TR-91-12

Complex Autoregressive Model and Its Properties

Iwao SEKITA* (sekita@etl.go.jp), Takio KURITA * and Nobuyuki OTSU †

Received 2 May 1991

Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba,
Ibaraki, 305, JAPAN

Abstract

This paper shows the following three properties of the CAR (complex autoregressive) model which were proposed by the authors for shape description and recognition. The CAR coefficients estimated by the least squares fit are shown to be the maximum likelihood estimates and also the estimates obtained by maximizing the mutual information between boundary point z_j and the previous m boundary points $\{z_{j-k}\}_{k=1}^{m}$ on the assumption that each prediction error is a complex Gaussian random variable. And determining the $(m+1)$-th complex autocorrelation coefficient r_{m+1} so as to maximize the entropy of the consecutive $(m+2)$ boundary points on the assumption that the points are a $(m+2)$-dimensional complex Gaussian random variable, is equivalent to determining the coefficient r_{m+1} by the CAR model of order m on the assumption of no correlation between the prediction error ϵ_j and the boundary point z_{j-m-1}.

*Mathematical Informatics Section, Information Science Division
†Currently, Director of Machine Understanding Division

1. Introduction

To recognize arbitrary shapes of patterns on a plane, many shape descriptors have been developed[1]. Kashyap et al.[2] and Dubois et al.[3] used the real AR (autoregressive) model to describe a boundary of a pattern and to recognize the shape. They approximated a boundary by angularly equispaced radius vectors between the boundary and its centroid, and fit a real AR model to a sequence of the vector lengths. However, some problems emerge when the model is applied to two-dimensional discrete-time signals (boundaries). For example, some boundary is sampled to multivalues by a radius vector. Furthermore, even though the lengths of the vectors are ordered[3] according to the tracing of a boundary, there are the possibility that different shapes give the same sequence of lengths.

We proposed the CAR (complex autoregressive) model and used the CAR coefficients and the CPARCOR (complex partial autocorrelation) coefficients for shape recognition[4],[5]. The coefficients are invariant to the similarity transformation (translation, scale and rotation) and choice of the starting point of tracing of a boundary.

In this paper, we show the following three properties of the CAR model. They are the following relations: the relation between the CAR coefficients estimated by the least squares fit and the maximum likelihood estimates, the relation between the coefficients and estimates by maximizing the mutual information, and the relation between the CAR model and the maximum entropy method.

2. CAR (complex autoregressive) model

Let (x_j, y_j) $(j = 0, 1, ..., N-1)$ be a sequence of boundary points obtained by sampling according to the order of tracing by a sequential boundary follower. For simplicity, we consider a closed boundary, namely the relations $x_{-j} = x_{N-j}$ and $y_{-j} = y_{N-j}$ hold.

To construct invariant features to the similarity transformation, we represent each boundary point by a complex number $z_j = x_j + iy_j$. Thus we have a sequence of complex numbers z_j's which correspond to the boundary. For this sequence of complex numbers, a CAR model of order m is defined by a linear combination of preceding m boundary points as a natural extension of the conventional real AR model. That is

$$z_j = \sum_{k=1}^{m} a_k z_{j-k} + \epsilon_j, \tag{1}$$

where ϵ_j is a complex white noise. The CAR coefficients $\{a_k\}_{k=1}^m$ are determined so that the mean squares error $\varepsilon^2(m, \boldsymbol{a})$

$$\varepsilon^2(m, \boldsymbol{a}) = \mathrm{E}_j(\bar{\epsilon}_j \epsilon_j) = \boldsymbol{a}^* R(m) \boldsymbol{a} - \boldsymbol{r}^* \boldsymbol{a} - \boldsymbol{a}^* \boldsymbol{r} + r_0 \tag{2}$$

is minimized, where

$$R(m) = \begin{bmatrix} r_0, & \bar{r}_1, & \bar{r}_2, & ..., & \bar{r}_{m-1} \\ r_1, & r_0, & \bar{r}_1, & ..., & \bar{r}_{m-2} \\ r_2, & r_1, & r_0, & ..., & \bar{r}_{m-3} \\ & & ..... & & \\ r_{m-1}, & r_{m-2}, & r_{m-3}, & ..., & r_0 \end{bmatrix},$$

$$\boldsymbol{a} = [a_1, a_2, a_3, ..., a_m]',$$
$$\boldsymbol{r} = [r_1, r_2, r_3, ..., r_m]',$$
$$r_k = \mathrm{E}_j(\bar{z}_{j-k} z_j).$$

The signs E_j, $\bar{}$, $'$ and * denote $(1/N)\sum_{j=0}^{N-1}$, the complex conjugate, the transpose and the complex conjugate and transpose, respectively. The variable r_k is the complex autocorrelation coefficient and $R(m)$ is an Hermitian matrix ($R^*(m) = R(m)$).

According to inner product in complex Euclidean space and the projection theory, the CAR coefficients $\{a_k\}_{k=1}^m$ minimizing the mean squares error (Eq.(2)) are obtained by solving

$$\boldsymbol{r} = R(m)\boldsymbol{a}, \tag{3}$$

and the least mean squares error $\hat{\varepsilon}^2$ is given by

$$\hat{\varepsilon}^2 = r_0 - \boldsymbol{r}^*\boldsymbol{a}. \tag{4}$$

The CAR coefficients have the following desirable properties for shape recognition. They are invariant to rotation around the origin and choice of the starting point in tracing a boundary. When directions (clockwise or counterclockwise) in tracing are different, the coefficients become complex conjugate of the other. The coefficients become invariant to translation and scale by, for example, setting the origin at the boundary centroid and by fixing the number of the boundary points. In this paper, we assume the origin of $\{z_j\}_{j=0}^{N-1}$ is set at their centroid, that is

$$\mathrm{E}_j z_j = 0. \tag{5}$$

3. Relation to the maximum likelihood estimation

The CAR coefficients estimated by the least squares fit are related to the maximum likelihood estimates[6]. Let $N_{C^2}(\bar{z}, z; \mu, \sigma^2)$ denote 2-dimensional Gaussian probability density function on $\bar{z}z$-plane, and both $\bar{z}$ and z are complex Gaussian random variables:

$$N_{C^2}(\bar{z}, z; \mu, \sigma^2) = \frac{1}{\pi\sigma^2}\exp\{-\frac{(\bar{z}-\bar{\mu})(z-\mu)}{\sigma^2}\}, \tag{6}$$

$$\mathrm{Prob}\{\bar{z} \in [\bar{z}, \bar{z}+d\bar{z}] \wedge z \in [z, z+dz]\} = N_{C^2}(\bar{z}, z; \mu, \sigma^2)\frac{d\bar{z}dz}{2i}.$$

We assume that a forward prediction error ϵ_j is a complex Gaussian random variable, where the mean of $\{\epsilon_j\}_{j=0}^{N-1}$ is 0 because the origin of $\{z_j\}_{j=0}^{N-1}$ is set their centroid and

$$\forall\{a_k\}_{k=1}^m, \quad \mathrm{E}_j\epsilon_j = \mathrm{E}_j(z_j - \sum_{k=1}^m a_k z_{j-k}) = (1 - \sum_{k=1}^m a_k)(\mathrm{E}_j z_j) = 0.$$

Then the likelihood $L(\boldsymbol{a}, \sigma^2)$ is given by

$$\begin{aligned} L(\boldsymbol{a}, \sigma^2) &= \prod_{j=0}^{N-1} N_{C^2}(\bar{\epsilon}_j, \epsilon_j; 0, \sigma^2) \\ &= (\frac{1}{\pi\sigma^2})^N \exp\{-\frac{1}{\sigma^2}\sum_{j=0}^{N-1}(\bar{\epsilon}_j\epsilon_j)\} \\ &= (\frac{1}{\pi\sigma^2})^N \exp\{-\frac{N\varepsilon^2(\boldsymbol{a})}{\sigma^2}\} \end{aligned}$$

and the log likelihood $l(\boldsymbol{a}, \sigma^2)$ is reduced to

$$l(\boldsymbol{a}, \sigma^2) = -N \ln(\pi\sigma^2) - \frac{N\varepsilon^2(\boldsymbol{a})}{\sigma^2}.$$

Thus, maximization of $l(\boldsymbol{a}, \sigma^2)$ requires that the CAR coefficients minimize the mean squares error. This means that the CAR coefficients $\hat{\boldsymbol{a}}$ obtained by the least squares fit are the maximum likelihood estimates. The variance σ^2 maximizing $l(\hat{\boldsymbol{a}}, \sigma^2)$ coincides with the least mean squares error $\hat{\varepsilon}^2$. In this condition, the maximum likelihood $\hat{L}$ and the maximum log likelihood $\hat{l}$ are respectively given by

$$\hat{L} = (\frac{1}{\pi\hat{\varepsilon}^2})^N \exp\{-N\},$$

$$\hat{l} = -N \ln(\pi\hat{\varepsilon}^2) - N.$$

4. Relation to the mutual information

Using a vector notation, Eq.(1) is expressed by

$$z_j = \boldsymbol{a}' \boldsymbol{z}_j + \epsilon_j,$$

and $\boldsymbol{z}_j = [z_{j-1}, z_{j-2}, ..., z_{j-m}]'$. Let us assume ϵ_j is independent of $\boldsymbol{z}_j$ and its density function is

$$p_\epsilon(\bar{\epsilon}, \epsilon) = N_{C^2}(\bar{\epsilon}, \epsilon; 0, \sigma^2),$$

and $\sigma^2 = \varepsilon^2(\boldsymbol{a})$ by the maximum likelihood method. Then mutual information $I[\{z_j\}; \{\boldsymbol{z}_j\}](\boldsymbol{a})$ is given by

$$I[\{z_j\}; \{\boldsymbol{z}_j\}](\boldsymbol{a}) = H[\{z_j\}] - H[\{z_j\}|\{\boldsymbol{z}_j\}](\boldsymbol{a}) = H[\{z_j\}] - H[\{\epsilon_j\}](\boldsymbol{a}), \tag{7}$$

and the entropy $H[\{\epsilon_j\}]$ is deduced to

$$H[\{\epsilon_j\}](\boldsymbol{a}) = -\int p_\epsilon(\bar{\epsilon}, \epsilon; \boldsymbol{a}) \ln(p_\epsilon(\bar{\epsilon}, \epsilon; \boldsymbol{a}))d\epsilon = \ln \pi\varepsilon^2(\boldsymbol{a}) + 1. \tag{8}$$

Since the entropy $H[\{z_j\}]$ is constant, the entropy $H[\{\epsilon_j\}](\boldsymbol{a})$ has to be minimum in order to maximize the mutual information. From Eq.(8), it is equivalent to minimizing the mean squares error $\varepsilon^2(\boldsymbol{a})$. Consequently, obtaining the CAR coefficients $\boldsymbol{a}$ so as to minimize the mean squares error $\varepsilon^2(\boldsymbol{a})$ is equivalent to obtaining $\boldsymbol{a}$ so as to maximize the mutual information I$[\{z_j\}; \{\boldsymbol{z}_j\}](\boldsymbol{a})$ under the above assumption.

5. Relation to the maximum entropy method

The CAR model is also related to the maximum entropy method as well as in the case of the real AR model[7]. Let $N_{C^{2m}}(\bar{\boldsymbol{z}}, \boldsymbol{z}; \boldsymbol{\mu}, \Sigma)$ denote $2m$-dimensional Gaussian probability

density function on $\bar{\boldsymbol{z}}\boldsymbol{z}$-plane, and both $\bar{\boldsymbol{z}}$ and $\boldsymbol{z}$ are m-dimensional complex Gaussian random variables:

$$N_{C^{2m}}(\bar{\boldsymbol{z}}, \boldsymbol{z}; \boldsymbol{\mu}, \Sigma) = \frac{1}{\pi^m |\Sigma|} \exp\{-(\bar{\boldsymbol{z}} - \bar{\boldsymbol{\mu}})' \Sigma^{-1} (\boldsymbol{z} - \boldsymbol{\mu})\}, \tag{9}$$

$$\text{Prob}\{\bar{\boldsymbol{z}} \in [\bar{\boldsymbol{z}}, \bar{\boldsymbol{z}} + d\bar{\boldsymbol{z}}] \wedge \boldsymbol{z} \in [\boldsymbol{z}, \boldsymbol{z} + d\boldsymbol{z}])\} = N_{C^{2m}}(\bar{\boldsymbol{z}}, \boldsymbol{z}; \boldsymbol{\mu}, \Sigma) \prod_{k=1}^{m} (\frac{d\bar{z}^k dz^k}{2i}),$$

where $\boldsymbol{\mu}$ is an m-dimensional complex mean vector, Σ is a $(m \times m)$-dimensional complex covariance matrix.

Let $\boldsymbol{z}_j$ be an $(m+2)$-dimensional complex random variable defined by

$$\boldsymbol{z}_j = [z_j, z_{j-1}, ..., z_{j-m-1}]',$$

and assume that the probability density function is given by $N_{C^{2(m+2)}}(\bar{\boldsymbol{z}}, \boldsymbol{z}; \mathbf{o}, R(m+2))$. It is noted here that $r_1 \simeq r_0$ in $R(m+2)$ usually holds for actual boundary points $\{z_j\}_{j=0}^{N-1}$ since $z_j \simeq z_{j-1}$.

Under this assumption, the entropy H of the distribution is given by

$$H = \ln(|R(m+2)|) + (m+2)\ln(\pi e). \tag{10}$$

Consider the following problem: when $\{r_k\}_{k=0}^{m}$ are exactly known, determine r_{m+1} from $\{r_k\}_{k=0}^{m}$ by the maximum entropy spectral analysis so as to maximize the entropy H. This is equivalent to determine r_{m+1} that maximizes $|R(m+2)|$. The determinant of R(m+2) is always non-negative real number, namely

$$|R(m+2)| \geq 0 \ (\forall m), \tag{11}$$

because $R(m+2)$ is represented by $\frac{1}{N} Z^*(m+2) Z(m+2)$, where

$$Z(m+2) = \begin{bmatrix} z_0, & z_{N-1}, & z_{N-2}, & z_{N-3}, & ..., & z_{N-m-1} \\ z_1, & z_0, & z_{N-1}, & z_{N-2}, & ..., & z_{N-m} \\ z_2, & z_1, & z_0, & z_{N-1}, & ..., & z_{N-m+1} \\ & & & ..... & & \\ z_{N-1}, & z_{N-2}, & z_{N-3}, & z_{N-4}, & ..., & z_{N-m-2} \end{bmatrix}.$$

Let $S(m+1)$ be

$$S(m+1) = \begin{bmatrix} r_1, & r_0, & \bar{r}_1, & \bar{r}_2, & ..., & \bar{r}_{m-1} \\ r_2, & r_1, & r_0, & \bar{r}_1, & ..., & \bar{r}_{m-2} \\ r_3, & r_2, & r_1, & r_0, & ..., & \bar{r}_{m-3} \\ & & & ..... & & \\ r_{m+1}, & r_m, & r_{m-1}, & r_{m-2}, & ..., & r_1 \end{bmatrix}.$$

Then the determinant $|R(m+2)|$ can be given by

$$|R(m+2)| = -|R(m)||r_{m+1} - \frac{u}{|R(m)|}|^2 + R_1, \tag{12}$$

where R_1 is a real number not including r_{m+1},

$$u = (-1)^{m+1} \sum_{\ell=1}^{m} r_\ell S(m+1)_{\ell,1}, \tag{13}$$

and $S(m+1)_{\ell,1}$ is the cofactor of the element $(\ell, 1)$ of $S(m+1)$. From Eq.(11), $|R(m+2)|$ becomes maximum when $r_{m+1} = u/|R(m)|$. That is

$$r_{m+1}|R(m)| - u = (-1)^m |S(m+1)| = 0.$$

Thus $|R(m+2)|$ becomes maximum when $|S(m+1)| = 0$. This also means that $|R(m+2)|$ becomes maximum when $|S^*(m+1)| = 0$.

Next, consider the CAR model of order m

$$z_j - \sum_{k=1}^{m} a_k z_{j-k} = \epsilon_j. \tag{14}$$

The relations

$$r_{k_1} - \sum_{k=1}^{m} a_k r_{k_1-k} = 0 : k_1 = 1, 2, ..., m \tag{15}$$

hold because $\boldsymbol{r} - R(m)\boldsymbol{a} = 0$. Suppose that $\mathrm{E}_j(\epsilon_j \bar{z}_{j-m-1}) = 0$. Multiplying $\bar{z}_{j-m-1}$ to the both sides of Eq.(14) and taking expectation, we also have

$$r_{m+1} - \sum_{k=1}^{m} a_k r_{m+1-k} = 0. \tag{16}$$

From (15), (16), $[1, a_1, a_2, ..., a_m]' \neq \mathbf{o}$ and $r_{-k} = \bar{r}_k$, the determinant must be zero, namely $|S(m+1)| = 0$. When we consider the backward predicted CAR model, the determinant $|S^*(m+1)|$ also must be 0 in the same manner as for the forward predicted model expressed by Eq.(14).

Consequently, determining r_{m+1} so as to maximize the entropy H in Eq.(10) is equivalent to determining it by the CAR model of order m on the assumption that $\mathrm{E}_j(\epsilon_j \bar{z}_{j-m-1}) = 0$.

Acknowledgements
The authors are grateful to the member of the Mathematical Informatics Section for valuable discussions especially to Dr. M. Tanaka for the expression of the Gaussian probability density function of a complex Gaussian random variable. They also express thanks to Dr. K. Tamura, Director of the Information Science Division, for his encouragement to continue the project.

References

[1] T. Pavlidis : "Algorithm for shape analysis of contours and waveforms," IEEE Trans. Pattern Anal. & Machine Intelli. , **PAMI-2**, 4, pp. 301-312 (1980).

[2] R. L. Kashyap and R. Chellappa : "Stochastic models for closed boundary analysis: representation and reconstruction," IEEE Trans. Inform. Theory, **IT-27**, 5, pp. 627-637 (1981).

[3] S. R. Dubois and F. H. Glanz : "An autoregressive model approach to two-dimensional shape classification," IEEE Trans. Pattern Anal. & Machine Intelli. , **PAMI-8**, 1, pp. 55-66 (1986).

[4] N. Otsu, and T. Kurita: "Proposal of complex autoregressive model as a shape descriptor," National Convention Records of IEICE Japan, **D-496** (1989). (in Japanese)

[5] I. Sekita, T. Kurita and N. Otsu: "Classification by a complex autoregressive model, " Trans. of IEICE Japan D-II, **J73-D-II**, 6, pp. 804-811 (1990). (in Japanese)

[6] T. Kurita, I. Sekita and N. Otsu: "Distance measures for planar shape analysis based on a complex autoregressive model, " Trans. of IEICE Japan D-II, **J73-D-II**, 9, pp. 1493-1503 (1990). (in Japanese)

[7] A. van den Bos : "Alternative interpretation of maximum entropy spectral analysis," IEEE Trans. Information Theory, **IT-17**, 7, pp. 493- 494 (1971).