated as a submanifold imbedded in M.

Let us define the α-representation $\ell^{(\alpha)}(\omega)$ of a power spectrum $S(\omega)$ by

$$\ell^{(\alpha)}(\omega) = \begin{cases} -\frac{1}{\alpha}\{S(\omega)\}^{-\alpha}, & \alpha \neq 0 , \\ \log S(\omega) , & \alpha = 0 . \end{cases} \tag{3.3}$$

(Remark: It is better to define the α-representation by $-(1/\alpha)[S(\omega)^{-\alpha} - 1]$. However, the following discussions are the same for both representations.) We pose the following regularity conditions on the members of M that $\ell^{(\alpha)}$ can be expanded into the Fourier series for any α as

$$\ell^{(\alpha)}(\omega) = \xi_0^{(\alpha)} + 2 \sum_{t>0} \xi_t^{(\alpha)} \cos\omega t , \tag{3.4}$$

where

$$\xi_t^{(\alpha)} = \frac{1}{2\pi} \int \ell^{(\alpha)}(\omega) \cos\omega t d\omega , \qquad t = 0, 1, 2, \ldots .$$

We may write $\ell^{(\alpha)}(\omega)$ specified by $\xi^{\alpha} = \xi_t^{(\alpha)}$ as $\ell^{(\alpha)}(\omega; \xi^{(\alpha)})$. The infinite number of parameters $\{\xi_t^{(\alpha)}\}$ specify a power function by

$$S(\omega; \xi^{(\alpha)}) = \begin{cases} [-\alpha \ell^{(\alpha)}(\omega; \xi^{(\alpha)})]^{-1/\alpha} , & \alpha \neq 0 \\ \exp \ell^{(0)}(\omega; \xi^{(0)}) \end{cases} \tag{3.5}$$

so that they are regarded as defining an infinite-dimensional coordinate system in M. We call $\{\xi_t^{(\alpha)}\}$ the α-coordinate system of M. Obviously, the -1-coordinates are given by the autocovariances, $\xi_t^{(-1)} = c_t$

The negative of the 1-coordinates $\xi_t^{(1)}$, which are the Fourier coefficients of $S^{-1}(\omega)$, are denoted by $\tilde{c}_t$ and are called the inverse autocovariances, $\xi_t^{(1)} = -\tilde{c}_t$.

3.2. Geometry of parametric and non-parametric time-series models

Let M_n be a set of the power spectra $S(\omega; u)$ which are smoothly specified by an n-dimensional parameter $u = (u^a)$, $a = 1, 2, \ldots, n$. Then, M_n becomes (under a certain regularity condition) a submanifold of

M. This M_n is called a parametric time-series model. However, the members of M specified by an infinite-dimensional parameter u, e.g., by the α-coordinates $\xi^{(\alpha)} = \{ \xi_t^{(\alpha)},\ t = 0, 1, \dots \}$ in the form $S(\omega, \xi^{(\alpha)})$. The following discussions are hence common to both the parametric and non-parametric models, irrespective of the dimensions n of parameters.

Let us consider the tangent space T_u at u of M or M_n, which is spanned by an infinite number of or n basis vectors $\partial_a = \partial/\partial u^a$ associated with the coordinate system u. The α-representation of ∂_a is the following function in ω,

$$\partial_a = (\partial/\partial u^a)\ell^{(\alpha)}(\omega; u) .$$

Hence, in M, the basis $\partial_t^{(\alpha)}$ associated with the α-coordinates $\xi_t^{(\alpha)}$ is

$$\partial_t^{(\alpha)} = \begin{cases} 1, & t = 0 \\ 2\cos\omega t, & t \neq 0 . \end{cases}$$

Let us introduce the inner product g_{ab} of ∂_a and ∂_b in T_u by

$$g_{ab}(u) = \langle \partial_a, \partial_b \rangle = E_\alpha[\partial_a \ell^{(\alpha)}(\omega; u)\partial_b \ell^{(\alpha)}(\omega; u)] ,$$

where E_α is the operator defined at u by

$$E_\alpha[a(\omega)] = \int \{S(w; u)\}^{2\alpha} a(\omega) d\omega .$$

The above inner product does not depend on α, and written as

$$\langle \partial_a, \partial_b \rangle = \int \partial_a \log S(\omega, u) \partial_b \log S(\omega, u)\, d\omega. \qquad (3.6)$$

We next define the α-covariant derivative $\nabla^{(\alpha)}_{\partial_a}\partial_b$ of ∂_b in the direction of ∂_a by the projection of $\partial_a\partial_b \ell^{(\alpha)}$ to T_u. Then, the components of the α-connection are given by

$$\Gamma^{(\alpha)}_{abc}(u) = \langle \nabla^{(\alpha)}_{\partial_a}\partial_b, \partial_c \rangle = \int S^{2\alpha}\partial_a\partial_b \ell^{(\alpha)}\partial_c \ell^{(\alpha)} d\omega . \qquad (3.7)$$

If we use 0-representation, it is given by

$$\Gamma^{(\alpha)}_{abc}(u) = \int (\partial_a\partial_b \log S - \alpha\partial_a \log S\partial_b \log S)\partial_c \log S \, d\omega .$$

From (3.4) and (3.7), we can easily see that the α-connection vanishes in M identically, if the α-coordinate system $\xi^{(\alpha)}$ is used. Hence, we have

__Theorem 3.1__. The non-parametric M is α-flat for any α. The α-affine coordinate system is given by $\xi^{(\alpha)}$. The two-coordinate systems $\xi^{(\alpha)}$ and $\xi^{(-\alpha)}$ are mutually dual.

Since our definitions of the geometrical structure in M or M_n are the same as those introduced before in a family of probability distributions on sample space X, except that $X = \{x_t\}$ is infinite-dimensional in the present time-series case, we can define the α-divergence from $S_1(\omega)$ to $S_2(\omega)$ in M. It is calculated as follows

__Theorem 3.2__. The α-divergence from S_1 to S_2 is given by

$$D(S_1, S_2) = \begin{cases} (1/\alpha^2)\int \{[S_2(\omega)/S_1(\omega)]^\alpha - 1 - \alpha\log[S_2/S_1]\}d\omega, & \alpha \neq 0 \\ \end{cases}$$

$$\left| \quad (1/2) \int [\log S_1(\omega) - \log S_2(\omega)]^2 d\omega , \qquad \alpha = 0 .\right.$$

3.3. α-flat models

An α-model M_n^α of order n is a parametric model such that the α-representation of the power spectrum of a member in M_n^α is specified by n + 1 parameters $u = (u^k)$, k = 0, 1, ..., n, as

$$\ell^{(\alpha)}(\omega; u) = u_0 + 2\sum_{k=1}^{n} u_k \cos k\omega .$$

Obviously, M_n^α is α-flat (and hence $-\alpha$-flat), and u is its α-affine coordinate system.

The AR-model M_n^{AR} of order n consists of the stochastic processes defined recursively by

$$\sum_{k=0}^{n} a_k x_{t-k} = \varepsilon_t$$

where ε_t is a white Gaussian process with unit variance and $a = (a_0, a_1, \ldots, a_n)$ is the (n+1)-dimensional parameter specifying the members of M_n^{AR}. Hence, it is an (n+1)-dimensional submanifold of M. The power spectrum $S(\omega; a)$ of the process specified by a is given by

$$S(\omega; a) = 1 / \left| \sum_{k=0}^{n} a_k e^{ik\omega} \right|^2 .$$

We can calculate the geometric quantities of M_n^{AR} in terms of the AR-coordinate system a from the above expression. Similarly, the MA-model M_n^{MA} of order n is defined by the processes

$$x_t = \sum_{k=0}^{n} b_k \varepsilon_{t-k}$$

where $b = (b_0, b_1, \ldots, b_n)$ is the MA-parameter. The power spectrum $S(\omega; b)$ of the process specified by b is

$$S(\omega;\ b) = \left| \sum_{k} b_k e^{ik\omega} \right|^2 .$$

The exponential model M_n^{EXP} of order n introduced by Bloomfield is composed of the following power spectra $S(\omega;\ e)$ parameterized by $e = (e_0, e_1, \ldots, e_n)$,

$$S(\omega;\ e) = \exp\left\{ e_0 + \sum_{k=0}^{n} e_k \cos k\omega \right\} .$$

It is easy to show that the 1-representation of $S(\omega;\ a)$ in M_n^{AR} is given by

$$\tilde{c}_k = \sum_{t=k}^{n} a_t a_{t-k} , \qquad k = 0, 1, \ldots, n$$

$$\tilde{c}_k = 0 , \qquad k > n$$

where

$$\ell^{(1)}(\omega;\ a) = - S^{-1}(\omega;\ a) = \sum_{k} \tilde{c}_k e^{i\omega k} .$$

This shows that M_n^{AR} is a submanifold specified by $\tilde{c}_k = 0$, $(k > n)$ in M. Hence, it coincides with $M_n^{(1)}$, although the coordinate system a is not 1-affine but curved. Similar discussions hold for M_n^{MA}.

Theorem 3.3. The AR-model M_n^{AR} coincides with $M_n^{(1)}$, and hence is ± 1-flat. The MA-model M_n^{MA} coincides with $M_n^{(-1)}$, and hence is also ± 1-flat. The exponential model M_n^{EXP} coincides with $M_n^{(0)}$, and 0-flat. Since it is metric, it is an (n+1)-dimensional Euclidean space with an orthogonal Cartesian coordinate system e.

3.4. α-approximation and α-projection

Given a parametric model $M_n = \{S(\omega; u)\}$, it is sometimes necessary to approximate a spectrum $S(\omega)$ by one belonging to M_n. For example, given a finite observations $x_1, \dots, x_T$ of $\{x_t\}$, one tries to estimate u in the parametric model M_n by obtaining first a non-parametric estimate $\hat{S}(\omega)$ based on $x_1, \dots, x_T$ and then approximating it by $S(\omega; \hat{u}) \in M_n$. The α-approximation of S is the one that minimizes the α-divergence $D_\alpha[S(\omega), S(\omega, u)]$, $u \in M$. It is well known that the -1-approximation is related to the maximum likelihood principle. As we have shown in §2, the α-approximation is given by the α-projection of $S(\omega)$ to M_n. We now discuss the accuracy of the α-approximation. To this end, we consider a family of nested models M_n such that $M_0 \subset M_1 \subset M_2 \subset \dots \subset M_\infty = M$. The $\{M_n^{AR}\}$, $\{M_n^{MA}\}$ and $\{M_n^{EXP}\}$ are nested models, in which M_0 is composed of the white noises of various powers.

Let $\{M_n^\alpha\}$ be a family of the α-flat nested models, and let $\hat{S}_n(\omega; \hat{u}_n) \in M_n$ be the $-\alpha$-approximation of $S(\omega)$, where $\hat{u}_n$ is the (n+1)-dimensional parameter given by

$$\min_{S_n \in M_n^\alpha} D_{-\alpha}\{S, S_n(\omega)\} = D_{-\alpha}\{S, S_n(\omega; \hat{u}_n)\} .$$

The error of the approximation by $\hat{S}_n \in M_n$ is measured by the $-\alpha$-divergence $D_{-\alpha}(S, \hat{S}_n)$. We define

$$E_n(S) = \min_{S_n \in M_n^\alpha} D_{-\alpha}(S, S_n) = D_{-\alpha}(S, \hat{S}_n) . \tag{3.8}$$

It is an intersting problem to know how $E_n(S)$ decreases as n increases.

<u>Theorem 3.4</u>. The approximation error $E_n(S)$ of S is decomposed as

$$E_n(S) = \sum_{k=n}^{\infty} D_{-\alpha}(\hat{S}_{k+1}, \hat{S}_k) . \tag{3.9}$$

Hence,

$$D_{-\alpha}(S, \hat{S}_0) = \sum_{n=0}^{\infty} D_{-\alpha}(\hat{S}_{n+1}, \hat{S}_n) .$$

The theorem is proved by the Pythagorean relation for the right triangle $\Delta S\hat{S}_n\hat{S}_0$ composed of the α-geodesic $\hat{S}_n\hat{S}_0$ included in M_n^α and $-\alpha$-geodesic $S\hat{S}_n$ intersecting at $\hat{S}_n$ perpendicularly. The theorem shows that the approximation error $E_n(S)$ is decomposed into the sum of the $-\alpha$-divergences of the successive approximations $\hat{S}_k$, $k = n+1, \ldots, \infty$, where $\hat{S}_\infty = S$ is assumed. Moreover, we can prove that the $-\alpha$-approximation of $\hat{S}_k$ in M_n^α $(n < k)$ is $\hat{S}_n$. In other words, the sequence $\{\hat{S}_n\}$ of the approximations of S has the following property that $\hat{S}_n$ is the best approximation of $\hat{S}_k$ $(k > n)$ and that the approximation error $E_n(S)$ is decomposed into the sum of the $-\alpha$-divergences between the further successive approximations.

Let us consider the family M_n^{AR} of the AR-models. It coincides with M_n^1. Let S_n be the -1-approximation of S. Let $c_t(S)$ and $\tilde{c}_t(S)$ be, respectively, the autocovariances and inverse autocovariances. Since c_t

and $\tilde{c}_t$ are the mutually dual -1-affine and 1-affine coordinate systems, the -1-approximation $\hat{S}_n$ of S is determined by the following relations

1) $c_t(\hat{S}_n) = c_t(S)$, $t = 0, 1, \ldots, n$

2) $\tilde{c}_t(\hat{S}_n) = 0$, $t = n+1, n+2, \ldots$.

This implies that the autocovariances of $\hat{S}_n$ is the same as those of S up to t = n, and that the inverse autocovariances $\tilde{c}_t$ of S_n vanish for $t > n$. Similar relations hold for any other α-flat nested models, where c_t and $\tilde{c}_t$ are replaced by the dual pair of α- and $-\alpha$-affine coordinates. Especially, since $\{M_n^{EXP}\}$ are the nested Euclidean submanifolds with the self-dual coordinates $\xi^{(0)}$, their properties are extremely simple.

We have shown some fundamental properties of α-flat nested parametric models. These properties seem to be useful for constructing the theory of estimation and approximation of time-series. The ARMA-model, which is not α-flat for any α, has also an interesting global and local geometrical properties, although we do not discuss about them here.

4. References

Amari, S. (1982a). Differential geometry of curved exponential families --- curvatures and information loss. Ann. Statist., 10, 357-387

Amari, S. (1982b). Geometrical theory of asymptotic ancillarity and conditional inference. Biometrika, 69, 1-17

Amari, S. (1983a). Comparisons of asymptotically efficient tests in terms of geometry of statistical structures. Invited paper at 44th Session of the ISI, Madrid

Amari, S. (1983b). Differential geometry of statistical inference, Probability Theory and Mathematical Statistics (ed. Ito, K. and and Prokhorov, J.V.), Springer Lecture Notes in Math., 1021, 26-40

Amari, S. (1984). Differential geometry of statistics, METR 84-1, Univ. Tokyo.

Amari, S. and Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family, Ann. Inst. Statist. Math., 35A, 1-24

Kumon, M. and Amari, S. (1983). Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference, Proc. Roy. Soc. London, A387, 429-458

Nagaoka, H. and Amari, S. (1982). Differential geometry of smooth families of probability distributions, METR 82-7, Univ. Tokyo