On Two Selected Topics Connected with Stochastic Systems Theory

YU. A. ROZANOV

International Institute for Applied Systems Analysis, Laxenburg, Austria

I. Markov State Representation. Considering a control of a stochastic system it is desirable to choose such a phase-state that the system evolution in time forms a Markov process.

Of course it is not always possible if one is restricted to use a state-space only of a certain type (say, a finite dimensional vector space). And if it is possible then there are many ways to do so. In [1]–[4] some relevant problems were considered.

Let us consider the situation in a case of Gaussian stationary processes.

Let $x(t)=\{x_\alpha(t)\}$ be an arbitrary family of univariate Gaussian stationary processes $x_\alpha(t)$, $-\infty<t<\infty$, $H^t(x)$ be a linear closure of all variables $x_\alpha(t)$, H^{-t} and H^{t+} be a linear closure correspondingly of all subspaces $H^s(x)$, $s\leqslant t$, and $H^s(x)$, $s\geqslant t$, in a Hilbert space of random variables with the usual inner product $(h_1,h_2)=Eh_1h_2$.

Let $P^{-t}(x)$ be the orthogonal projector onto $H^{-t}(x)$. The process $x(t)$ is Markovian if

$$P^{-t}(x)H^{t+}(x)=H^t(x). \tag{1}$$

Let us say that the Markov process $x(t)$ gives a *Markov state representation* for a stationary process $y(t)$ if

$$H^t(y)\subseteq H^t(x). \tag{2}$$

In a case of finite dimensional vector-processes it means that

$$y(t)=Cx(t) \tag{2'}$$

where C is a constant matrix.

There is a feeling that for any given $y(t)$ it must be in some sense minimal process $x(t)$ provided the Markov state representation. How can we describe this minimal Markov process if it exists?

APPLIED MATHEMATICS & OPTIMIZATION, Vol. 3, No. 1

Let us consider a class of all Markov processes $x(t)$ satisfying the condition (2) with the same "innovation" as $y(t)$ namely such that

$$H^{-t}(x)=H^{-t}(y)=H^{-t}. \tag{3}$$

We have

$$H^{t}(y)\subseteq H^{t}(x), \quad H^{t+}(y)\subseteq H^{t+}(x)$$

and

$$P^{-t}H^{t+}(y)\subseteq P^{-t}H^{t+}(x)=H^{t}(x)$$

where $P^{-t}=P^{-t}(x)=P^{-t}(y)$. Thus for the Markov process $x(t)$

$$P^{-t}H^{t+}(y)\subseteq H^{t}(x) \tag{4}$$

where $P^{-t}H^{t+}(y)$ is the subspace "splitting" the future $H^{t+}(y)$ and the past $H^{-t}(y)$ of the process $y(t)$, $-\infty<t<\infty$.

Let us form a stationary process $Y(t)$, $-\infty<t<\infty$, with

$$H^{t}(Y)=P^{-t}H^{t+}(y)$$

say

$$Y(t)=\{V_t h, h\in P^{-0}H^{0+}(y)\} \tag{5}$$

where V_t, $-\infty<t<\infty$, means a family of unitary operators in our Hilbert space, generated by equations

$$x_\alpha(t+s)=V_t x_\alpha(s); \ -\infty<s,t<\infty.$$

Obviously

$$H^{t}(y)\subseteq H^{-t}\cap H^{t+}(y)\subseteq H^{t}(Y) \tag{6}$$

and the stationary process $Y(t)$ has the same innovation as $y(t)$:

$$H^{-t}(Y)=H^{-t}.$$

Let us show that $Y(t)$ is a Markov process.

Indeed for any $h\in H^{0}(Y)$ there is $h^{t}\in H^{0+}(y)$ such that

$$P^{-0}h^{+}=h, \quad h^{+}-h\perp H^{-0}$$

and so

$$V_t(h^{+}-h)=V_t h^{+}-V_t h\perp H^{-t}.$$

$$H^{-0}\subseteq H^{-t}, \quad V_t h^{+}-V_t h\perp H^{-0}, \quad t\geqslant 0.$$

We have

$$P^{-0}(V_t h^+ - V_t h) = 0,$$

$$P^{-0}(V_t h) = P^{-0}(V_t h^+) \in H^0(Y)$$

because for $t \geqslant 0$

$$V_t h^+ \in H^{0+}(y) \quad \text{if} \quad h^+ \in H^{0+}(y).$$

Thus

$$P^{-0} H^{0+}(Y) = H^0(Y),$$

$$P^{-t} H^{t+}(Y) = H^t(Y)$$

what we needed to prove.

As a result we obtain the following.

Theorem 1. *There is the Markov process $Y(t)$ which space $H^t(Y)$ coincides with the minimal subspace splitting the future $H^{t+}(y)$ and the past $H^{-t}(y)$ of $y(t)$. This process $Y(t)$ gives the minimal, Markov state representation for $y(t)$ in the sense that*

$$H^t(Y) \subseteq H^t(x) \tag{6}$$

for any other Markov process $x(t)$ provided the relation (2).

Let us find now a condition for existence of *finite-dimensional* Markov state representation.

Let us consider a finite-dimensional Markov vector-process $x(t)$. In a case of discrete time $t = 0, \pm 1, \ldots$ we have

$$x(t) - Ax(t-1) = \sigma u(t) \tag{7}$$

where A, σ are constant matrices of a proper size and $u(t)$ means the corresponding innovation process for the pure non-deterministic process. The spectral transfer matrix-function can be found as

$$\phi_{xu} = (e^{i\lambda} I - A)^{-1} \sigma \tag{8}$$

and if $y(t) = Cx(t)$ then $\phi_{yu} = C\phi_{xu}$ so the spectral densities matrix $\phi_{yy} = \phi_{yu} \cdot \phi_{yu}^*$ of the process $y(t)$ must be a rational function of $z = e^{i\lambda}$.

Similarly in a case of continuous time t we have

$$dx(t) - ax(t) = \sigma du(t) \tag{7'}$$

and

$$\phi_{xu} = (i\lambda I - a)^{-1} \sigma \tag{8'}$$

so $\phi_{yu} = C\phi_{xu}$ and the spectral density $f_{yy} = \phi_{yu}\phi_{yu}^*$ is a rational function of $z = i\lambda$.

As well known for a such type of stationary process $y(t)$ there is the explicit finite-dimensional Markov state representation. For example the corresponding Markov process $x(t)$ can be taken as follows. Let

$$t = 0, \pm 1, \dots$$

and

$$y(t) = \int_{-\pi}^{\pi} e^{i\lambda t}\phi_{yu}(e^{-i\lambda})\, d\phi_u(\lambda)$$

be a spectral representation of $y(t)$ upon its multi-dimensional innovation process $u(t)$ with ϕ_{yu} as the maximal (rational) factor in the factorization

$$\phi_{yu}\phi_{yu}^* = f_{yy}$$

see [5]. Let

$$Q(z) = \sum_{k=0}^{n} q_k z^k$$

be a polynomial with a non-degenerated coefficient $q_0(\det q_0 \neq 0)$ such thai

$$\phi_{yu} Q = \sum_{k=0}^{m} c_k z^k$$

is also polynomial matrix. Say if ϕ_{yu} is polynomial itself then one can take $Q = I$. Let us set

$$x_0(t) = \int_{-\pi}^{\pi} e^{i\lambda t} Q(e^{-i\lambda})^{-1}\, d\Phi_u(\lambda)$$

and

$$x(t) = \{x_k(t)\}$$

where

$$x_k(t) = x_0(t-k); \qquad k = 0, 1, \dots, r-1 \quad (r = \max m, n)$$

Obviously the process $x(t)$ is Markovian because the equation of the type (7) holds:

$$x_0(t) - q_0^{-1}\sum_{k=0}^{n-1} q_{k+1} x_k(t-1) = q_0^{-1} u(t)$$

$$x_k(t) - x_{k-1}(t-1) = 0; \qquad k = 1, \dots, r-1$$

$$\sum_{k=0}^{m} c_k x_k(t) = \int_{-\pi}^{\pi} e^{i\lambda t}\left[\sum_{k=0}^{m} c_k e^{-i\lambda k}\right] Q(e^{-i\lambda})^{-1} d\Phi_u(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{i\lambda t}\phi_{yu}(e^{-i\lambda})\, d\Phi_u(\lambda) = y(t)$$

which gives us the representation (2′).

One can proceed similarly in a case of continuous time t.

As an additional result we obtained the following.

***Theorem* 2.** *The minimal subspace splitting the future and the past is finite dimensional if and only if the process $y(t)$ has a spectral densities matrix with rational components.*[1]

As well known for univariate process $y(t)$ with the spectral density

$$f = \left|\frac{P}{Q}\right|^2$$

(where P/Q is the outer factor) the minimal splitting subspace is generated by functions

$$\frac{e^{i\lambda(t-k)}}{Q}; \qquad k = 0, \ldots, r-1 \tag{9}$$

in a case of the discrete time and

$$\frac{(i\lambda)^k e^{i\lambda t}}{Q}; \qquad k = 0, \ldots, r-1 \tag{9'}$$

for the continuous time where r is a maximal degree of the polynomial P, Q (see, for example [6]).

The explicit description of minimal splitting subspace for multidimensional processes with rational spectrum is still it seems an open problem.

Another open problem is concerning analysis of the Markov state representations by means of processes $x(t)$ with different innovations $u(t)$, $-\infty < t < \infty$, i.e. such that

$$H^{-t}(x) \supseteq H^{-t}(y).$$

(Note that the innovation type and the richness of the past $H^{-t}(x)$ can be characterized completely by the inner factor in the corresponding factorization $f_{yy} = \phi_{yu} \cdot \phi_{yu}^*$).

---

[1]Cf. with [6] where similar result was obtained for univariate processes in a quite complicated analytical way.

II. Innovation Continuity. Let $x(t)$ be a random process on an interval of the real line and $\mathcal{C}_t$ be a complete σ-algebra generated by the variables $x(s)$, $s \leqslant t$. The σ-algebras $\mathcal{C}_t$ is growing when t is increasing. We consider the question on continuity of the $\mathcal{C}_t$ growth which is quite important for different approaches to the stochastic optimal control as well as for the general theory of random processes (see for example [7]).

Let $H_t = L^2(\mathcal{C}_t)$ be the subspace of all random variables h, $Eh^2 < \infty$, measurable with respect to the σ-algebra $\mathcal{C}_t$. It is convenient to treat H_t as the subspace in the Hilbert space of all random variables h, $Eh^2 < \infty$, with the inner product $E(h_1 \cdot h_2)$. Because of the obvious correspondence between $\mathcal{C}_t$ and H_t we consider mainly the family H_t treated as a function of t.

Let us set

$$H_{t-0} = \overline{\bigcup_{s<t} H_s}, \qquad H_{t+0} = \bigcap_{u>t} H_u$$

the former one means the closure of all subspaces H_s, $s<t$. We have

$$H_{t-0} \subseteq H_t \subseteq H_{t+0}$$

and it can be a gap between H_{t-0} and H_t as well as between H_t and H_{t+0}. Say it occurs if t is a fixed point of discontinuity of the random process $x(\cdot)$. So considering conditions for the family H_t to be continuous we assume that the process $x(t)$ in a metric phase-space R is stochastically continuous:

$$\lim_{s \to t} P\{\rho(x(s), x(t)) \geqslant \epsilon\} = 0 \tag{10}$$

for any $\epsilon > 0$; here $\rho(x_1, x_2)$ means the distance between points $x_1, x_2 \in R$.

One can verify that under this assumption

$$H_{t-0} = H_t \tag{11}$$

Let us remind that a probabilities distribution in a metric space is regular namely for any measurable set B

$$\inf_{F \subseteq B} P(B \setminus F) = 0$$

where inf is over all closed sets F, $F \subseteq B$. For any closed set F we have

$$\inf_{G \supseteq F} P(G \setminus F) = 0$$

where inf is over all open sets G, $G \supseteq F$, with boundaries δG of zero probability $P(\delta G) = 0$; say one can take the proper

$$G = \{x : \rho(x, F) < r\}, \quad r \to 0,$$

with the boundaries

$$\delta G \subseteq \{x : \rho(x,F)=r\}$$

which are disjoint for different r so $P(\delta G)=0$ except not more than a countable number of r. Thus any event $\{x(t)\in B\}$ can be approximated by a proper event $\{x(t)\in G\}$ with $P\{x(t)\in\delta G\}=0$. The event $\{x(t)\in G\}$ itself can be approximated by events $\{x(s)\in G\}$, $s<t$. Indeed

$$P\{x(t)\in G, x(s)\notin G\} \leqslant P\{x(t)\in G\cdot F\}+P\{\rho(x(t),x(s))\geqslant\epsilon\}\to 0$$

if we consequently take

$$F=\{x:\rho(x,R\cdot G)\geqslant\epsilon\}$$

and $s\to t-0$, $\epsilon\to 0$. Applying it to the other open set $G'=R\cdot(G\cup\delta G)$ with the same boundary $\delta G'=\delta G$ we obtain

$$P\{x(t)\notin G, x(s)\in G\}\leqslant P\{x(t)\in G', x(s)\notin G'\}\to 0.$$

Thus the σ-algebra $\mathfrak{A}_t$ generated by the events $\{x(s)\in B\}$, $s\leqslant t$, coincides with the σ-algebra $\mathfrak{A}_{t-0}$ generated by the events $\{x(s)\in B\}$, $s<t$.

We apply now the same arguments considering the right-continuity of the family H_t:

$$H_{t+0}=H_t. \tag{12}$$

Generally this property does not hold even for very smooth process $x(t)$; moreover it can be arbitrary type of discontinuity (see for example [8]).

But it holds for a case of stochastically continuous processes with independent increments (which are more and more usable in martingale approach to the stochastic optimal control). Apart of the Wiener process case we don't know where this fact can be found though it looks like one of the classical results of the probabilities theory being the direct generalization of the famous 0-1 low.

Theorem 3. *For stochastically continuous process with independent increments*

$$H_{t-0}=H_t=H_{t+0}.$$

The proof of the theorem is based on the following

Lemma. *The orthogonal complement $H_u\ominus H_t$ in H_u to the subspace H_t, $t<u$, is a linear closure of variables*

$$h=h_t h_{t_1}(A_1)\cdots h_{t_n}(A_n) \tag{13}$$

where $h_t\in H_t$ and

$$h_{t_k}(A_k)=1_{A_k}-E\,1_{A_k},$$

1_{A_k} *are indicators of the events*

$$A_{k-}=\{x(t_k)-x(t_{k-1})\in B_k\};\quad t_0=t<t_1<\cdots<t_n\leqslant u.$$

Indeed the subspace H_u coincides with the linear closure of elements

$$h=1_{A_1}\cdots 1_{A_n}\in H_t$$

(where A_k are the events of the type $A_k=\{x(t_k)\in B_k\}$; $t_1<\cdots<t_n\leqslant t$) together with elements of the type (13) and the last are obviously orthogonal to the subspace H_t.

As it was actually shown in the case of stochastically continuous process by the proof of the equation (11) any event

$$A=\{x(t_1)-x(t)\in B\}$$

can be approximated by the event

$$A'=\{x(t_1)-x(t+\delta)\in B\},\qquad \delta\to+0,$$

and therefore any element h of the type (13) can be approximated by similar element h' which is obtained from h by the substitution of the second factor:

$$h_{t_1}(A_1)\to h_{t_1}(A').$$

Thus any element h_{t+0} from the orthogonal complement $H_{t+0}\ominus H_t$ in H_{t+0} to the subspace H_t as an element of the subspace $H_u\ominus H_t$ can be approximated by the linear combination of the proper elements h'. According to the lemma they belong to the subspaces $H_u\ominus H_{t+\delta}$ so they are orthogonal to $H_{t+0}\subseteq H_{t+\delta}$. Because the element h_{t+0} belongs to H_{t+0} we conclude that $h_{t+0}=0$ and thus $H_{t+0}=H_t$.

References

[1] R. E. KALMAN, Linear Stochastic Filtering Theory–Reappraisal and Outlook, in *Proceedings of Symposium on System Theory,* Polytechnic Institute of Brooklyn (1965), pp. 197–205.

[2] P. FAURRE, Réalisations Markoviennes de processus stationnaires, IRIA Report No. 13, Institut de Recherche d'Informatique et d'automatique, Le Chesnay, 1973.

[3] H. AKAIKE, Markovian Representation of Stochastic Processes by Canonical Variables, *SIAM J. on Control,* **13**, 1 (1975), 162.

[4] H. AKAIKE, Stochastic Theory of Minimal Realization, *IEEE Transactions on Automatic Control,* **19**, 6 (1974), 667.

[5] YU. A. ROZANOV, Stationary Random Processes, Moscow, Fismatgiz, 1963.

[6] I. A. IBRAGIMOV, YU. A. ROZANOV, Gaussian Random Processes, Moscow, Nauka, 1970.

[7] Y. LIPTZER, A. N. SCHIRYAEV, Statistics of Random Processes, Moscow, Nauka, 1974.

[8] YU. A. ROZANOV, Innovation Processes, Moscow, 1973.

(Received February 24, 1976)