Liet. matem. rink., T. 40, spec. nr., 408–414

Segmentation of multivariate autoregressive sequences

Antanas LIPEIKA, Joana LIPEIKIENĖ (MII)
e-mail: lipeika@ktl.mii.lt

1. Statement of the problem

We consider the output $\vec{X}(n)$, $n = \ldots, 1, 2, \ldots, N, \ldots$, of a linear discrete time r-dimensional time-variant dynamic system, the input of which is the sequence of Gaussian r-dimensional independent random variables $\vec{V}(n)$, $n = \ldots, 1, 2, \ldots, N, \ldots$, with zero mean and unit covariance matrix. The system satisfies stability conditions and its structure is described by an autoregressive (AR) model

$$\vec{X}(n) = -A_1(n)\vec{X}(n-1) - A_2(n)\vec{X}(n-2) - \ldots - A_p(n)\vec{X}(n-p) + B(n)\vec{V}(n), \quad (1)$$

where p is an order of the system; $\vec{X}(n) = (x_1(n), \ldots, x_r(n))^T$; $\vec{V}(n) = (v_1(n), \ldots, v_r(n))^T$; T – a sign of transposition; $A_1(n), \ldots, A_p(n), B(n)$ are matrixes of dimension $r \times r$; $\det B(n) \neq 0$. Parameters of the system $Q(n) = (A_1(n), \ldots, A_p(n), B(n))$ at unknown points of time $u = (u_1, \ldots, u_M)$, $u_1 < u_2 < \ldots u_M$ abruptly change their values, but are constant and known between change points.

Let us denote

$$Q_j = (A_1^{(j)}, \ldots, A_p^{(j)}, B_j), \quad j = 1, 2, \ldots, M+1,$$

where

$$A_i^{(j)} = \begin{pmatrix} a_{11}^{(j)}(i) \ldots a_{1r}^{(j)}(i) \\ \ldots\ldots\ldots\ldots \\ a_{r1}^{(j)}(i) \ldots a_{rr}^{(j)}(i) \end{pmatrix}, \quad B_j = \begin{pmatrix} b_1^{(j)} & 0 & \ldots & 0 \\ \ldots\ldots\ldots\ldots & & & \\ 0 & 0 & \ldots & b_r^{(j)} \end{pmatrix}.$$

Then

$$Q(n) = \begin{cases} Q_1, & n = \ldots, 1, 2, \ldots, u_1 \\ Q_2, & n = u_1 + 1, \ldots, u_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ Q_j, & n = u_{j-1} + 1, \ldots, u_j \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ Q_{M+1}, & n = u_M + 1, \ldots, N, \ldots \end{cases} \quad (2)$$

Let us assume $u_1 > p+1$ and denote $u_0 = p+1$, $u_{M+1} = N$. The problem consists in obtaining the estimates $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_M)$ of change-points $u = (u_1, \ldots, u_M)$, using realization $\vec{X} = (\vec{X}(1), \ldots, \vec{X}(n))$ of a random sequence $\vec{X}(n)$.

We will solve this problem using maximum likelihood estimate

$$\hat{u} = \arg \max_{\substack{u \\ p<u_1<\ldots<u_M<N}} L(u|\vec{X}), \tag{3}$$

where $L(u|\vec{X})$ is logarithmic likelihood function of change points $u = (u_1, \ldots, u_M)$.

2. Derivation of conditional distribution density of the observable random sequence

To maximize the likelihood function $L(u|\vec{X})$ we need to know the conditional distribution $p(\vec{X}|u)$.

Since input of the system (1) is a sequence of Gaussian r-dimensional independent random variables (independent of u) with zero mean and unit covariance matrix and the system (1) is linear, the conditional distribution density $p(\vec{X}|u)$ is also Gaussian.

We can also express $p(\vec{X}|u)$ in terms of conditional distributions

$$\begin{aligned} p(\vec{X}|u) &= p(\vec{X}(1), \vec{X}(2), \ldots, \vec{X}(p)) \\ &\quad \times \prod_{n=p+1}^{N} p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(1)). \end{aligned} \tag{4}$$

As output of the system (1) at time instant n depends only on p previous outputs $x(n-1), \ldots, x(n-p)$, we can rewrite (4) in the form

$$\begin{aligned} p(\vec{X}|u) &= p(\vec{X}(1), \vec{X}(2), \ldots, \vec{X}(p)) \\ &\quad \times \prod_{n=p+1}^{N} p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p)). \end{aligned}$$

Taking into account change points u and the condition $p < u_1 < u_2 \ldots < u_M < N$ we have

$$\begin{aligned} p(\vec{X}|u) &= p(\vec{X}(1), \vec{X}(2), \ldots, \vec{X}(p)) \\ &\quad \times \prod_{n=p+1}^{u_1} p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p)) \\ &\quad \times \prod_{n=u_1+1}^{u_2} p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p)) \times \ldots \\ &\quad \times \prod_{n=u_M+1}^{N} p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p)). \end{aligned} \tag{5}$$

Since $p(\vec{X}|u)$ is Gaussian distribution density, $p(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p))$ is also Gaussian distribution density with conditional mean

$$M(n) = E\left[\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p))\right] \tag{6}$$

and conditional covariance matrix

$$K(n) = E[(\vec{X}(n) - M(n))^2|u, \vec{X}(n-1), \ldots, \vec{X}(n-p))].$$

Taking into account (1) we can write (6) as

$$M(n) = -A_1(n)\vec{X}(n-1) - \ldots - A_p(n)\vec{X}(n-p)$$

and

$$p\left(\vec{X}(n)|u, \vec{X}(n-1), \ldots, \vec{X}(n-p)\right) = \frac{(\det\ K(n))^{-\frac{1}{2}}}{(2\pi)^{r/2}} \exp\left\{-\frac{1}{2}\left(\vec{X}(n) - M(n)\right)^T K(n)^{-1}\left(\vec{X}(n) - M(n)\right)\right\}, \tag{7}$$

where

$$(\vec{X}(n) - M(n)) = \vec{X}(n) + A_1(n)\vec{X}(n-1) + \ldots + A_p(n)\vec{X}(n-p).$$

Using (2) we can denote

$$K_j = K(n)\Big|_{Q(n)=Q_j} = \begin{pmatrix} (b_1^{(j)})^2 0 \ldots 0 \\ \ldots\ldots\ldots\ldots \\ 0\,0 \ldots (b_r^{(j)})^2 \end{pmatrix}, \tag{8}$$

$$(\det K_j)^{-\frac{1}{2}} = (\det K(n))^{-\frac{1}{2}}\Big|_{Q(n)=Q_j} = (b_1^{(j)})^{-1}(b_2^{(j)})^{-1} \ldots (b_r^{(j)})^{-1}, \tag{9}$$

$$K_j^{-1} = K(n)^{-1}\Big|_{Q(n)=Q_j} = \begin{pmatrix} (b_1^{(j)})^{-2} 0 \ldots 0 \\ \ldots\ldots\ldots\ldots \\ 0 \quad 0 \ldots (b_r^{(j)})^{-2} \end{pmatrix}$$

and

$$R(n, j|\vec{X}) = \left(\vec{X}(n) - M(n)\right)\Big|_{Q(n)=Q_j} = \vec{X}(n) + A_1^{(j)}\vec{X}(n-1) + \ldots + A_p^{(j)}\vec{X}(n-p). \tag{10}$$

Then, substituting (7) into (5) and taking into account (2) and (8)–(10) we can write

$$p(\vec{X}|u) = p(\vec{X}(1), \vec{X}(2), \ldots, \vec{X}(p)) \times (2\pi)^{\frac{-r(N-p)}{2}}$$
$$\times \prod_{j=1}^{M+1} (\det\ K_j)^{\frac{-(u_j - u_{j-1})}{2}} \times \prod_{j=1}^{M+1} \prod_{n=u_{j-1}+1}^{u_j} R(n, j|\vec{X})^T K_j^{-1} R(n, j|\vec{X}). \quad (11)$$

Taking the logarithm of the $p(\vec{X}|u)$, from (11) finally we have

$$\log p(\vec{X}|u) = \log p(\vec{X}(1), \vec{X}(2), \ldots, \vec{X}(p))$$
$$- \frac{r(N-p)}{2} \log (2\pi) - \frac{1}{2} \sum_{j=1}^{M+1} (u_j - u_{j-1}) \log (\det\ K_j)$$
$$- \frac{1}{2} \sum_{j=1}^{M+1} \sum_{n=u_{j-1}+1}^{u_j} R(n, j|\vec{X})^T K_j^{-1} R(n, j|\vec{X}). \quad (12)$$

3. Remark

For the given realization $\vec{X}$ of the random sequence instead of maximizing (12) we can maximize objective function $\Theta(u|\vec{X})$, which differs from (12) by an additive constant not depending on u, i.e.,

$$\hat{u} = \arg \max_{\substack{u \\ p<u_1<\ldots<u_M<N}} \log\ p(\vec{X}|u) = \arg \max_{\substack{u \\ p<u_1<\ldots<u_M<N}} \Theta(u|\vec{X}), \quad (13)$$

where

$$\Theta(u|\vec{X}) = L_1(u_1|\vec{X}) + L_2(u_2|\vec{X}) + \ldots + L_M(u_M|\vec{X}). \quad (14)$$

Each of the functions $L_i(u_i|\vec{X})$, $i = 1, 2, \ldots, M$ depends only on one variable, corresponding to unknown change point u_i and can be expressed as

$$L_i(n|\vec{X}) = L_i(n-1|\vec{X}) + \log \frac{\det\ K_{i+1}}{\det\ K_i} + [R(n, i+1|\vec{X})^T K_{i+1}^{-1} R(n, i+1|\vec{X})$$
$$- R(n, i|\vec{X})^T K_i^{-1} R(n, i|\vec{X})], \quad i = 1, 2, \ldots, M, \quad n = p+1, \ldots, N \quad (15)$$

with initial conditions $L_i(n|\vec{X}) = 0$, $i = 1, 2, \ldots, M$.

4. Maximization of the likelihood function

The function $\Theta(u|\vec{X})$ as well as the function $L(u|\vec{X})$ is defined only for discrete time arguments and one possible way to find location of its global maximum is full search in

M dimensional space, restricted by the condition $p < u_1 < u_2 < \ldots < u_M < N$. Amount of computations using full search is of order N^M. Such amount of computation is not acceptable in many practical applications. Therefore an effective search algorithm is needed for $\Theta(u|\vec{X})$ maximization.

Since the function $\Theta(u|\vec{X})$ is sum of partial functions $L_i(u_i|\vec{X})$, $i = 1, 2, \ldots, M$ and each of these partial functions depends only on one variable, we can use the dynamic programming method (Bellman, 1957) to determine location of the global maximum of this function.

Dynamic programming method is rather the methodology than the method. The principle of optimality, which is the basis of a class of computational algotithms for the above optimization problem, according to Bellmam, is "An optimal policy has the property that, whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision".

Maximization procedure fully depends on restrictions in the range of the maximization variables. In our case the restrictions are

$$p < u_1 < u_2 < \ldots < u_M < N.$$

Let us begin from maximization of $L_1(u_1|\vec{X})$, $p < u_1 < u_2$. We can see that maximum of $L_1(u_1|\vec{X})$ depends only on u_2 . So we can define the function $g_1(u_2|\vec{X})$ such that

$$g_1(u_2|\vec{X}) = \max_{\substack{u_1 \\ p<u_1<u_2}} L_1(u_2|\vec{X}), \quad u_2 = p+2, \ldots, N$$

We can see that the global maximum of the of the function $g_1(u_2|\vec{X})$ (and also of the function $L_1(u_1|\vec{X})$) depends only on u_2 (one variable) for all possible values of $u_3, u_4, \ldots, u_M$, satisfying the condition $u_2 < u_3 < u_4 < \ldots < u_M < N$.

Now let us maximize $L_1(u_1|\vec{X}) + L_2(u_2|\vec{X})$, $p < u_1 < u_2 < u_3$. We can see that maximum of $L_1(u_1|\vec{X}) + L_2(u_2|\vec{X})$ depends only on u_3 . So we can define the function $g_2(u_3|\vec{X})$ such that

$$g_2(u_3|\vec{X}) = \max_{\substack{u_2 \\ p+1<u_2<u_3}} [L_2(u_2|\vec{X}) + g_1(u_2|\vec{X})], \quad u_3 = p+3, \ldots, N.$$

The global maximum of the function $g_2(u_3|\vec{X})$ (and also of the function $L_1(u_1|\vec{X}) + L_2(u_2|\vec{X})$) depends only on u_3 (one variable) for all possible values of $u_4, u_5, \ldots, u_M$, satisfying the condition $u_3 < u_4 < u_5 < \ldots < u_M < N$.

Similarilly we can maximize the function (Lipeika, 2000)

$$L_1(u_1|\vec{X}) + L_2(u_2|\vec{X}) + \ldots + L_i(u_i|\vec{X}), \quad p < u_1 < u_2 < u_3 < \ldots < u_i < u_{i+1}$$

as the maximum of $L_1(u_1|\vec{X}) + L_2(u_2|\vec{X}) + \ldots + L_i(u_i|\vec{X})$ depends only on u_{i+1} . So we can define the function $g_i(u_{i+1}|\vec{X})$ such that

$$g_i(u_{i+1}|\vec{X}) = \max_{\substack{u_i \\ p+i-1<u_i<u_{i+1}}} [L_i(u_i|\vec{X}) + g_{i-1}(u_i|\vec{X})], \quad u_{i+1} = p+i+1, \ldots, N. \tag{16}$$

The global maximum of the function $g_i(u_{i+1}|\vec{X})$ (and also of the function $L_1(u_1|\vec{X}) + L_2(u_2|\vec{X}) + \ldots + L_i(u_i|\vec{X})$) depends only on u_{i+1} (one variable) for all possible values of $u_{i+2}, u_{i+3}, \ldots, u_M$, satisfying the condition $u_{i+1} < u_{i+2} < u_{i+3} < \ldots < u_M < N$.

For maximization of

$$\Theta(u|\vec{X}) = L_1(u_1|\vec{X}) + L_2(u_2|\vec{X}) + \ldots + L_M(u_M|\vec{X})$$

we have

$$g_M(u_{M+1}|\vec{X}) = \max_{\substack{u_M \\ p+M-1<u_M<u_{M+1}}} [L_M(u_M|\vec{X}) + g_{M-1}(u_M|\vec{X})], \quad u_{M+1} = p+M+1, \ldots, N, \tag{17}$$

where u_{M+1} is an additional variable, which means the possible length of the available realization of the random sequence. Since the length of the realization is N, the value $g_M(N|\vec{X})$ is the global maximum of the $\Theta(u|\vec{X})$. The functions $g_i(n|\vec{X})$ are not decreasing functions with respect to $n = p+1, \ldots, N$.

The maximum likelihood estimate $\hat{u} = [\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_M]$ of the change points u is obtained in the following way

$$\hat{u}_k = \min [\arg \max_{\substack{n \\ p+k \leqslant n \leqslant \hat{u}_{k+1}}} g_k(n|\vec{X})], \quad k = M, M-1, \ldots, 2, 1, \tag{18}$$

where, for convenience, we made a notation $\hat{u}_{M+1} = N$.

For further reduction of computation amount, we can compute the functions $g_i(u_{i+1}|\vec{X})$, $i = 1, \ldots, M$ recursively

$$g_1(u_2|\vec{X}) = \max [g_1(u_2-1|\vec{X}), L_1(u_2-1|\vec{X})], \quad u_2 = p+3, \ldots, N$$

with the initial condition $g_1(p+2|\vec{X}) = L_1(p+1|\vec{X})$.

And for $i = 2, \ldots, M$

$$g_i(u_{i+1}|\vec{X}) = \max \left\{ g_i(u_{i+1}-1|\vec{X}), [g_{i-1}(u_{i+1}-1|\vec{X}) + L_i(u_{i+1}-1|\vec{X})] \right\}, \quad u_{i+1} = p+i+2, \ldots, N$$

with the initial conditions

$$g_i(p+i+1|\vec{X}) = L_i(p+i|\vec{X}) + g_{i-1}(p+i|\vec{X}), \quad i = 2, \ldots, M.$$

It is interesting to note, that by assuming $B(i) = I$, $i = 1, 2, \ldots, M+1$ (unit matrix) and using the same optimization method we obtain least square (LS) estimates of change points $u = [u_1, u_2, \ldots, u_M]$. This solution enables to reduce amount of computations from the order N^M (full search) to the order $N \times M$ (dynamic programming), where N is a length of an observable random sequence and M is the number of change points in this sequence.

References

[1] R.E. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, New Jersey, USA (1957).
[2] A. Lipeika, Optimal Segmentation of Random Sequences, *Informatica*, 11(3), 243–256 (2000).

Daugiamačių autoregresinių sekų segmentacija

A. Lipeika, J. Lipeikienė

Darbe nagrinėjama daugiamačių autoregresinių sekų segmentacija naudojant maksimalaus tikėtinumo įvertį. Pasiūlytas efektyvus tikėtinumo funkcijos maksimumo paieškos būdas.