Ann. Inst. Statist. Math.
Vol. 43, No. 4, 621–646 (1991)

DIFFERENTIAL GEOMETRICAL STRUCTURES RELATED TO FORECASTING ERROR VARIANCE RATIOS

DAMING XU

Department of Mathematics, University of Oregon, Eugene, OR 97403-1222, U.S.A.

(Received February 6, 1989; revised July 16, 1990)

Abstract. Differential geometrical structures (Riemannian metrics, pairs of dual affine connections, divergences and yokes) related to multi-step forecasting error variance ratios are introduced to a manifold of stochastic linear systems. They are generalized to nonstationary cases. The problem of approximating a given time series by a specific model is discussed. As examples, we use the established scheme to discuss the AR (1) approximations and the exponential smoothing of ARMA series for multi-step forecasting purpose. In the process, some interesting results about spectral density functions are derived and applied.

Key words and phrases: Riemannian metric, affine connection, divergence, spectral density, forecasting error variance ratio, yoke.

1. Introduction

Differential geometrical methods in statistics have been clearly established (Barndorff-Nielson *et al.* (1986), Amari *et al.* (1987), Kass (1989)). In Barndorff-Nielsen *et al.* (1986), it was pointed out that applications of differential geometry to time series models raised special problems, and they particularly posed the following questions: are differential geometric notions useful in connection with prediction of future observations and with behavior under incorrect models? Amari (1984, 1986, 1987*a*, 1987*b*) discussed differential geometrical structures on a manifold of linear systems. He introduced a Riemannian metric, a system of affine connections—α-connections, and corresponding α-divergences.

For a stationary zero-mean Gaussian process

$$X_t = \sum_{u=0}^{\infty} \kappa_u \epsilon_{t-u}, \tag{1.1}$$

where $\kappa_0 = 1$, all the κ_u's are real, $\sum_{u=0}^{\infty} \kappa_u^2 < \infty$ and $\{\epsilon_t\}$ is a series of Gaussian white noise with variance σ_ϵ^2, it is well known that its spectral density function is

$$S(\omega) = \frac{\sigma_\epsilon^2}{2\pi} h(\omega) \tag{1.2}$$

for $\omega \in [-\pi, \pi]$, where

$$h(\omega) = \left| \sum_{u=0}^{\infty} \kappa_u e^{-iu\omega} \right|^2. \tag{1.3}$$

The infinite-dimensional manifold L considered by Amari consists of all the $\{X_t\}$'s in (1.1) with its spectral density $S(\omega)$ continuous and satisfying

$$0 < S(\omega) < \infty. \tag{1.4}$$

Since we are only interested in the stochastic properties of $\{X_t\}$, which are completely described by its spectral density, our manifold L is identified as the set of all the continuous $S(\omega)$'s which satisfy (1.2)–(1.4). (Here several things are a little different from Amari's. In order to relate our geometry to forecasting problems, we quit Amari's restriction that ϵ_t's variance is equal to 1 and instead assume that $\kappa_0 = 1$. Also we have an extra factor, $1/2\pi$, for spectral density. All these differences do not affect the geometric structures discussed here.) Amari appropriately introduced

$$g_{uv} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \partial_u \log S(\omega)\, \partial_v \log S(\omega)\, d\omega$$

as the Riemannian metric at $S(\omega)$, where ∂_t denotes the partial derivative $\partial/\partial c_t$ and $\{c_t\}$ are the coordinates. Thus the spectral density plays a role similar to that of the likelihood function in the i.i.d. case. Among his α-connections, the most important two are the $+1$-connection and the -1-connection. The corresponding divergences are the $+1$-divergence

$$D_1(S_1, S_2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \frac{S_2}{S_1} - 1 - \log \frac{S_2}{S_1} \right\} d\omega$$

and the -1-divergence

$$D_{-1}(S_1, S_2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \frac{S_1}{S_2} - 1 - \log \frac{S_1}{S_2} \right\} d\omega.$$

Since $D_1(S_1, S_2) = D_{-1}(S_2, S_1)$, we only discuss D_{-1} in the following.

One of the primary goals in time series modelling is forecasting. How are our geometric structures related to forecasting problems? Let σ_{pq}^2 be the one-step forecasting error variance when S_q is used for forecasting of a series generated by S_p. The ratio $r_{12} = \sigma_{12}^2/\sigma_{11}^2$, called the one-step forecasting error variance ratio of S_2 versus S_1, equals 1 when $S_2 = S_1$, and is bigger than 1 when $S_2 \neq S_1$. And $\log r_{12}$ is some sort of "distance" measure of S_2 from S_1, which indicates how well S_2 approximates S_1 for the purpose of one-step prediction. As shown by Lemma 1.1 in Xu (1988), we have

$$\inf_{\sigma_1^2/\sigma_2^2} D_{-1}(S_1, S_2) = \log r_{12}, \tag{1.5}$$

where σ_p^2 is the noise variance of S_p. Since $D_{-1}(S_1, S_2)$ depends on σ_1^2 and σ_2^2 only through σ_1^2/σ_2^2, $\inf_{\sigma_1^2/\sigma_2^2} D_{-1}(S_1, S_2)$ is the same as $\inf_{\sigma_2^2} D_{-1}(S_1, S_2)$, and hence it can be understood as the -1-divergence from S_1 of the set of all the systems in L which are the same as S_2 except for possible different noise variances. In this way, Amari's $+1$- and -1-geometric structures are related to the one-step forecasting problem. In Xu (1988) and Tiao and Xu (1990), we have elucidated that not only the one-step forecasting error variance ratio but also the multi-step forecasting error variance ratios play a role in describing the discrepancy between time series systems. Thus a natural question is whether we can find differential geometrical connections and divergences on L which relate to the multi-step forecasting error variance ratios. The present paper aims at giving an answer to this question.

Unfortunately, we can not directly deal with the multi-step forecasting error variance ratios, because they solely are not necessarily capable of determining divergences. A trivial example is that the two-step forecasting error variance ratios between any two zero-mean MA (1) time series systems (see Box and Jenkins (1976) for the succinct notations AR, MA and ARIMA etc.) are always equal to 1, hence the corresponding divergence, if it could be determined, would be equal to 0, no matter whether these two systems are the same or not. One possible way to cope with this problem is to first find differential geometrical connections and divergences related to

$$\frac{\lambda\sigma_{12}^{(1)2} + (1-\lambda)\sigma_{12}^{(l)2}}{\lambda\sigma_{11}^{(1)2} + (1-\lambda)\sigma_{11}^{(l)2}}, \tag{1.6}$$

where $0 < \lambda \leq 1$, l is a positive integer, and $\sigma_{pq}^{(l)2}$ is the l-step forecasting error variance when the S_q system is used for prediction of a series generated by S_p. The limiting positions of geodesics and projections as $\lambda \to 0$ are what we need for l-step forecasting problems. (Here we have an open question: Do the limiting positions of geodesics and projections exist? In Sections 5 and 6 we will see that the answer is positive for some interesting examples.) The Riemannian metric, divergence, a pair of dual affine connections and a pair of dual affine coordinate systems (Spivak (1979), Amari (1987a), Lauritzen (1987)) corresponding to (1.6) are called the (l, λ)-metric, the (l, λ)-divergence, the (l, λ)- and $(l, \lambda)^*$-connections and the (l, λ)- and $(l, \lambda)^*$-coordinates, respectively.

In Section 2, some preparative results are discussed. We introduce the Riemannian metric related to (1.6) in Section 3, and give the corresponding affine coordinates, potential functions and connections, justify the divergence and show the duality in Section 4. Some interesting results about spectral density functions are derived. They play an important role in establishing our differential geometric framework and also have their independent interest. Section 5 is devoted to the problem of approximating a given time series by a specific model, and the method developed is used to discuss the AR (1) approximations. The generalization of the geometrical structures to nonstationary cases is discussed in Section 6, with the exponential smoothing for multi-step forecasting as an example. Section 7 gives some concluding remarks. The yokes and the related differential geometrical structures are mentioned.

Throughout this paper, i is reserved for the imaginary unit, i.e. $\sqrt{-1}$, l the number of steps in forecasting and λ the weight of linear combinations of the forecasting error variances. We will use $\{\theta^j, j = 0, 1, \dots\}$ to denote the (l, λ)-coordinates, $j, k, m, \dots$ for their indices, $\{\eta_\alpha, \alpha = 0, 1, \dots\}$ and $\alpha, \beta, \gamma, \dots$ the $(l, \lambda)^*$-coordinates and their indices, and $\{c_t, t = 0, 1, \dots\}$ and $t, u, v, \dots$ the general coordinates and their indices. 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. Thus $a_j\theta^j$ automatically means $\sum_{j=0}^{\infty} a_j\theta^j$.

2. Some preparative results

Suppose

$$S_1(\omega) = \frac{\sigma_1^2}{2\pi} f(\omega) \tag{2.1}$$

and

$$S_2(\omega) = \frac{\sigma_2^2}{2\pi} g(\omega) \tag{2.2}$$

are two systems in L, where

$$f(\omega) = \left| \sum_{t=0}^{\infty} \xi_t \, e^{-it\omega} \right|^2 \tag{2.3}$$

and

$$g(\omega) = \left| \sum_{t=0}^{\infty} \zeta_t \, e^{-it\omega} \right|^2. \tag{2.4}$$

In this section, we are going to give a spectral expression of $\sigma_{12}^{(l)2}$, the l-step forecasting error variance when S_2 is used for forecasting of a series generated by S_1, get some useful properties of spectral density functions and intuitively derive an appropriate quantity which will play the same role in our geometric structure related to (1.6) as D_{-1} in the structure related to one-step forecasting error variance ratio.

The expression in the following lemma can be proved similarly as the one for $l = 1$ given in Grenander and Rosenblatt ((1957), p. 261, (2)).

LEMMA 2.1. *The l-step forecasting error variance by using S_2 for prediction of a series generated by S_1, where S_1 and S_2 are given in* (2.1)–(2.4), *is*

$$\sigma_{12}^{(l)2} = \frac{\sigma_1^2}{\pi} \int_0^{\pi} \frac{|\sum_{s=0}^{l-1} \zeta_s e^{-is\omega}|^2 f(\omega)}{g(\omega)} \, d\omega. \tag{2.5}$$

If we call $h(\omega)$ in (1.3) the structural factor of $S(\omega)$, Lemma 2.2 says that the integral of the logarithm of structural factor of any S in L is 0.

LEMMA 2.2. *For any S in L,*

$$\int_0^\pi \log h(\omega)\,d\omega = 0, \tag{2.6}$$

where $h(\omega)$ is as in (1.2) *and* (1.3).

PROOF. Since we have

$$\sigma_\epsilon^2 = 2\pi \exp\left\{\frac{1}{\pi}\int_0^\pi \log S(\omega)\,d\omega\right\} = \sigma_\epsilon^2 \exp\left\{\frac{1}{\pi}\int_0^\pi \log h(\omega)\,d\omega\right\} \tag{2.7}$$

from Kolmogorov (1941) (see also Grenander and Rosenblatt ((1957), p. 69, etc.)), (2.5) follows.

By Lemma 2.1, (1.6) equals

$$\frac{1}{\pi}\int_0^\pi \left[\lambda + (1-\lambda)\left|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}\right|^2\right]\frac{f(\omega)}{g(\omega)}\,d\omega \Big/ \left[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2\right],$$

or

$$\begin{aligned}\log\frac{\lambda\sigma_{12}^{(1)2} + (1-\lambda)\sigma_{12}^{(l)2}}{\lambda\sigma_{11}^{(1)2} + (1-\lambda)\sigma_{11}^{(l)2}} &= \log\left\{\frac{1}{\pi}\int_0^\pi \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\zeta_s^2}\frac{f(\omega)}{g(\omega)}\,d\omega\right\} \\ &\quad - \log\frac{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\zeta_s^2}.\end{aligned} \tag{2.8}$$

Let us imitate D_{-1} to consider a quantity of the form

$$\frac{1}{\pi}\int_0^\pi \left\{K_1\frac{S_1}{S_2} - 1 - \log\left(K_2\frac{S_1}{S_2}\right)\right\}d\omega, \tag{2.9}$$

where K_1 and K_2 do not depend on σ_1^2 and σ_2^2. Since

$$\int_0^\pi \log\frac{f(\omega)}{g(\omega)}\,d\omega = 0,$$

by Lemma 2.2, the minimum of (2.9) over σ_1^2/σ_2^2 is

$$\log\left\{\frac{1}{\pi}\int_0^\pi K_1\frac{f(\omega)}{g(\omega)}\,d\omega\right\} - \log K_2. \tag{2.10}$$

Comparing (2.10) with (2.8), we see that in order to make the minimum of (2.9) equal the logarithm of (1.6), we should choose (We have some other choices of K_1

and K_2 which also satisfy this requirement. But there are some other restrictions, and we will see later that our candidate $D_{l,\lambda}$ in (2.11) is a right one.)

$$K_1 = \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1} \zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \zeta_s^2}$$

and

$$K_2 = \frac{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \xi_s^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \zeta_s^2}.$$

That is, we should employ

$$\text{(2.11)} \quad D_{l,\lambda}(S_1, S_2) = \frac{1}{\pi}\int_0^{\pi} \left\{ \left[\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1} \zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \zeta_s^2} \right] \frac{S_1(\omega)}{S_2(\omega)} - 1 - \log\frac{S_1(\omega)}{S_2(\omega)} - \log\frac{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \xi_s^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1} \zeta_s^2} \right\} d\omega,$$

which is called the D-function of S_1 and S_2 for the moment. When $l = 1$ or $\lambda = 1$, $D_{l,\lambda}(S_1, S_2)$ reduces to $D_{-1}(S_1, S_2)$. Therefore, $D_{l,\lambda}$ is a generalization of D_{-1}. Its minimization result, which is formally expressed in the next lemma, is a generalization of Lemma 1.1 in Xu (1988).

LEMMA 2.3. *Let $D_{l,\lambda}(S_1, S_2)$ be the D-function of S_1 and S_2. Then*

$$\text{(2.12)} \qquad \inf_{\sigma_2^2} D_{l,\lambda}(S_1, S_2) = \inf_{\sigma_1^2/\sigma_2^2} D_{l,\lambda}(S_1, S_2) = \log\frac{\lambda\sigma_{12}^{(1)2} + (1-\lambda)\sigma_{12}^{(l)2}}{\lambda\sigma_{11}^{(1)2} + (1-\lambda)\sigma_{11}^{(l)2}}.$$

Lemma 2.4 will play an important role in our establishing the differential geometrical structures.

LEMMA 2.4. *For any S in L, expressed as in* (1.2) *and* (1.3), *we have*

$$\text{(2.13)} \qquad \partial_t \sum_{s=0}^{l-1} \kappa_s^2 = \frac{1}{\pi}\int_0^{\pi} \left| \sum_{s=0}^{l-1} \kappa_s e^{-is\omega} \right|^2 \partial_t \log h(\omega)\, d\omega,$$

where $\partial_t = \partial/\partial c_t$ and $\{c_t\}$ is any coordinate system of S.

PROOF. From (2.12), $\sigma_{12}^{(1)2} \geq \sigma_{11}^{(1)2}$ and $\sigma_{12}^{(l)2} \geq \sigma_{11}^{(l)2}$, it is trivial that

$$D_{l,\lambda}(S_1, S_2) \geq 0$$

for any S_1 and S_2. But $D_{l,\lambda}(S_1, S_2) = 0$ when $S_1 = S_2$. So we must have

$$\text{(2.14)} \qquad \partial_t D_{l,\lambda}(S, S_2)|_{S_2 = S} = 0$$

for any S. Now

$$\text{(2.15)}\qquad \partial_t D_{l,\lambda}(S,S_2) = \frac{1}{\pi}\int_0^\pi \left\{ \left[\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\zeta_s^2}\right]\frac{\partial_t S(\omega)}{S_2(\omega)} - \frac{\partial_t S(\omega)}{S(\omega)} - \frac{(1-\lambda)\partial_t \sum_{s=0}^{l-1}\kappa_s^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2}\right\} d\omega,$$

hence (2.14) is equivalent to

$$\frac{1}{\pi}\int_0^\pi \left[\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2 - \sum_{s=0}^{l-1}\kappa_s^2\right]\partial_t \log S(\omega)\, d\omega = \partial_t \sum_{s=0}^{l-1}\kappa_s^2.$$

The left hand side of the above equality is equal to

$$\frac{1}{\pi}\int_0^\pi \left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2 \partial_t \log h(\omega)\, d\omega$$

because of

$$\text{(2.16)}\qquad \frac{1}{\pi}\int_0^\pi \left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2 d\omega = \sum_{s=0}^{l-1}\kappa_s^2$$

and Lemma 2.2. Thus (2.13) is shown.

3. Our (l,λ)-metric

One of the important properties of D_{-1} is that its second derivatives with respect to certain coordinate yield a Riemannian metric tensor. We will elucidate in this section that our D-function also possesses this property, and hence our (l,λ)-metric can be derived. Let us review Amari's -1-coordinate system and verify a lemma first.

Amari's -1-coordinates $c_t^{(-1)}$, $t = 0, 1, \ldots$ are

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

where

$$\text{(3.1)}\qquad e_0(\omega) = 1, \qquad e_t(\omega) = \sqrt{2}\cos\omega t, \qquad t \geq 1.$$

The -1-coordinates relate to the autocovariances of X_t:

$$c_0^{(-1)} = E(X_u^2), \qquad c_t^{(-1)} = \sqrt{2}E(X_u X_{t+u}), \qquad t \geq 1.$$

Therefore, for any S in L we have

$$\text{(3.2)}\qquad \partial_j\partial_k S = 0,$$

where $\partial_j = \partial/\partial\theta^j$ and $\{\theta^j \equiv c_j^{(-1)}\}$ are the -1-coordinates of S. This equality, (2.7) and Lemma 2.4 will be used in proving the following lemma.

LEMMA 3.1. *Suppose that S, ∂_t and $\{c_t\}$ are as in Lemma 2.4. Then*

$$\begin{aligned}(3.3)\quad \int_0^{\pi}\Bigg\{2\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_t \log h(\omega)\, \partial_u \log h(\omega) \\ - \partial_t\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_u \log h(\omega) - \partial_u\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_t \log h(\omega)\Bigg\}\, d\omega\end{aligned}$$

is nonnegative definite, and

$$\begin{aligned}(3.4)\quad \int_0^{\pi} \partial_t\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_u \log h(\omega)\, d\omega \\ = \int_0^{\pi} \partial_u\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_t \log h(\omega)\, d\omega.\end{aligned}$$

PROOF. Consider

$$G_l(S_1, S_2) = \frac{1}{\pi}\int_0^{\pi}\left\{\frac{|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\sum_{s=0}^{l-1}\xi_s^2}\frac{S_1(\omega)}{S_2(\omega)} - 1 - \log\frac{S_1(\omega)}{S_2(\omega)}\right\} d\omega.$$

By the same way as dealing with $D_{l,\lambda}(S_1, S_2)$, we can show that $G_l(S_1, S_2) = 0$ when $S_1 = S_2$,

$$\inf_{\sigma_2^2} G_l(S_1, S_2) = \inf_{\sigma_1^2/\sigma_2^2} G_l(S_1, S_2) = \log r_{12}^{(l)},$$

where $r_{12}^{(l)} = \sigma_{12}^{(l)2}/\sigma_{12}^{(1)2}$ is the l-step forecasting error variance ratio of S_2 versus S_1, and hence $G_l(S_1, S_2)$ is a contrast function (Eguchi (1983)). So

$$(3.5)\quad -\frac{\partial}{\partial c_t}\frac{\partial}{\partial c_u'} G_l(S, S')\Bigg|_{S'=S}$$

is symmetric and nonnegative definite, where $\{c_t'\}$ are the coordinates of S'. The equality (3.4) is an easy consequence of the symmetry of (3.5). Because of (3.4), (2.16) and (2.13), it is not difficult to see that (3.5) is positively proportional to

$$\begin{aligned}(3.6)\quad &\partial_t \log\sigma_\epsilon^2\, \partial_u \log\sigma_\epsilon^2 \\ &+ \frac{1}{\pi}\int_0^{\pi}\Bigg\{\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_t \log h(\omega)\, \partial_u \log h(\omega) \\ &- \frac{1}{2}\partial_t\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_u \log h(\omega) \\ &- \frac{1}{2}\partial_u\Bigg|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\Bigg|^2 \partial_t \log h(\omega)\Bigg\}\, d\omega\Bigg/\sum_{s=0}^{l-1}\kappa_s^2.\end{aligned}$$

Using $c_0 = \sigma_\epsilon^2$ and $c_t = \kappa_t$ for $t \geq 1$ in (3.6), we see that (3.3) should be nonnegative definite for this coordinate system, which implies that it is nonnegative definite for any coordinate system because that it is a tensor.

The following theorem, which will be proved based on Lemma 3.1, makes it possible for us to introduce our (l, λ)-metric.

THEOREM 3.1. *Suppose that S, ∂_t and $\{c_t\}$ are as in Lemma 2.4, $\{\theta^j\}$ are the -1-coordinates of S and $D_{l,\lambda}$ is the D-function. Then*

$$-\frac{1}{\pi}\int_0^\pi \partial_t\left\{\frac{\lambda+(1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{\left[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]S(\omega)}\right\}\partial_u S(\omega)\,d\omega \tag{3.7}$$

is positive definite, and

$$\partial_j\partial_k D_{l,\lambda}(S,S_2) = -\frac{1}{\pi}\int_0^\pi \partial_j\left\{\frac{\lambda+(1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{\left[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]S(\omega)}\right\}\partial_k S(\omega)\,d\omega \tag{3.8}$$

for any S_2 in L.

PROOF. By (2.6), (2.13), (2.16) and (3.4), (3.7) is the same as

$$\begin{aligned}
&\frac{1}{\pi[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2]}\int_0^\pi\left\{\left[\lambda+(1-\lambda)\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2\right]\right.\\
&\qquad \partial_t\log S(\omega)\ \partial_u\log S(\omega) - \frac{1-\lambda}{2}\partial_t\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2\partial_u\log h(\omega)\\
&\qquad \left.-\frac{1-\lambda}{2}\partial_u\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2\partial_t\log h(\omega)\right\}d\omega\\
&\qquad +\partial_t\log\left[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]\partial_u\log\left[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right].
\end{aligned} \tag{3.9}$$

Using Lemmas 2.2 and 2.4, we see that (3.9) is positively proportional to

$$2Q_1 + \frac{1-\lambda}{\pi}Q_2,$$

where Q_2 is as in (3.3), which is nonnegative definite, and

$$
\begin{aligned}
Q_1 =& \frac{\lambda}{\pi}\int_0^{\pi} \partial_t \log h(\omega)\, \partial_u \log h(\omega)\, d\omega + \Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big]\partial_t \log \sigma_\epsilon^2\, \partial_u \log \sigma_\epsilon^2 \\
&+ (1-\lambda)\Big[\partial_t \log \sigma_\epsilon^2\, \partial_u \sum_{s=0}^{l-1}\kappa_s^2 + \partial_u \log \sigma_\epsilon^2\, \partial_t \sum_{s=0}^{l-1}\kappa_s^2\Big] \\
&+ (1-\lambda)^2 \partial_t \sum_{s=0}^{l-1}\kappa_s^2\, \partial_u \sum_{s=0}^{l-1}\kappa_s^2 \Big/ \Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big] \\
=& \frac{\lambda}{\pi}\int_0^{\pi} \partial_t \log h(\omega)\, \partial_u \log h(\omega)\, d\omega + \Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big] \\
&\partial_t \log\Big\{\sigma_\epsilon^2\Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big]\Big\}\, \partial_u \log\Big\{\sigma_\epsilon^2\Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big]\Big\},
\end{aligned}
$$

which is trivially positive definite. Therefore, (3.7) is positive definite.

From (2.11) and (3.2),

$$
\partial_j \partial_k D_{l,\lambda}(S, S_2) = -\frac{1}{\pi}\partial_j \partial_k \int_0^{\pi} \log\Big\{\Big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\Big] S(\omega)\Big\}\, d\omega. \tag{3.10}
$$

The right hand side of (3.10) is the same as that of (3.8) by (3.2) and the fact that $D_{l,\lambda}$ is a contrast function.

Now let us formally define our metric tensor.

DEFINITION 3.1. The (l,λ)-metric tensor $\{g_{tu}(l,\lambda), t, u = 0, 1, \ldots\}$ at S as in (1.2), is given by (The equality (3.11) can be rewritten as $g_{tu} = -(\partial/\partial c_t)\cdot(\partial/\partial c'_u) D_{l,\lambda}(S, S')\big|_{S'=S}$, where $\{c'_t\}$ are the coordinates of S'. Therefore, this is the same as the D-metric tensor defined in Eguchi (1983). But we need to point out here that we can not ensure that the ρ-metric tensor for any contrast function ρ is positive definite. The counterexamples are trivial. So we still need Theorem 3.1 in order to equip g as a Riemannian metric.)

$$
g_{tu}(l,\lambda) = -\frac{1}{\pi}\int_0^{\pi} \partial_t \Big\{\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{\big[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\big] S(\omega)}\Big\} \partial_u S(\omega)\, d\omega. \tag{3.11}
$$

4. The (l,λ)-coordinates, the (l,λ)-connections and their duals

In this section we will introduce the (l,λ)- and $(l,\lambda)^*$-coordinates and the (l,λ)- and $(l,\lambda)^*$-connections, show their duality and verify the divergence.

DEFINITION 4.1. For any $S \in L$, its -1-coordinates will also be called its (l, λ)-coordinates, denoted as $\{\theta^j, j = 0, 1, \dots\}$. That is,

$$\theta^j = 2\int_0^\pi S(\omega)e_j(\omega)\,d\omega, \tag{4.1}$$

where $e_j(\omega)$ is as in (3.1). The $(l, \lambda)^*$-coordinates of S, denoted as $\{\eta_\alpha, \alpha = 0, 1, \dots\}$, are defined by

$$\eta_\alpha = -\frac{1}{2\pi^2}\int_0^\pi \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2]S(\omega)}e_\alpha(\omega)\,d\omega. \tag{4.2}$$

It is trivial that $\sum_{j=0}^\infty \theta^j\theta^j < \infty$ and $\sum_{\alpha=0}^\infty \eta_\alpha\eta_\alpha < \infty$ from the Parseval's identity and the definition of the manifold L. We also need to show that S is uniquely determined by $\{\eta_\alpha\}$. Suppose that S_1 and S_2, as given in (2.1)–(2.4), have the same $\{\eta_\alpha\}$. Then

$$\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\xi_s e^{-is\omega}|^2}{\sigma_1^2[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2]f(\omega)} = \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\sigma_2^2[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\zeta_s^2]g(\omega)},$$

hence

$$\begin{aligned} 1 &= \frac{1}{\pi}\int_0^\pi \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\xi_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2}\,d\omega \\ &= \frac{\sigma_1^2}{\pi\sigma_2^2}\int_0^\pi \frac{[\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2]f(\omega)}{[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\zeta_s^2]g(\omega)}\,d\omega, \end{aligned}$$

or

$$\frac{\lambda\sigma_{12}^{(1)2} + (1-\lambda)\sigma_{12}^{(l)2}}{\lambda\sigma_{22}^{(1)2} + (1-\lambda)\sigma_{22}^{(l)2}} = 1 \tag{4.3}$$

by (2.5). Similarly,

$$\frac{\lambda\sigma_{21}^{(1)2} + (1-\lambda)\sigma_{21}^{(l)2}}{\lambda\sigma_{11}^{(1)2} + (1-\lambda)\sigma_{11}^{(l)2}} = 1. \tag{4.4}$$

From (4.3) and (4.4), we must have

$$\lambda\sigma_{12}^{(1)2} + (1-\lambda)\sigma_{12}^{(l)2} = \lambda\sigma_{11}^{(1)2} + (1-\lambda)\sigma_{11}^{(l)2} \tag{4.5}$$

or

$$\lambda\sigma_{21}^{(1)2} + (1-\lambda)\sigma_{21}^{(l)2} = \lambda\sigma_{22}^{(1)2} + (1-\lambda)\sigma_{22}^{(l)2}, \tag{4.6}$$

because if (4.5) and (4.6) are both false, the left hand sides in (4.5) and (4.6) will be both greater than their corresponding right hand sides, and either (4.3) or (4.4) can not be true. Now $S_1 = S_2$ can be derived from either (4.5) or (4.6). Thus $\{\eta_\alpha\}$ satisfies the conditions for a coordinate system.

The (l,λ)-potential functions are defined in the next definition. They are related to the "(l,λ)-entropy" H.

DEFINITION 4.2. For any $S \in L$, as given in (1.2), the (l,λ)-potential functions $\psi_{l,\lambda}$ and $\phi_{l,\lambda}$ are given by

$$\psi_{l,\lambda}(S) = -H(S) - \frac{3}{2} - \log 2\pi \tag{4.7}$$

and

$$\phi_{l,\lambda}(S) = H(S) + \frac{1}{2} + \log 2\pi\,, \tag{4.8}$$

where

$$H(S) = \frac{1}{\pi}\int_0^\pi \log\left\{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]S(\omega)\right\}d\omega. \tag{4.9}$$

Theorem 4.1 shows that $\{\theta^j\}$ and $\{\eta_\alpha\}$ are mutually dual, and $\psi_{l,\lambda}$ and $\phi_{l,\lambda}$ are truly the potential functions.

THEOREM 4.1. *Let $\psi_{l,\lambda}(S)$, $\phi_{l,\lambda}(S)$, $\{\theta^j\}$, $\{\eta_\alpha\}$ and $g_{jk}(l,\lambda)$ be the (l,λ)-potential functions at $S \in L$, the (l,λ)- and $(l,\lambda)^*$-coordinates of S and the (l,λ)-metric at S in the (l,λ)-coordinates, respectively. Then*

$$\psi_{l,\lambda}(S) + \phi_{l,\lambda}(S) - \theta^t\eta_t = 0, \tag{4.10}$$

$$\eta_j = \partial_j\psi_{l,\lambda}(S), \tag{4.11}$$

$$\theta^\alpha = \partial^\alpha\phi_{l,\lambda}(S), \tag{4.12}$$

$$\partial_j\partial_k\psi_{l,\lambda}(S) = g_{jk}(l,\lambda) = \frac{\partial\eta_k}{\partial\theta^j}, \tag{4.13}$$

$$\partial^\alpha\partial^\beta\phi_{l,\lambda}(S) = g^{\alpha\beta}(l,\lambda) = \frac{\partial\theta^\beta}{\partial\eta_\alpha} \tag{4.14}$$

and

$$\langle\partial_j, \partial^\alpha\rangle = \delta_j^\alpha, \tag{4.15}$$

where $\partial_j = \partial/\partial\theta^j$, $\partial^\alpha = \partial/\partial\eta_\alpha$, $\{g^{\alpha\beta}(l,\lambda)\}$ is the inverse of $\{g_{jk}(l,\lambda)\}$, $\langle\,,\,\rangle$ is the inner product induced by the Riemannian metric $g_{tu}(l,\lambda)$ and δ_j^α is the Kronecker delta.

PROOF. From (4.1), (4.2) and the fact that $\theta^t\eta_t = \sum_{t=0}^\infty \theta^t\eta_t$ is convergent,

$$\theta^t\eta_t = -\frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{[\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega'}|^2]S(\omega)}{[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2]S(\omega')}\sum_{t=0}^\infty e_t(\omega)e_t(\omega')\,d\omega'\,d\omega.$$

Since

$$\sum_{t=0}^{\infty} e_t(\omega)e_t(\omega') = \pi\delta(\omega - \omega'), \tag{4.16}$$

where $\delta(\omega)$ is the generalized function δ, we have

$$\theta^t \eta_t = -\frac{1}{\pi}\int_0^{\pi} \frac{\left[\lambda + (1-\lambda)\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2\right]S(\omega)}{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]S(\omega)}\,d\omega = -1. \tag{4.17}$$

But

$$\psi_{l,\lambda}(S) + \phi_{l,\lambda}(S) = -1 \tag{4.18}$$

from (4.7) and (4.8). Thus (4.10) comes out of (4.17) and (4.18).

Since $\{\theta^j\}$ are the -1-coordinates of S, we have

$$\partial_j S(\omega) = e_j(\omega)/2\pi. \tag{4.19}$$

By (4.7), (4.9) and (4.19),

$$\partial_j \psi_{l,\lambda}(S) = -\frac{(1-\lambda)\partial_j \sum_{s=0}^{l-1}\kappa_s^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2} - \frac{1}{2\pi^2}\int_0^{\pi} \frac{e_j(\omega)}{S(\omega)}\,d\omega. \tag{4.20}$$

But by (2.13), (2.16), (2.7) and (4.19),

$$\begin{aligned} \partial_j \sum_{s=0}^{l-1}\kappa_s^2 &= \frac{1}{\pi}\int_0^{\pi}\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2 \left\{\partial_j \log S(\omega) - \partial_j \log\sigma_\epsilon^2\right\} d\omega \\ &= \frac{1}{2\pi^2}\int_0^{\pi}\left\{\left|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}\right|^2 - \sum_{s=0}^{l-1}\kappa_s^2\right\}\frac{e_j(\omega)}{S(\omega)}\,d\omega. \end{aligned} \tag{4.21}$$

Substituting (4.21) into (4.20) and noticing (4.2), we get (4.11). From (4.10) and (4.11),

$$\begin{aligned} 0 &= \partial^\alpha \psi_{l,\lambda}(S) + \partial^\alpha \phi_{l,\lambda}(S) - \partial^\alpha(\theta^u \eta_u) \\ &= \frac{\partial\theta^u}{\partial\eta_\alpha}\eta_u + \partial^\alpha\phi_{l,\lambda}(S) - \theta^\alpha - \frac{\partial\theta^u}{\partial\eta_\alpha}\eta_u, \end{aligned}$$

which results in (4.12). The first equality in (4.13) is verified by (4.7), (4.9), (3.8), (3.10) and (3.11), and implies the second one by combining with (4.11). From (4.12),

$$\partial^\alpha\partial^\beta\phi_{l,\lambda}(S) = \frac{\partial\theta^\beta}{\partial\eta_\alpha}.$$

Since $\{\partial\theta^\beta/\partial\eta_\alpha\}$ is the inverse of $\{\partial\eta_k/\partial\theta^j\}$, (4.14) is derived. We can easily verify (4.15) by (4.14) and the fact that $\{g^{\alpha\beta}(l,\lambda)\}$ is the inverse of $\{g_{jk}(l,\lambda)\}$.

Now we can justify that our D-function is the (l,λ)-divergence because of the next theorem.

THEOREM 4.2. *Let $D_{l,\lambda}(S,S')$, $\{\theta^j\}$ and $\{\eta'_\alpha\}$ be the D-function of S and S', the (l,λ)-coordinates of S and the $(l,\lambda)^*$-coordinates of S', respectively. Then*

$$D_{l,\lambda}(S,S') = \psi_{l,\lambda}(S) + \phi_{l,\lambda}(S') - \theta^t \eta'_t. \tag{4.22}$$

PROOF. Let

$$S'(\omega) = \frac{\sigma_{\epsilon'}^2}{2\pi} h'(\omega) \quad \text{and} \quad h'(\omega) = \left|\sum_{u=0}^{\infty} \kappa'_u e^{-iu\omega}\right|^2$$

with $\kappa'_0 = 1$ and $\sum_{u=0}^{\infty} {\kappa'_s}^2 < \infty$. By the similar argument as (4.17), we have

$$\theta^t \eta'_t = -\frac{1}{\pi}\int_0^\pi \frac{\left[\lambda + (1-\lambda)\left|\sum_{s=0}^{l-1} \kappa'_s e^{-is\omega}\right|^2\right] S(\omega)}{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1} {\kappa'_s}^2\right] S'(\omega)}\, d\omega. \tag{4.23}$$

But from (4.7)–(4.9),

$$\begin{aligned}\psi_{l,\lambda}(S) + \phi_{l,\lambda}(S') = &-\frac{1}{\pi}\int_0^\pi \log\left\{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1} \kappa_s^2\right] S(\omega)\right\} d\omega \\ &+ \frac{1}{\pi}\int_0^\pi \log\left\{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1} {\kappa'_s}^2\right] S'(\omega)\right\} d\omega - 1,\end{aligned}$$

which, combined with (2.11) and (4.23), implies (4.22).

In order to get global properties, except for the Riemannian metric $g_{jk}(l,\lambda)$, we also need to define mutual relations between any two tangent spaces at two neighboring points. This can be accomplished by defining an affine correspondence between two tangent spaces, which is called an affine connection (Amari (1985)) in differential geometry. An affine connection is specified by defining $\nabla_{\partial_t}\partial_u$, the rate at which ∂_u "intrinsically" changes in the direction ∂_t or the covariant derivative of ∂_u in the direction ∂_v. Equivalently, it can be described by its components (Amari (1987*a*)) $\Gamma_{tuv} = \langle \nabla_{\partial_t}\partial_u, \partial_v\rangle$. The following definition gives the components of our (l,λ)- and $(l,\lambda)^*$-connections.

DEFINITION 4.3. The components of the (l,λ)- and $(l,\lambda)^*$-connections for the coordinate system $\{c_t\}$ are given by

$$\Gamma_{tuv} = -\frac{1}{\pi}\int_0^\pi \partial_t \partial_u S(\omega) \partial_v \left\{\frac{\lambda + (1-\lambda)\left|\sum_{s=0}^{l-1} \kappa_s e^{-is\omega}\right|^2}{\left[\lambda + (1-\lambda)\sum_{s=0}^{l-1} \kappa_s^2\right] S(\omega)}\right\} d\omega \tag{4.24}$$

and

$$\Gamma^*_{tuv} = -\frac{1}{\pi}\int_0^\pi \partial_t\partial_u\left\{\frac{\lambda+(1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{\left[\lambda+(1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2\right]S(\omega)}\right\}\partial_v S(\omega)\,d\omega. \tag{4.25}$$

The above definition is justified by the next theorem, where the facts that $\{\theta^j\}$ and $\{\eta_\alpha\}$ are correspondingly the affine coordinates with respect to the (l,λ)- and $(l,\lambda)^*$-connections and L is flat (i.e. the Riemann-Christoffel curvature vanishes identically) with any of these connections are also shown.

THEOREM 4.3. *Γ_{tuv} and Γ^*_{tuv} are torsion free and mutually dual connections. The manifold L is (l,λ)- and $(l,\lambda)^*$-flat. The (l,λ)-coordinates $\{\theta^j\}$ and the $(l,\lambda)^*$-coordinates $\{\eta_\alpha\}$ are the affine coordinates with respect to the (l,λ)- and $(l,\lambda)^*$-connections, correspondingly.*

PROOF. Since

$$\Gamma_{tuv} = -\left.\frac{\partial^3 D_{l,\lambda}(S,S')}{\partial c_t\partial c_u\partial c'_v}\right|_{S'=S} \quad \text{and} \quad \Gamma^*_{tuv} = -\left.\frac{\partial^3 D_{l,\lambda}(S,S')}{\partial c'_t\partial c'_u\partial c_v}\right|_{S'=S},$$

where $\{c_t\}$ and $\{c'_t\}$ are the coordinates of S and S', respectively, the first claim is concluded in Eguchi (1983) and Amari ((1985), p. 98). The second and third claims can be verified by $\Gamma_{jkm}=0$ and $\Gamma^*_{\alpha\beta\gamma}=0$.

Now we have completed the establishing of the differential geometric structures related to (1.6).

5. Approximation by a model

Given a model M and a system $S \in L$, we want to get $\hat{S} \in M$ which is the nearest to S in some divergence measure among all the systems in M.

DEFINITION 5.1. If $\hat{S}$ satisfies $D_{l,\lambda}(S,\hat{S}) = \min_{S'\in M} D_{l,\lambda}(S,S')$ or $D_{l,\lambda}(\hat{S},S) = \min_{S'\in M} D_{l,\lambda}(S',S)$, then $\hat{S}$ is correspondingly called the (l,λ)- or the $(l,\lambda)^*$-approximation of S by M.

By (2.12) and Definition 5.1, it is clear that $D_{l,\lambda}(S,\hat{S})$, where $\hat{S}\equiv\hat{S}(l,\lambda)$ is the (l,λ)-approximation of S by M, is the minimum of the logarithm of (1.6) over M (here we take S as S_1 and any system in M as S_2). Therefore, if $\hat{S}(l,\lambda)$ tends to some $\hat{S}(l)$ in M as $\lambda\to 0$, $\hat{S}(l)$ will have the minimum l-step forecasting error variance ratio, i.e. $\hat{S}(l)$ is a best approximation of S by M for the purpose of l-step prediction.

We are going to use M_q to denote the submanifold consisting of all MA (q) systems in L. This submanifold is characterized by

$$\theta^j \equiv c_j^{(-1)} = 0 \tag{5.1}$$

for any $j > q$. Let $A_p^{(l,\lambda)}$ be the submanifold characterized by

$$\eta_\alpha = 0 \tag{5.2}$$

for any $\alpha > p + l - 1$, where $\{\eta_\alpha\}$ are the $(l,\lambda)^*$-coordinates. Then all AR (p) systems are included in $A_p^{(l,\lambda)}$. A submanifold is said to be completely flat if the directions of its tangent spaces remain fixed all over this submanifold (i.e. if the Euler-Schouten curvature vanishes identically). In the next theorem it is shown that both $A_p^{(l,\lambda)}$ and M_q are completely flat with respect to the (l,λ)- and $(l,\lambda)^*$-connections.

THEOREM 5.1. *The submanifolds M_q and $A_p^{(l,\lambda)}$ are both completely (l,λ)- and $(l,\lambda)^*$-flat in L.*

PROOF. This is trivial because of the results in Theorem 4.3 and the fact that both (5.1) and (5.2) are linear restrictions.

The explicit forms of the (l,λ)-approximation when our model is $A_p^{(l,\lambda)}$ and the $(l,\lambda)^*$-approximation when our model is MA (q) are given in Theorem 5.2.

THEOREM 5.2. *The (l,λ)-approximation $\hat{S}_p$ of S by $A_p^{(l,\lambda)}$ is unique. Its (l,λ)-coordinates $\{\hat{\theta}^j\}$ and $(l,\lambda)^*$-coordinates $\{\hat{\eta}_\alpha\}$ satisfy*

$$\hat{\theta}^j = \theta^j \tag{5.3}$$

for $j = 0, 1, \ldots, p + l - 1$ and $\hat{\eta}_\alpha = 0$ for $\alpha > p + l - 1$, where $\{\theta^j\}$ are the (l,λ)-coordinates of S. The approximation error evaluated by the (l,λ)-divergence of $\hat{S}_p$ from S is

$$D_{l,\lambda}(S, \hat{S}_p) = H(\hat{S}_p) - H(S). \tag{5.4}$$

The $(l,\lambda)^*$-approximation $\hat{S}_q^*$ of S by MA (q) is unique. Its (l,λ)-coordinates $\{\hat{\theta}^{*j}\}$ and $(l,\lambda)^*$-coordinates $\{\hat{\eta}_\alpha^*\}$ satisfy

$$\hat{\theta}^{*j} = 0 \tag{5.5}$$

for $j > q$ and

$$\hat{\eta}_\alpha^* = \eta_\alpha \tag{5.6}$$

for $\alpha = 0, 1, 2, \ldots, q$, where $\{\eta_\alpha\}$ are the $(l,\lambda)^*$-coordinates of S. The approximation error evaluated by the (l,λ)-divergence of S from $\hat{S}_q^*$ is

$$D_{l,\lambda}(\hat{S}_q^*) = H(S) - H(\hat{S}_q^*). \tag{5.7}$$

PROOF. Since

$$D_{l,\lambda}(S,\hat{S}_p) = \min_{S' \in A_p^{(l,\lambda)}} D_{l,\lambda}(S,S') \quad \text{and}$$

$$D_{l,\lambda}(S,S') = \psi_{l,\lambda}(S) + \phi_{l,\lambda}(S') - \sum_{t=0}^{p+l-1} \theta^t \eta'_t$$

by (4.22) and (5.2), where $\{\eta'_t\}$ are the $(l,\lambda)^*$-coordinates of S', (5.3) is derived from

$$\frac{\partial}{\partial \eta'_j} D_{l,\lambda}(S,S') \bigg|_{S'=\hat{S}_p} = 0$$

and (4.12). The uniqueness comes from the previous theorem and Theorem 3.9 in Amari (1985), and (5.4) is derived by (4.22), (4.7), (4.8) and $\theta^t \hat{\eta}_t = \hat{\theta}^t \hat{\eta}_t = -1$. The equations (5.5)–(5.7) can be shown similarly.

For a sequence of $A_p^{(l,\lambda)}$ models or MA models, the following theorem shows that the approximation errors are decomposed additively corresponding to each dimension of the model. This result can be proved in the same way as for Theorem 10 in Amari (1987*b*).

THEOREM 5.3. *Let $\{\hat{S}_p, p = 0,1,\dots\}$ and $\{\hat{S}^*_q, q = 0,1,\dots\}$ be the sequences of the (l,λ)-approximations of S by $A_p^{(l,\lambda)}$ and the $(l,\lambda)^*$-approximations of S by M_q, respectively. Then, $\hat{S}_n$ is also the (l,λ)-approximation of $\hat{S}_k$ by $A_n^{(l,\lambda)}$ and $\hat{S}^*_n$ is also the $(l,\lambda)^*$-approximation of $\hat{S}^*_k$ by M_n, when $k > n$. The approximation errors satisfy the additive relation*

$$D_{l,\lambda}(S,\hat{S}_n) = D_{l,\lambda}(S,\hat{S}_k) + D_{l,\lambda}(\hat{S}_k,\hat{S}_n) \quad \textit{and}$$
$$D_{l,\lambda}(\hat{S}^*_n,S) = D_{l,\lambda}(\hat{S}^*_k,S) + D_{l,\lambda}(\hat{S}^*_n,\hat{S}^*_k).$$

In the rest of this section we are going to discuss the (l,λ)-approximation by AR (1) of any $S \in L$ for $l = 1,2$.

Suppose that the autocovariance of S at lag t is v_t. Our approximate AR (1) model is

$$(1 - \delta B) X_t = a_t, \tag{5.8}$$

where B is the backward shift operator, $BX_t = X_{t-1}$. Its $(1,\lambda)^*$-coordinates are

$$\eta'_0 = -\frac{1+\delta^2}{\sigma_a^2}, \qquad \eta'_1 = \frac{\sqrt{2}\delta}{\sigma_a^2}$$

and $\eta'_\alpha = 0$ for $\alpha > 1$, where σ_a^2 is the noise variance of a_t, and its $(1,\lambda)$-entropy is $H = \log(\sigma_a^2/2\pi)$. Therefore, in order to get the $(1,\lambda)$-approximation of S by (5.8), we need to minimize

$$\log \sigma_a^2 + \frac{(1+\delta^2)v_0}{\sigma_a^2} - \frac{2\delta v_1}{\sigma_a^2} \tag{5.9}$$

according to (4.22) and (4.8). It is trivial that the minimizers of (5.9) turn out to be

$$\delta = \frac{v_1}{v_0} \tag{5.10}$$

and

$$\sigma_a^2 = \frac{v_0^2 - v_1^2}{v_0}, \tag{5.11}$$

and the approximation error measured by the $(1, \lambda)$-divergence from S of (5.8) when (5.10) and (5.11) are taken is

$$\log \frac{v_0^2 - v_1^2}{\sigma_\epsilon^2 v_0},$$

where σ_ϵ^2 is the noise variance of S. In other words, the smallest one-step forecasting error variance ratio of AR (1) system vs. S is $(v_0^2 - v_1^2)/(\sigma_\epsilon^2 v_0)$. No wonder that λ does not play any role here.

The $(2, \lambda)^*$-coordinates and the $(2, \lambda)$-entropy of (5.8) are

$$\begin{aligned}
\eta_0' &= -\frac{1 + \delta^2[\lambda + (1-\lambda)\delta^2]}{\sigma_a^2[1 + (1-\lambda)\delta^2]}, \\
\eta_1' &= \frac{\sqrt{2}\lambda\delta}{\sigma_a^2[1 + (1-\lambda)\delta^2]}, \\
\eta_2' &= \frac{\sqrt{2}(1-\lambda)\delta^2}{\sigma_a^2[1 + (1-\lambda)\delta^2]}
\end{aligned}$$

and $\eta_\alpha' = 0$ for $\alpha > 2$, and $H = \log\{\sigma_a^2[1 + (1-\lambda)\delta^2]/2\pi\}$, respectively. Since the $(2, \lambda)$-coordinates of S are still its -1-coordinates: $\theta^0 = v_0$, $\theta^j = \sqrt{2}\, v_j$ for $j \geq 1$, the parameters of the $(2, \lambda)$-approximation of S by (5.8) should minimize

$$\begin{aligned}
&\log\{\sigma_a^2[1 + (1-\lambda)\delta^2]\} - \eta_0' v_0 - \sqrt{2}\eta_1' v_1 - \sqrt{2}\eta_2' v_2 \\
&\quad = \log\{\sigma_a^2[1 + (1-\lambda)\delta^2]\} \\
&\quad\quad + \frac{1}{\sigma_a^2[1 + (1-\lambda)\delta^2]}\{v_0 + v_0[\lambda + (1-\lambda)\delta^2]\delta^2 \\
&\quad\quad - 2\lambda v_1 \delta - 2(1-\lambda) v_2 \delta^2\}.
\end{aligned} \tag{5.12}$$

Hence, they satisfy

$$\sigma_a^2[1 + (1-\lambda)\delta^2] = v_0 + v_0[\lambda + (1-\lambda)\delta^2]\delta^2 - 2\lambda v_1 \delta - 2(1-\lambda) v_2 \delta^2, \tag{5.13}$$

where the right hand side is positive when λ is small enough, and δ should be the minimizer of

$$\begin{aligned}
&v_0[\lambda + (1-\lambda)\delta^2]\delta^2 - 2\lambda v_1 \delta - 2(1-\lambda) v_2 \delta^2 \\
&\quad = v_0(1-\lambda)\delta^4 + [\lambda v_0 - 2(1-\lambda) v_2]\delta^2 - 2\lambda v_1 \delta.
\end{aligned} \tag{5.14}$$

It is not difficult to show that as $\lambda \to 0$ this minimizer tends to 0 when $v_2 \le 0$, to $\sqrt{v_2/v_0}$ when $v_2 > 0$ and $v_1 > 0$, and to $-\sqrt{v_2/v_0}$ when $v_2 > 0$ and $v_1 < 0$, and σ_a^2 in (5.13) tends to $v_0 - 2v_2\delta^2 + v_0\delta^4$. Here we see that the limiting position of $(2,\lambda)$-approximation as $\lambda \to 0$ exists, and this limiting system S', which is $X_t = a_t$ when $v_2 \le 0$, $(1 - \sqrt{v_2/v_0}B)X_t = a_t$ when $v_2 > 0, v_1 > 0$, and $(1 + \sqrt{v_2/v_0}B)X_t = a_t$ when $v_2 > 0, v_1 < 0$, has the smallest two-step forecasting error variance ratio vs. S:

$$\exp\left\{\lim_{\lambda\to 0} D_{2,\lambda}(S,S')\right\} = \begin{cases} \dfrac{v_0}{\sigma_\epsilon^2(1+\kappa_1^2)} & \text{if } v_2 \le 0; \\ \dfrac{v_0^2 - v_2^2}{v_0\sigma_\epsilon^2(1+\kappa_1^2)} & \text{if } v_2 > 0, \end{cases}$$

where κ_1^2 is as given in (1.3).

6. Nonstationary cases

Until now we suppose that our systems are stationary. Occasionally we need to consider nonstationary cases. Suppose, for example, our two systems are

$$(1-\rho_1 B)Y_t = X_{1t} \tag{6.1}$$

and

$$(1-\rho_2 B)Y_t = X_{2t} \tag{6.2}$$

respectively, where $-1 < \rho_1 \le 1$, $-1 < \rho_2 \le 1$, X_{1t} and X_{2t} are stationary invertible zero-mean Gaussian time series having $S_1(\omega)$ and $S_2(\omega)$ in L, given in (2.1)–(2.4), as their spectral densities, correspondingly, and Y_t, X_{1t} and X_{2t} in (6.1) and (6.2) are all 0 when $t \le 0$. Then, we are dealing with the systems which might have a nonstationary unit root $+1$. Now $U_t = (1-B)Y_t$ is stationary (in fact it is asymptotically stationary, but its asymptotic behavior is the same as that of the corresponding stationary time series, so we just take it as stationary, especially we use the spectral density function of the corresponding stationary time series as the spectral density function of this asymptotically stationary time series), no matter Y_t is generated by (6.1) or (6.2) and no matter what the ρ_1 value or the ρ_2 value is. It satisfies

$$U_t = \frac{1-B}{1-\rho_1 B}X_{1t} \tag{6.3}$$

or

$$U_t = \frac{1-B}{1-\rho_2 B}X_{2t}, \tag{6.4}$$

and its spectral density function is, correspondingly,

$$\frac{|1-e^{-i\omega}|^2}{|1-\rho_1 e^{-i\omega}|^2}S_1(\omega) \quad \text{or} \quad \frac{|1-e^{-i\omega}|^2}{|1-\rho_2 e^{-i\omega}|^2}S_2(\omega).$$

Let $\hat{Y}_{t+l|t}$ denote the l-step ahead predictor when we use the system (6.2) while $\{Y_t\}$ is in fact generated by (6.1) and $Y_1, Y_2, \ldots, Y_t$ have been observed, and

$\hat{U}_{t+l|t}$ the l-step ahead predictor when we use (6.4) while $\{U_t\}$ satisfies (6.3) and $U_1, U_2, \ldots, U_t$, or equally $Y_1, Y_2, \ldots, Y_t$ have been observed. Then

$$\hat{Y}_{t+l|t} - Y_{t+l} = \sum_{u=t+1}^{t+l} (\hat{U}_{u|t} - U_u). \tag{6.5}$$

Similar to Grenander and Rosenblatt ((1957), p. 261, (2)), we have

$$\begin{aligned}(6.6)\quad &\hat{U}_{u|t} - U_u \\ &\asymp - \int_{-\pi}^{\pi} \frac{e^{i(u-t)\omega} \sum_{s=0}^{u-t-1} e^{-is\omega} \left[\sum_{v=0}^{s} \zeta_v \rho_2^{s-v} - \sum_{v=0}^{s-1} \zeta_v \rho_2^{s-1-v}\right]}{c(e^{-i\omega})} dz(\omega),\end{aligned}$$

where "$\asymp$" means "'asymptotically' represented by" (it exactly means that when $u-t$ is a fixed integer, the left hand side of (6.6) tends to the right hand side in mean square as $t \to \infty$),

$$|c(e^{-i\omega})|^2 = \frac{g(\omega)|1 - e^{-i\omega}|^2}{|1 - \rho_2 e^{-i\omega}|^2},$$

and $z(\omega)$ is the spectral process with orthogonal increments corresponding to $\{U_t\}$. It is not difficult to get

$$\hat{Y}_{t+l|t} - Y_{t+l} \asymp - \int_{-\pi}^{\pi} \frac{e^{il\omega} \sum_{s=0}^{l-1} e^{-is\omega} \sum_{u=0}^{s} \zeta_u \rho_2^{s-u}}{c(e^{-i\omega})} dz(\omega)$$

from (6.5) and (6.6). Thus

$$\begin{aligned}\sigma_{12}^{(l)2} &\equiv \lim_{t\to\infty} E\big[\hat{Y}_{t+l|t} - Y_{t+l}\big]^2 \\ &= \frac{\sigma_1^2}{\pi} \int_0^{\pi} \frac{|\sum_{s=0}^{l-1} e^{-is\omega} \sum_{u=0}^{s} \zeta_u \rho_2^{s-u}|^2 |1 - \rho_2 e^{-i\omega}|^2 f(\omega)}{|1 - \rho_1 e^{-i\omega}|^2 g(\omega)} d\omega,\end{aligned} \tag{6.7}$$

which plays the same role as (2.5), and the geometrical problems can be similarly discussed.

The other cases of nonstationarity can be treated in a similar way.

As an application, let us consider the following example. Suppose that $\{Y_t\}$ is generated by

$$(1 - \varrho B)\Big(1 - \sum_{u=1}^{p} \varphi_t B^u\Big) Y_t = \Big(1 - \sum_{u=1}^{q} \vartheta_u B^u\Big) a_t, \qquad t \geq 1 \tag{6.8}$$

denoted as the $(\varrho, \varphi, \vartheta)$ system, and we use $(1 - B)Y_t = (1 - \rho B)\epsilon_t$, $t \geq 1$ called the $(1, \rho)$ system, to approximate it, where $-1 < \varrho \leq 1$, $-1 < \rho \leq 1$, all the roots of

$$1 - \sum_{u=1}^{p} \varphi_u z^u = 0 \quad \text{and} \quad 1 - \sum_{u=1}^{q} \vartheta_u z^u = 0$$

are out of the unit disk, $\{a_u, u \geq 1\}$ and $\{\epsilon_u, u \geq 1\}$ are two white Gaussian noise series, and Y_u, a_u and ϵ_u are all 0 when $u \leq 0$. The $(1,\rho)$ system provides exponential smoothing predictors with ρ as the damping constant. We want to find the best ρ, i.e. the best damping constant, in the sense that it attains the smallest l-step forecasting error variance ratio versus the true system (6.8).

This problem has been discussed in Tiao and Xu (1990), and also in Cox (1961) for the case when $l = 1$ and (6.8) is AR (1). In Xu (1988) and Tiao and Xu (1990), the special case when (6.8) is ARMA (1,1) was treated in detail and used to elucidate that not only the one-step forecasting error variance ratio but also the multi-step forecasting error variance ratios play a role in describing the discrepancy between time series systems and that the optimal solutions for multi-step forecastings are different from that for one-step purpose, which is obtained from the method of maximum likelihood. It was also pointed out there that the popular ARIMA (0,1,1) model is fairly "robust" with respect to ARMA (1,1) series, i.e. for any step prediction purposes the ARIMA (0,1,1) model, which gives the exponential smoothing predictors, provides very efficient forecast for ARMA (1,1) series over a wide range of parameter space. This explains, from one aspect, why the Box-Jenkins approach (Box and Jenkins (1976)) is so successful.

Let

$$f(\omega) \equiv \left|\sum_{t=0}^{\infty} e^{-it\omega} \sum_{v=0}^{t} \xi_v\right|^2 = \frac{|1-e^{-i\omega}|^2 |1-\sum_{t=1}^{q} \vartheta_t e^{-it\omega}|^2}{|1-\varrho e^{-i\omega}|^2 |1-\sum_{t=1}^{p} \varphi_t e^{-it\omega}|^2}, \tag{6.9}$$

where $\xi_t = \xi_t(\varrho, \varphi, \vartheta)$. Then the l-step forecasting error variance by using the $(1,\rho)$ system when $\{Y_t\}$ is generated by the $(\varrho, \varphi, \vartheta)$ system is

$$\sigma_{12}^{(l)2} = \frac{\sigma_a^2}{\pi} \int_0^{\pi} \frac{|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}|^2 f(\omega)}{|1-\rho e^{-i\omega}|^2}\, d\omega \tag{6.10}$$

according to (6.7). Now we have

$$\partial_t \left[\sum_{s=0}^{l-1} \left(\sum_{v=0}^{s} \xi_v\right)^2\right] = \frac{1}{\pi} \int_0^{\pi} \left|\sum_{s=0}^{l-1} e^{-is\omega} \sum_{v=0}^{s} \xi_v\right|^2 \partial_t \log f(\omega)\, d\omega$$

corresponding to (2.13), where $\partial_t = \partial/\partial c_t$ and $\{c_t\}$ are any coordinates system of $S_1(\omega) \equiv (\sigma_a^2/2\pi) f(\omega)$. The divergence measures which we need to consider are

$$\begin{aligned} D_{l,\lambda}(\varrho, \varphi, \vartheta; \rho) = \frac{1}{\pi} \int_0^{\pi} \Bigg\{ & \frac{\lambda + (1-\lambda)|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}|^2}{1+(l-1)(1-\lambda)(1-\rho)^2} \frac{S_1(\omega)}{S_2(\omega)} - 1 \\ & - \log \frac{S_1(\omega)}{S_2(\omega)} - \log \frac{\lambda + (1-\lambda)\sum_{s=0}^{l-1}(\sum_{v=0}^{s} \xi_v)^2}{1+(l-1)(1-\lambda)(1-\rho)^2} \Bigg\} d\omega \end{aligned}$$

and

$$\begin{aligned} D^*_{l,\lambda}(\varrho, \varphi, \vartheta; \rho) = \frac{1}{\pi} \int_0^{\pi} \Bigg\{ & \frac{\lambda + (1-\lambda)|\sum_{s=1}^{l-1} e^{-is\omega} \sum_{v=0}^{s} \xi_v|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}(\sum_{v=0}^{s} \xi_v)^2} \frac{S_2(\omega)}{S_1(\omega)} - 1 \\ & - \log \frac{S_2(\omega)}{S_1(\omega)} - \log \frac{1+(l-1)(1-\lambda)(1-\rho)^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}(\sum_{v=0}^{s} \xi_v)^2} \Bigg\} d\omega, \end{aligned}$$

where $S_2(\omega) = (\sigma_\epsilon^2/2\pi)|1-\rho e^{-i\omega}|^2$. The best ρ for l-step forecasting purpose is described in the following theorem.

THEOREM 6.1. *Let $\rho^{(l)}$ and $\rho^{*(l)} \in (-1,1]$ be defined by*

$$r_l(\varrho,\varphi,\vartheta;\rho^{(l)}) = \min_{\rho\in(-1,1]} r_l(\varrho,\varphi,\vartheta;\rho) \tag{6.11}$$

and

$$r_l^*(\varrho,\varphi,\vartheta;\rho^{*(l)}) = \min_{\rho\in(-1,1]} r_l^*(\varrho,\varphi,\vartheta;\rho),$$

where $r_l(\varrho,\varphi,\vartheta;\rho)$ and $r_l^*(\varrho,\varphi,\vartheta;\rho)$ are the l-step forecasting error variance ratio of the $(1,\rho)$ system from the $(\varrho,\varphi,\vartheta)$ system and the l-step forecasting error variance ratio of the $(\varrho,\varphi,\vartheta)$ system from the $(1,\rho)$ system, respectively, $f(\omega)$ as in (6.9). Then

(i) *when $\varrho \in (-1,1)$*

$$\rho^{*(l)} = 1, \tag{6.12}$$

and when $\varrho = 1$, $\rho^{*(l)}$ is the solution of

$$\frac{\rho+(l-1)(1-\rho^2)}{1+(l-1)(1-\rho^2)} = \frac{\displaystyle\int_0^\pi \frac{\cos\omega|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2}{f(\omega)}\,d\omega}{\displaystyle\int_0^\pi \frac{|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2}{f(\omega)}\,d\omega}; \tag{6.13}$$

(ii) *$\rho^{(l)}$ is the minimizer of*

$$G(\rho) = \int_0^\pi \frac{|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}|^2 f(\omega)}{|1-\rho e^{-i\omega}|^2}\,d\omega \tag{6.14}$$

on $(-1,1]$.

PROOF. (i) Here

$$r_l^*(\varrho,\varphi,\vartheta;\rho) = \frac{1}{\pi}\int_0^\pi \frac{|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2|1-\rho e^{-i\omega}|^2}{[1+(l-1)(1-\rho)^2]f(\omega)}\,d\omega.$$

When $\varrho \in (-1,1)$, we have

$$\int_0^\pi \frac{|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2|1-\rho e^{-i\omega}|^2}{f(\omega)}\,d\omega \begin{cases} = \infty & \text{if } \rho\in(-1,1); \\ < \infty & \text{if } \rho = 1. \end{cases}$$

Hence (6.12) is shown. We can apply Theorem 5.2 to the case when $\varrho = 1$. Now

$$f(\omega) = \frac{|1-\sum_{t=1}^{q}\vartheta_t e^{-it\omega}|^2}{|1-\sum_{t=1}^{p}\varphi_t e^{-it\omega}|^2},$$

and we have (after letting $\lambda \to 0$)

$$\int_0^\pi \frac{|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}|^2}{\sigma_\epsilon^2[1+(l-1)(1-\rho)^2]|1-\rho e^{-i\omega}|^2}\,d\omega = \int_0^\pi \frac{|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2}{\sigma_a^2\left[\sum_{s=0}^{l-1}(\sum_{v=0}^{s}\xi_v)^2\right]f(\omega)}\,d\omega \tag{6.15}$$

and

$$\int_0^\pi \frac{|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}|^2\cos\omega}{\sigma_\epsilon^2[1+(l-1)(1-\rho)^2]|1-\rho e^{-i\omega}|^2}\,d\omega = \int_0^\pi \frac{|\sum_{s=0}^{l-1} e^{-is\omega}\sum_{v=0}^{s}\xi_v|^2\cos\omega}{\sigma_a^2\left[\sum_{s=0}^{l-1}(\sum_{v=0}^{s}\xi_v)^2\right]f(\omega)}\,d\omega \tag{6.16}$$

from (5.6). But

$$\left|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}\right|^2 = 1+(l-1)(1-\rho)^2+2(1-\rho)\sum_{s=1}^{l-1}[1+(1-\rho)(l-1-s)]\cos\omega,$$

$$\int_0^\pi \frac{\cos n\omega}{1-2\rho\cos\omega+\rho^2}\,d\omega = \frac{\pi\rho^n}{1-\rho^2}$$

when n is nonnegative integer and

$$\sum_{s=1}^{l-1}[1+(1-\rho)(l-1-s)]\rho^s = (l-1)\rho,$$

hence

$$\int_0^\pi \frac{\left|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}\right|^2}{|1-\rho e^{-i\omega}|^2}\,d\omega = \frac{\pi}{1-\rho^2}\{1+(l-1)(1-\rho^2)\}. \tag{6.17}$$

Similarly,

$$\int_0^\pi \frac{\left|1+(1-\rho)\sum_{s=1}^{l-1} e^{-is\omega}\right|^2\cos\omega}{|1-\rho e^{-i\omega}|^2}\,d\omega = \frac{\pi}{1-\rho^2}\{\rho+(l-1)(1-\rho^2)\}. \tag{6.18}$$

Therefore (6.13) is derived from (6.15)–(6.18).

(ii) Since

$$r_l(\varrho,\varphi,\vartheta;\rho) = \frac{G(\rho)}{\pi\sum_{s=0}^{l-1}(\sum_{v=0}^{s}\xi_v)^2},$$

where $G(\rho)$ is as in (6.14), the conclusion is trivial.

7. Concluding remarks

There are some ways to create richer families of differential geometrical structures. Let, for example,

$$\Gamma^{(\delta)}_{tuv} = \frac{1+\delta}{2}\Gamma_{tuv} + \frac{1-\delta}{2}\Gamma^{*}_{tuv},$$

where Γ_{tuv} and Γ^{*}_{tuv} are as in (4.24) and (4.25). Each connection in this family still corresponds to the Riemannian metric $g_{tu}(l,\lambda)$ in (3.11), and $\Gamma^{(\delta)}_{tuv}$ and $\Gamma^{(-\delta)}_{tuv}$ are dual.

If, instead of the divergences, we use the notion of yokes, introduced by Barndorff-Nielson (Blæsild (1987)), which are essentially the contrast functions with positive definite metric tensors, we can choose

$$K_1 = \frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2}$$

and $K_2 = 1$ in (2.9), and the D-function, which is a yoke, is

$$D_{l,\lambda}(S_1,S_2) = \frac{1}{\pi}\int_0^{\pi}\left\{\left[\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\zeta_s e^{-is\omega}|^2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\xi_s^2}\right]\frac{S_1(\omega)}{S_2(\omega)} - 1 - \log\frac{S_1(\omega)}{S_2(\omega)}\right\}d\omega.$$

The corresponding metric, dual coordinates, potential functions and dual connections are

$$g_{tu}(l,\lambda) = -\frac{1}{\pi}\int_0^{\pi}\partial_t\left\{\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{S(\omega)}\right\}\partial_u\left\{\frac{S(\omega)}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2}\right\}d\omega,$$

$$\theta^j = \frac{2}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2}\int_0^{\pi} S(\omega)e_j(\omega)\,d\omega,$$

$$\eta_\alpha = -\frac{1}{2\pi^2}\int_0^{\pi}\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{S(\omega)}e_\alpha(\omega)\,d\omega,$$

$$H(S) = \frac{1}{\pi}\int_0^{\pi}\log S(\omega)\,d\omega,$$

$$\Gamma_{tuv} = -\frac{1}{\pi}\int_0^{\pi}\partial_t\partial_u\left\{\frac{S(\omega)}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2}\right\}\partial_v\left\{\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{S(\omega)}\right\}d\omega$$

and

$$\Gamma^{*}_{tuv} = -\frac{1}{\pi}\int_0^{\pi}\partial_t\partial_u\left\{\frac{\lambda + (1-\lambda)|\sum_{s=0}^{l-1}\kappa_s e^{-is\omega}|^2}{S(\omega)}\right\}\partial_v\left\{\frac{S(\omega)}{\lambda + (1-\lambda)\sum_{s=0}^{l-1}\kappa_s^2}\right\}d\omega,$$

respectively. And Theorems 4.1–4.3 are still valid.

In the i.i.d. case, we have some universality of the α-geometry if we require the invariance under the transformations of the random variable and the parameter (Amari (1985)). But for time series case, the invariance under the transformation of the random variable seems no longer reasonable. Then, which kind of requirements should we pose and what kind of geometrical structures can be naturally introduced? This is one of the problems which we need to further investigate.

The establishing of differential geometrical structures related to forecasting error variance ratios certainly has meaningful impact on time series analysis. In Sections 5 and 6, we have discussed the divergences of AR (1) system from any S in L and the divergences of an $(1, \rho)$ system from an $(\varrho, \varphi, \vartheta)$ system. There are some other popular models. For example, AR (p) model, the double exponential smoothing model, the seasonal multiplicative ARIMA $(0,0,1) \times (0,1,1)_{12}$ model and ARIMA $(0,1,1) \times (0,1,1)_{12}$ model. We can get the divergences of these models from their "neighboring" models by the general formula (2.5). It would be interesting and useful to know the shapes of these popular models in the manifold of systems, to know if a smaller set of ARIMA models can for practical purposes be used as approximations to other model. If a smaller set of more robust ARIMA models could be developed, it would help in making choices and might help explain why these models seem to work in practice.

In Xu (1988) and Tiao and Xu (1990), we have shown that the approximations will be different for different step forecastings. This is also displayed by the examples in Sections 5 and 6. One problem which we are interested in is then how the approximation varies as the number of forecasting steps changes. That is, we need to study the function relationship of the (l, λ)-approximations on l.

When the noises are not normal, Amari (1986, 1987b) pointed out that we should also use higher-order moments to describe the differential geometric properties. Another open problem is how to introduce metrics, connections, divergences, etc. in this case.

Acknowledgements

The author is grateful to the referees for their many valuable comments and suggestions. Some discussions with Paul Vos are helpful.

References

Amari, S. (1984). Differential geometry of systems, *Lecture Notes*, **528**, Research Institute for Mathematical Sciences, 235–253, Kyoto University, Kyoto.

Amari, S. (1985). Differential-geometrical methods in statistics, *Lecture Notes in Statist.*, **28**, Springer, New York.

Amari, S. (1986). Geometrical theory on manifolds of linear systems, Technical Report Metr 86-1, Department of Mathematical Engeneering and Instrumentation Physics, University of Tokyo, Tokyo.

Amari, S. (1987*a*). Differential geometrical theory of statistics, *Differentia Geometry in Statistical Inference* (Amari *et al.* (1987) below), 19–94.

Amari, S. (1987*b*). Differential geometry of a parametric family of invertible linear systems—riemannian metric, dual affine connections, and divergence, *Math. Systems Theory*, **20**, 53–82.

Amari, S., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L. and Rao, C. R. (1987). *Differential Geometry in Statistical Inference*, IMS Lecture Notes-Monograph Series, Vol. 10, Hayward, California.

Barndorff-Nielsen, O. E., Cox, D. R. and Reid, N. (1986). The role of differential geometry in statistical theory, *Internat. Statist. Rev.*, **54**, 83–96.

Blæsild, P. (1987). Yokes: elemental properties with statistical applications, *Geometrization in Statistical Theory* (ed. C. T. J. Dodson), 193–198, ULDM Publications, Lancaster.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis*, Holden-Day, San Francisco.

Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods, *J. Roy. Statist. Soc. Ser. B*, **23**, 414–422.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimator in a curved exponential family, *Ann. Statist.*, **11**, 793–803.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Wiley, New York.

Kass, R. E. (1989). The geometry of asymptotic inference, *Statist. Sci.*, **4**, 188–234.

Kolmogorov, A. N. (1941). Stationary sequences in Hilbert space, *Bulletin of Moscow State University*, **2**(6), 1–40, Moscow.

Lauritzen, S. L. (1987). Statistical manifoles, *Differectial Geometry in Statistical Inference* (Amari *et al.* (1987) above), 163–216.

Spivak, M. (1979). *Differential Geometry*, 2nd ed., Publish or Perish, Houston.

Tiao, G. C. and Xu, D. (1989). Robustness of MLE for multi-step predictions: the exponential smoothing case (manuscript).

Xu, D. (1988). Some divergence measures for time series models and their applications, Ph. D. Thesis, University of Chicago (unpublished).