

Curvature form on statistical model manifolds and its application to Bayesian analysis

Fuyuhiko Tanaka

*Department of Mathematical Informatics, University of Tokyo, Tokyo, Japan,
PRESTO, Japan Science and Technology Agency, Saitama, Japan
Email Address: ftanaka@stat.t.u-tokyo.ac.jp*

Received: Received March 18, 2012; Accepted March 18, 2012
Published online: 1 January 2012

Abstract:

In the Bayesian analysis with a statistical model, it is inevitable to determine a prior distribution of the unknown parameter. Since we encounter more and more complicated models in practical use, we need simple criteria by which we know whether there exists a certain class of prior on the statistical model. Recently, Takeuchi and Amari obtained the geometrical condition that a statistical model admits an α -parallel prior, one generalization of well-known Jeffreys prior. Matsuzoe, Takeuchi and Amari studied extensively the geometric condition in a curved exponential family. We formulate their result in terms of differential two form called curvature form on statistical model manifolds, which seems more suitable to evaluation of global properties of statistical model. While the trace of two form vanishes in general class of statistical model including exponential family, it does not vanish in the autoregressive moving average model, which is very fundamental and practically important in time series analysis.

Keywords: time series analysis; objective prior; information geometry

1 Introduction

Recently we encounter more and more complicated statistical models requiring numerical computation in various fields. Bayesian analysis could be helpful to practical user dealing with these models. However, as already known, there exists nontrivial problem of choosing a default prior distribution on the unknown parameter. We here call this kind of prior as a *noninformative prior* or an *objective prior*.

As one candidate of the objective prior, we focus on the α -parallel prior. Historically speaking, it was proposed first by Hartigan [8, 9, 10]. Later, Takeuchi and Amari [18] clarified an interesting connection between the information geometrical properties of the statistical model and the existence of the α -parallel prior. Then, Matsuzoe *et al.* [13] further investigate the condition and obtain the geometrical result for curved exponential family. Although their arguments are restricted to i.i.d. models, we expect that similar consequences hold in the asymptotic setting of time series model.

In the present paper, we formulate their result in terms of differential forms and obtain more concise form of their geometric condition. It is shown that a vanishing two form implies the existence of α -parallel priors on a given statistical model. Then we apply their arguments to non

i.i.d. cases. Generally speaking, it is extremely difficult to deal with non i.i.d. cases theoretically on prior selection. Even in the autoregressive moving average (ARMA) models [6], which is very fundamental and elementary time series model, prior selection has not been discussed theoretically so much. In the present paper, we focus on the Bayesian estimation of the spectral density of stationary Gaussian time series models according to Komaki [12]. Technically, it makes the problem more tractable and formally analogous arguments proceed to some extent.

After the extension of i.i.d. arguments to the above setting, as a nontrivial example we investigate the ARMA models, where information geometrical quantities are explicitly calculated [15, 12]. Thus, we obtain the explicit form of the trace of curvature two form in the ARMA models. As a consequence, we show that there exists no α -parallel prior except for the Jeffreys prior in the ARMA models.

Structure of the present paper is as follows. First, an α -parallel prior on the statistical model manifold is defined and some necessary and sufficient conditions of the existence are reviewed. Then, we define the trace 2-form and another necessary and sufficient condition is given. Then, we consider Bayesian estimation of unknown spectral densities and extend the i.i.d. arguments to stationary Gaussian time series. As a typical example, we take the ARMA model manifolds and calculate the trace 2-form, which is written in a very simple form. Finally, we mention the consequence of the main result, mainly from the statistical viewpoint.

2 Alpha parallel prior and trace 2-form

In the statistical model manifolds of dimension d , the affine volume element is defined by d -form (differential form of degree d). When model manifolds have good properties, such a volume element can be regarded as an extension of the invariant measure, which yields a prior distribution on the parameter space. Here we briefly review the above argument according to Takeuchi and Amari [18].

Let us consider the d -dimensional orientable smooth manifold M with an affine connection ∇ . We shall say that an affine connection ∇ is *locally equiaffine* if around each point x of M there is a parallel volume element, that is, a nonvanishing d -form ω such that $\nabla \omega = 0$ on a neighborhood of each x .

Definition 1. *By an equiaffine connection ∇ on M we mean a torsion-free affine connection that admits a parallel volume element ω on M . If ω is a volume element on M such that $\nabla \omega = 0$, then we say that (∇, ω) is an affine structure on M .*

Now we assume that a statistical model manifold is simply connected. Then, for locally equiaffine connection ∇ , there exists a volume element ω defined on M such that $\nabla \omega = 0$ on M . In a statistical model manifold $M := \{p(x|\theta) : \int p(x|\theta)dx = 1, p(x|\theta) \geq 0, \theta \in \Theta \subseteq \mathbb{R}^d\}$, for an arbitrary $\alpha \in \mathbb{R}$ fixed, a (symmetric) affine connection $\overset{(\alpha)}{\nabla}$ is naturally defined on M by

$$\Gamma_{jk}^{(\alpha)i} := \Gamma_{jk}^{(e)i} + \frac{1-\alpha}{2} T_{ljk} g^{il}, \quad \Gamma_{jk}^{(e)i} := \Gamma_{l;jk}^{(e)i} g^{il},$$

$$g_{ij} := E[\partial_i l \partial_j l], \quad \Gamma_{l;jk}^{(e)} := E[\partial_l l \partial_j \partial_k l], \quad T_{ljk} := E[\partial_l l \partial_j \partial_k l],$$

where g^{ij} is the inverse matrix of the Fisher metric g_{ij} , $l := \log p(x|\theta)$ denotes the log likelihood function and $E[\]$ denotes the expectation with respect to the observation x (see, e.g., Amari and Nagaoka [3] for details). Note that we used Einstein's summation convention. It is shown that $\overset{(\alpha)}{\nabla}$ is equiaffine for some $\alpha \neq 0$ if and only if it is equiaffine for all $\alpha \in \mathbb{R}$ [18]. Thus, we shall say that a statistical model manifold M is statistically equiaffine if the above equivalent conditions are satisfied. In the statistically equiaffine manifolds, we may represent the α -parallel volume element ω as

$$\omega = \pi(\theta)d\theta^1 \wedge \dots \wedge d\theta^d$$

for a certain coordinate $\theta = (\theta^1, \dots, \theta^d) \in \Theta \subseteq \mathbb{R}^d$. Since π is positive on the whole manifold, we take this as a prior distribution on the parameter space Θ .

Definition 2. *In a statistically equiaffine manifold, for fixed $\alpha \in \mathbb{R}$, we call the above form of π an α -parallel prior.*

Note that it could be an improper prior. For properties of α -parallel prior, see Takeuchi and Amari [18]. When $\alpha = 1$, 1-parallel prior is so-called "MLE prior" proposed by Hartigan [10].

We also note that there always exists a $\overset{(0)}{\nabla}$ -parallel volume element $\omega \propto \sqrt{g(\theta)}d\theta^1 \wedge \dots \wedge d\theta^d$, where g is the determinant of the Fisher metric, the invariant volume element in a Riemannian manifold (M, g_{ij}) . This prior distribution $\pi \propto \sqrt{g(\theta)}$ is called the *Jeffreys prior*, well-known in Bayesian statistics. As Jeffreys himself pointed out, it is not necessarily reasonable to adopt the Jeffreys prior as an objective prior in a higher dimensional parametric model. (See, for example, Robert [16] and references therein.)

Now we consider the necessary and sufficient condition that there exists an α -parallel prior ($\alpha \neq 0$) on the statistical model manifold. Hartigan derived the following condition

$$\partial_j \log \pi(\theta) = q_j(\theta), \quad q_j(\theta) = g^{ik} \overset{(e)}{\Gamma}_{k;ij},$$

which is a necessary and sufficient condition for the existence of the MLE prior. Later, Takeuchi and Amari pointed out the above condition is invariant under reparametrization and they derived more geometrical condition, that is,

$$\partial_i T_j - \partial_j T_i = 0, \quad T_i := T_{ikl} g^{kl},$$

is a necessary and sufficient condition that there exists an α -parallel prior ($\alpha \neq 0$).

In the present paper, we take another form of the above condition. Before proceeding, we introduce some notions like the connection 1-form and the curvature 2-form (for general definitions of them, see, e.g., Kobayashi and Nomizu [11]). For simplicity, we adopt concise

definitions using the coordinate vectors $\left\{ \partial_j := \frac{\partial}{\partial \theta^j} \right\}$. (Note that usual statistical model manifolds

are covered with only one coordinate system.) The connection 1-form is defined by $\omega^k_j := \Gamma_{ij}^k d\theta^i$, where $\{\Gamma_{ij}^k\}$ are affine connection coefficients and $\{d\theta^i\}$ are dual basis of coordinate vectors, i.e.,

$d\theta^l(\partial_m) = \delta_m^l$. Note that $\{\omega^k_j\}$ are d^2 1-forms. Then, so-called curvature 2-form is defined by $\Omega^k_j := d\omega^k_j + \omega^k_l \wedge \omega^l_j$. Now we define the trace of the curvature form, the sum of the diagonal components.

Definition 3. We call $\Xi := \text{Tr}\Omega := \Omega^j_j$ as a trace 2-form in the present paper.

For another coordinate system, say $\{\xi^k\}$, we obtain the following transformation rule.

$$\Omega(\xi)^k_j = \Lambda^l_j \Omega(\theta)^m_l (\Lambda^{-1})^k_j, \Lambda^i_j := \frac{\partial \theta^i}{\partial \xi^j}.$$

Thus, it is easily seen that a trace 2-form is invariant under the coordinate transformation. In general, if we take an invariant polynomial f on matrices like $f(A) = \text{Tr}A, \det A$, then the corresponding differential form $f(\Omega)$ becomes an invariant differential form over the manifold. In other words, such differential forms are always independent of parametrization.

$$\Xi^{(\alpha)} = -\frac{\alpha}{2} dT,$$

where $T := T_i d\theta^i$, which is called the *Tchebychev form* in affine geometry [17]. We obtain the following proposition.

Theorem 4. In the statistical model, for fixed α , there exists an α -prior distribution if and only if $\Xi^{(\alpha)} = 0$ on the model.

The above statement is one geometrical representation of Hartigan’s condition using differential forms, which yields a coordinate free expression. For trace 2-form with respect to α connection, the parameter α is only a multiplication factor. Thus we also obtain the following.

Corollary 5. In the statistical model, there exists α -prior distributions for all α if and only if the *Tchebychev form* is closed, $dT = 0$ on the model.

The above statement is the same one derived by Matsuzoe *et al.* (Proposition 3.2 in [13].)

3 Trace 2-form of the ARMA(p,q) model

In the Bayesian analysis of the time series model, it is a considerable obstacle to select an objective prior that is based on a certain justification. In econometrics, we often see the Bayesian analysis using an ad hoc prior (see, e.g., Zellner [21]). Even in the most simple model like the AR(1) model, objective prior selection is very challenging as is discussed in Phillips [14].

Seeking for an objective prior based on a certain theoretical argument, Berger and Yang [4] focused on the reference prior, which was proposed first by Bernardo [5] in the i.i.d. cases. They managed to derive the explicit form of the prior in the AR(1) model, but it seems more difficult to obtain the reference prior when $p \geq 2$.

Here, we deal with ARMA models as an application of the above theory. First we define geometrical quantities in a parametric family of spectral densities according to Amari [2]. Then, we calculate information geometrical quantities on the ARMA models by using the root coordinate. As a result, we obtain the explicit form of the trace two form in the ARMA models.

3.1 Information geometry on the spectral densities

When we have time series data subject to an unknown stationary Gaussian process, the estimation of the spectral density is equivalent to that of the original stochastic process. Although we do not enter the general theory of the estimation of spectral densities, the Fisher metric on the parametric families of spectral densities is given below. The Fisher metric of a model specified by a parametric family of spectral densities $M := \{S(\omega | \theta) : \theta \in \Theta\}$, where θ is a finite-dimensional parameter, is defined by

$$g_{ij} := g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j}\right) = \int_{-\pi}^{\pi} \frac{d\omega}{4\pi} \partial_i \log S(\omega | \theta) \partial_j \log S(\omega | \theta) \quad (1)$$

(Amari [2]).

The above metric is defined such that it coincides with the usual Fisher information as the sample size goes to infinity. Other geometrical quantities are defined in the same manner. For our purpose, we only present the following tensor.

$$T_{ijk} := \int \frac{d\omega}{2\pi} \partial_i \log S(\omega | \theta) \partial_j \log S(\omega | \theta) \partial_k \log S(\omega | \theta).$$

Note that α -connection is determined by g_{ij}, T_{ijk} like i.i.d. cases.

We expect that asymptotic theoretical arguments based on the information geometrical quantities like Fisher metric (1) in the i.i.d. cases are applied to parametric models of stationary Gaussian processes with large length of time series data. Thus, from the viewpoint of the prior selection, it is significant to investigate the existence of the α -parallel prior on the parametric models of spectral densities. In the present paper, as a typical model of stationary time series, we deal with the ARMA(p, q) model. It is already known that there exists the α -parallel prior on the AR(p) models (ARMA($p, 0$) model) and MA(q) models (ARMA($0, q$) model) because they are e(m)-flat (affine connection vanishes). However, as far as the author knows, it has not been investigated yet in the proper ARMA(p, q) models ($p, q > 0$).

3.2 Geometrical quantities on the ARMA(p, q) model manifold

Here, we briefly summarize the ARMA(p, q) model. It consists of random variables $\{X_t\}$ satisfying

$$X_t = -\sum_{i=1}^p a_i X_{t-i} + \sum_{j=0}^q b_j W_{t-j},$$

where $\{W_t\}$ is a Gaussian white noise with mean 0 and variance σ^2 . For basic notions and notations concerning the ARMA(p, q) model see [6].

The explicit form of the spectral density of the ARMA(p, q) model is

$$S(\omega | a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2) = \frac{\sigma^2 |M_b(z)|^2}{2\pi |L_a(z)|^2}, \quad z = e^{i\omega}, \quad (2)$$

where $L_a(z)$ and $M_b(z)$ are the characteristic polynomials and satisfy

$$L_a(Z)X_t = M_b(Z)W_t,$$

where Z is the shift operator that is defined by $ZX_t = X_{t+1}$ and

$$L_a(z) := \sum_{i=0}^p a_i z^{-i}, \quad M_b(z) := \sum_{j=0}^q b_j z^{-j} \text{ with } a_0 = b_0 = 1.$$

Now we adopt another coordinate system. Equation $z^p L_a(z) = z^p + a_1 z^{p-1} + \dots + a_{p-1} z + a_p$ is a polynomial of degree p and has p complex roots, z_1, z_2, \dots, z_p (Note that $|z_i| < 1$ from the stationarity condition). Since a_1, a_2, \dots, a_p are all real, it consequently has the conjugate roots.

Thus, these roots are rearranged in the order like, $z_1, \dots, z_s, z_{s+1}, \dots, z_{2s} \in \mathbb{C}$,

$z_{2s+1}, \dots, z_{2s+r} \in \mathbb{R}$, $2s + r = p$ and $z_{s+j} = \bar{z}_j$ ($1 \leq j \leq s$) (for simplicity, we assume that there are no multiple roots). The roots z_1, z_2, \dots, z_p correspond to the original parameter a_1, a_2, \dots, a_p one-to-one. Likewise, $w^q M_b(w) = w^q + b_1 w^{q-1} + \dots + b_{q-1} w + b_q$ is a polynomial of degree q and has q complex roots, w_1, w_2, \dots, w_q . Note that $|w_i| < 1$ from the invertibility condition. The same argument follows. Now we introduce a coordinate system $\theta = (\theta^0, \theta^1, \dots, \theta^{p+q})$ using these roots

$$\theta^0 := \sigma^2, \quad \theta^1 := z_1, \dots, \theta^p := z_p, \theta^{p+1} := w_1, \dots, \theta^{p+q} := w_q.$$

The formal complex derivatives are defined by

$$\frac{\partial}{\partial \theta} := \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial \bar{\theta}} := \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right),$$

where x and y are both real part and imaginary part of θ . Since the conjugate complex coordinates θ^i and $\bar{\theta}^i$ correspond to x_i and y_i one-to-one, each quantity is evaluated in the original real coordinate if necessary. (See, for example, Gunning and Rossi [7].)

In the coordinate system given above, the Fisher metric on the ARMA(p, q) model g_{IJ} is written in the following way:

$$g_{IJ} = \begin{pmatrix} g_{00} & \cdots & g_{0i} & \cdots \\ \vdots & & \vdots & \\ g_{i0} & \vdots & g_{ij} & \vdots \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix} \quad \text{and} \quad \begin{cases} g_{00} &= \frac{1}{2(\theta^0)^2} = \frac{1}{2\sigma^4} \\ g_{0i} &= g_{i0} = 0 \\ g_{ij} &= \frac{\varepsilon_i \varepsilon_j}{1-\theta^i \theta^j} \end{cases}, \quad (3)$$

where ε_i is defined by

$$\varepsilon_i = \begin{cases} +1, & 1 \leq i \leq p, \\ -1, & p+1 \leq i \leq p+q. \end{cases}$$

In the above coordinate, we easily obtain

$$T_{ijk} = \varepsilon_i \varepsilon_j \varepsilon_k \left\{ \frac{2\theta^i}{(1-\theta^i \theta^j)(1-\theta^i \theta^k)} + \frac{2\theta^j}{(1-\theta^j \theta^k)(1-\theta^j \theta^i)} + \frac{2\theta^k}{(1-\theta^k \theta^i)(1-\theta^k \theta^j)} \right\}.$$

3.3 Trace 2-form of the ARMA(p,q) model

Now we investigate the existence of the α -parallel prior for the ARMA(p, q) model by calculating its trace 2-form. The calculation of trace 2-form seems straightforward in a specific model, but we emphasize that it is not trivial in higher dimensional models. First we show the explicit form of the inverse matrix of the Fisher information g_{IJ} , which is given by

$$g^{00} = 2(\theta^0)^2, \quad g^{0i} = g^{i0} = 0,$$

and

$$g^{mh} = \varepsilon_m \varepsilon_h \frac{(1-\theta^m \theta^h) \prod_{l \neq h} (1-\theta^l \theta^m) \prod_{l \neq m} (1-\theta^l \theta^h)}{\prod_{l \neq h} (\theta^h - \theta^l) \prod_{l \neq m} (\theta^m - \theta^l)}.$$

Dealing with the summation of the terms multiplied with the inverse matrix like $\sum_{j,k} T_{ijk} g^{jk}$ is very cumbersome. By making use of some calculation techniques developed by author [19, 20], we obtain the following result.

Theorem 6. For the ARMA(p, q) model, the trace 2-form is given by

$$\Xi^{(\alpha)} = -4\alpha \sum_{i=1}^p \sum_{j=1}^q \frac{dw_j \wedge dz_i}{(1-z_i w_j)^2},$$

where $|z_i| < 1$ and $|w_j| < 1$.

From theorem 6, we see that the trace 2-form on the ARMA(p, q) model manifold vanishes when $p = 0$ or $q = 0$. It implies that there exist α -parallel priors on the AR(p) model and the

MA(q) model. On contrary, there is no α -parallel prior ($\alpha \neq 0$) in the proper ARMA(p, q) model (i.e., $p > 0$ and $q > 0$).

4 Concluding Remarks

We introduced the trace 2-form as a geometrical quantity on the statistical model. Until now, statistical applications of differential forms have not been considered so much, while other geometrical notions like metric, geodesics, curvature have been investigated considerably [1, 3]. Although our approach to the existence of the α -parallel prior is not outstanding, and equivalent to others already known, but it implies the possibility of applying the differential form to the statistical methods mainly related to the global properties of the statistical model.

Practically more and more complicated statistical models requiring numerical computation appear in various fields and the model manifolds may have nontrivial topology or other global properties. It is known that the differential form is a useful tool to analyse the global properties of differential manifolds. Thus, further development of the analysis of statistical manifolds based on the differential form is expected to become important.

Acknowledgment

This work is supported by JST PRESTO program and Kakenhi for Young Researchers (B). The author is grateful to Prof. Matsuzoe for kind notification of their own work and the reference on the Tchebychev form.

References

- [1] S. Amari, *Differential geometrical methods in statistics*. Springer-Verlag, 1985.
- [2] S. Amari, Differential geometry of a parametric family of invertible linear systems - Riemannian metric, dual affine connections, and divergence. *Math. Syst. Theory*, 20 (1987), 53-82.
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*. AMS, Oxford, 2000.
- [4] J. Berger and R. Y. Yang, Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, 10 (1994), 461-482.
- [5] J. M. Bernardo, Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B*, 41 (1979), 113-147.
- [6] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. Springer-Verlag, New York, 1991.
- [7] C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables*. Englewood Cliffs, NJ, Prentice-Hall, 1965.
- [8] J. A. Hartigan, Invariant prior densities. *Ann. Math. Statist.*, 35 (1964), 836-845.
- [9] J. A. Hartigan, The asymptotically unbiased density. *Ann. Math. Statist.*, 36 (1965), 1137-1152.
- [10] J. A. Hartigan, The maximum likelihood prior. *Ann. Statist.*, 26, no.6 (1998), 2083-2103.
- [11] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I*. Wiley, New York, 1969.
- [12] F. Komaki, Estimating method for parametric spectral densities. *J. Time Ser. Anal.*, 20 (1999), 31-50.
- [13] H. Matsuzoe, J. Takeuchi, and S. Amari, Equiaffine structures on statistical manifolds and Bayesian statistics. *Differential Geom. Appl.*, 24 (2006), 567-578.
- [14] P. C. B. Phillips, To criticize the critics: an objective Bayesian analysis of stochastic trends. *J. Appl. Econ.*, 6 (1991), 333-364.
- [15] N. Ravishanker, E. L. Melnick, and C. L. Tsai, Differential geometry of ARMA models. *J. Time Ser. Anal.*, 11 (1990), 259-0274.

-
- [16] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, 2001.
- [17] U. Simon, A. Schwenk-Schellschmidt, and H. Viesel, *Introduction to the Affine Differential Geometry of Hypersurfaces*, Lecture Notes of the Science University of Tokyo, 1991.
- [18] J. Takeuchi and S. Amari, α -parallel prior and its properties. *IEEE. Trans. Info. Theory*, 51, no.3 (2005), 1011-1023.
- [19] F. Tanaka and F. Komaki, The sectional curvature of AR model manifolds. *Tensor*, 64 (2003), 131-143.
- [20] F. Tanaka, Superharmonic priors for autoregressive models. *Mathematical Engineering Technical Reports*, 2009-18, (2009) 1-20.
- [21] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*. J. Wiley, New York, 1971.
-