KYBERNETIKA ČÍSLO 4, ROČNÍK 6/1970

Statistical Problems in Hilbert Spaces*

Application to Filtering Theory

ALAIN BENSOUSSAN

Some Engineering Problems lead to Estimation Problems in the case when the "observation" is not a true stochastic process. This is a motivation to look for a generalization of the ordinary Estimation Theory in the case of more general stochastic concepts. In this article, we show that it is possible to obtain some results in that way, for Linear Estimation Theory.

I. GENERAL DEFINITIONS AND CONCEPTS

The concept we shall use most frequently is that of Linear Random Functional. Let Φ be a separable Hilbert space, and Φ' the dual space of Φ.

Definition 1.1. A linear random functional is a linear mapping from Φ into a space of real random variables, denoted by Mes $[\omega, \nu; R]$, where ω is a sample space provided with a probability law ν.

If $\varphi_* \in \Phi'$, the corresponding real random variable is $Y_{\varphi_*}(\omega)$. A simple example of linear random functional on Φ' is the following: Let $Y(\omega)$ be a Φ-valued random variable, we set

$$Y_{\varphi_*}(\omega) = \langle Y(\omega), \varphi_* \rangle . \tag{1.1}$$

If Φ is a finite dimensional linear space, any linear random functional on Φ' will be of the shape (1.1). Of course, it is not always the case, if Φ is an infinite dimensional space. Thus, the concept of linear random functional on Φ' appears as a generalization of the concept of Φ-valued random variable.

Definition 1.2. We shall say that $Y_{\varphi_*}(\omega)$ has a mathematical expectation, if

$$\forall \varphi_* , \quad \mathsf{E} Y_{\varphi^*} < + \infty$$

* Paper presented at the Session of the Czechoslovak Cybernetic Society on November 12, 1969.

and $\varphi_* \to \mathsf{E}Y_{\varphi_*}$ is a continuous mapping from Φ' into R. In that case, there exists some $\bar{Y} \in \Phi$, such that

$$\mathsf{E}Y_{\varphi_*} = \langle \bar{Y}, \varphi_* \rangle, \quad \forall \varphi_* \in \Phi'. \tag{1.2}$$

The vector $\bar{Y}$ will be called the mathematical expectation of Y_{φ_*}.

Definition 1.3. We shall say that the linear random functional has a covariance operator $\Lambda \in \mathscr{L}(\Phi'; \Phi)$ if

$$\forall \varphi_*^1, \varphi_*^2 \in \Phi', \tag{1.3}$$

$$\mathrm{Cov}\,(Y_{\varphi_*^1}, Y_{\varphi_*^2}) = \mathsf{E}\{(Y_{\varphi_*^1} - \mathsf{E}Y_{\varphi_*^1})(Y_{\varphi_*^2} - \mathsf{E}Y_{\varphi_*^2})\} < +\infty$$

and $\varphi_*^1, \varphi_*^2 \to \mathrm{Cov}\,(Y_{\varphi_*^1}, Y_{\varphi_*^2})$ is continuous from $\Phi' \times \Phi' \to R$. In that case, there exists some $\Lambda \geqq 0$ and self adjoint, such that

$$\langle \Lambda\varphi_*^1, \varphi_*^2 \rangle = \mathrm{Cov}\,(Y_{\varphi_*^1}, Y_{\varphi_*^2}). \tag{1.4}$$

Example. *The white noise process*

Let be $\Phi = \Phi' = L^2(0, T)$. Let us consider a linear random functional $Y_\varphi(\omega)$ on Φ, with $\bar{Y} = 0$ and $\Lambda = 1$. We, thus have

$$\mathsf{E}(Y_{\varphi_1}, Y_{\varphi_2}) = \int_0^T \varphi_1(t)\, \varphi_2(t)\, \mathrm{d}t. \tag{1.5}$$

If, for each φ, Y_φ is a Gaussian random variable, the family Y_φ has the same properties as the family of stochastic integrals

$$\int_0^T \varphi(t)\, \mathrm{d}Z(t; \omega),$$

where $Z(t; \omega)$ is the Wiener process. It is well known that there does not exist a stochastic process $Y(t; \omega)$ such that

$$\int_0^T \varphi(t)\, \mathrm{d}Z(t; \omega) = \int_0^T \varphi(t)\, Y(t; \omega)\, \mathrm{d}t. \tag{1.6}$$

This means that, there does not exist a Φ-valued random variable $Y(\omega)$ such that

$$Y_\varphi(\omega) = (\varphi, Y(\omega)).$$

The linear random functional $Y_\varphi(\omega)$ is called the white noise process. In other words, the white noise process is not a Φ-valued random variable.

Definition 1.4. *Image of a linear random functional by an affine continuous mapping*

Let Φ, Ψ be two Hilbert spaces and

$$g(\varphi) = \alpha\varphi + m$$

an affine continuous mapping from Φ into Ψ. The mapping α belongs to $\mathscr{L}(\Phi, \Psi)$, and m is an element of Ψ. Let $Y_{\varphi_*}(\omega)$ a linear random functional on Φ', we define a linear random functional on Ψ' by

$$Z_{\psi_*}(\omega) = Y_{\alpha^*\psi_*}(\omega) + \langle m, \psi_* \rangle\,, \tag{1.7}$$

where α^* is the adjoint operator of α.

If $Y_{\varphi_*}(\omega)$ has a mathematical expectation $\bar{Y}$ and a covariance operator Λ, then $Z_{\psi_*}(\omega)$ has the same properties, and one has

$$\bar{Z} = \alpha\bar{Y} + m\,, \tag{1.8}$$

$$\mathrm{Cov}\,(Z) = \alpha\Lambda\alpha^*\,. \tag{1.9}$$

II. LINEAR ESTIMATION THEORY

Let Φ and Ψ be two Hilbert spaces. We consider a linear random functional on $\Phi' \times \Psi'$, denoted by $T_{\{\varphi_*,\psi_*\}}(\omega)$, having a mathematical expectation $(\bar{X}, \bar{Y}) \in \Phi \times \times \Psi$ and a covariance operator $Q \in \mathscr{L}(\Phi' \times \Psi'; \Phi \times \Psi)$. The covariance operator Q can be written as an operator matrix

$$Q = \begin{pmatrix} Q_{11}, & Q_{12}^* \\ Q_{12}, & Q_{22} \end{pmatrix}. \tag{2.1}$$

We set

$$X_{\varphi_*}(\omega) = T_{\{\varphi_*,0\}}(\omega) \tag{2.2}$$

and

$$Y_{\psi_*}(\omega) = T_{\{0,\psi_*\}}(\omega) \tag{2.3}$$

and therefore

$$T_{\{\varphi_*,\psi_*\}}(\omega) = X_{\varphi_*}(\omega) + Y_{\psi_*}(\omega)\,. \tag{2.4}$$

We want to give a meaning to the sentence "Estimate X_{φ_*} by Y_{ψ_*}" in such a way that, in the case when X_{φ_*} and Y_{ψ_*} are true random variables, the definition will be consistent with the ordinary one.

Definition 2.1. An estimate of X_{φ_*} will be an affine continuous mapping from Ψ into Φ, of the following shape

$$Y \to \hat{X}_\Lambda = \bar{X} + \Lambda(Y - \bar{Y}) \tag{2.5}$$

where Λ belongs to $\mathscr{L}(\Psi; \Phi)$. The estimate $\hat{X}_\Lambda$ depends only on the choice of Λ.

To $\hat{X}_\Lambda$, we shall associate the estimation error, which is an affine continuous mapping from $\Phi \times \Psi$ into Φ, defined by

$$(X, Y) \to \varepsilon_\Lambda = X - \bar{X} - \Lambda(Y - \bar{Y}) . \tag{2.6}$$

In the case when we have random variables $X(\omega)$, $Y(\omega)$ on Φ and Ψ, which are, respectively the variable to be estimated, and the observation, the formulas (2.5) and (2.6) define an estimate $\hat{X}_\Lambda(\omega)$ and an estimation error $\varepsilon_\Lambda(\omega)$. Of course, in that case, we get the ordinary definition. However, the definition (2.5) and (2.6) can easily be extended to the case when we have, not true random variables on Φ and Ψ, but linear random functionals X_{φ_*} and Y_{ψ_*} on Φ' and Ψ'. In that case, the estimate $\hat{X}_\Lambda$ will be a linear random functional on Φ', obtained by taking the image of Y_{ψ_*} by the mapping (2.5). In the same manner the estimation error will be a linear random functional on Φ', obtained by taking the image of $(X_{\varphi_*}, Y_{\psi_*})$ by the mapping (2.6).

Now, we would like to choose a "best" estimate. Our decision variable is Λ, and we have to define a criterion. We can first notice that the linear random functional "Estimation error" has a covariance operator Γ_Λ, which is given by

$$\langle \Gamma_\Lambda \varphi_*^1, \varphi_*^2 \rangle = \left\langle Q \begin{pmatrix} \varphi_*^1 \\ -\Lambda^* \varphi_*^1 \end{pmatrix}, \begin{pmatrix} \varphi_*^2 \\ -\Lambda^* \varphi_*^2 \end{pmatrix} \right\rangle , \quad \forall \varphi_*^1, \varphi_*^2 \in \Phi . \tag{2.7}$$

We, then, have the

Proposition 2.1. *Let us assume that* Q_{22} *is invertible, then we have*

$$\Gamma_\Lambda \geqq \Gamma_{\Lambda_0} \tag{2.8}$$

(*in the sense of the positive self adjoint operators order relationship*) *where*

$$\Lambda_0 = Q_{12}^* Q_{22}^{-1} , \tag{2.9}$$

and if $\Lambda \neq \Lambda_0$, *then*

$$\Gamma_\Lambda > \Gamma_{\Lambda_0}$$

(*in the sense that there exists at least one* φ *such that*

$$(\Gamma_\Lambda \varphi, \varphi) > (\Gamma_{\Lambda_0} \varphi, \varphi)) .$$

Remark 2.1. Since we have assumed that Q_{22} is invertible, it is impossible to assume at the same time that Y_{ψ_*} is a true random variable, unless Ψ is a finite dimensional linear space.

Let us consider the two relationships

$$\frac{dy}{dt} + A(t)\, y = f(t) + B(t)\, \xi(t)\,, \tag{3.1}$$

$$y(0) = y_0 + \zeta\,,$$

$$z(t) = C(t)\, y(t) + \eta(t) \tag{3.2}$$

where

$$t \in [0, T]\,, \quad A(.) \in L^\infty(0, T; \mathscr{L}(R^n; R^n))\,,$$

$$B(.) \in L^\infty(0, T; \mathscr{L}(R^m; R^n))\,, \quad C(.) \in L^\infty(0, T; \mathscr{L}(R^n; R^p))\,,$$

$$f(.) \in L^2(0, T; R^n)\,, \quad y_0 \in R^n\,.$$

We suppose that the triplet

$$\{\zeta, \xi(.), \eta(.)\}$$

belongs to the Hilbert space

$$\Phi = R^n \times L^2(0, T; R^m) \times L^2(0, T; R^p)\,. \tag{3.3}$$

We shall say that (3.1) defines the dynamics of a physical system, and (3.2) is the observation available on the system; $\xi(.)$ is an input, and ζ, $\eta(.)$ are disturbances. In the completely deterministic case, the disturbances are fixed elements, and thus, the choice of an element μ of Φ defines completely $y(t)$ at each time t, and $z(.)$. Let γ_t be the family of operators from $\Phi \to R^n$ defined by

$$\gamma_t(\mu) = y(t) \tag{3.4}$$

(the value of the continuous function y at time t) and δ the mapping from $\Phi \to$ $\to L^2(0, T; R^p)$ defined by

$$\delta(\mu) = z(.)\,. \tag{3.5}$$

To define a stochastic system analogous to (3.1), (3.2), we shall consider that the triplet μ is no longer a given element of Φ, but a linear random functional $\mu_\varphi(\omega)$ on $\Phi = \Phi'$. Since the mappings γ_t and δ are affine continuous, we also can define the state of the system at time t, as the image $\gamma_t(\mu_\varphi)$, and the observation as $\delta(\mu_\varphi)$. Thus the state of the system at time t is a linear random functional on R^n, and the observation is a linear random functional on $L^2(0, T; R^p)$. Since R^n is finite dimensional, the state of the system at time t is a R^n-valued random variable. Therefore, there exists a stochastic process $Y(t; \omega)$, with values in R^n, such that

$$\gamma_t(\mu_\varphi)_\psi\,(\omega) = (Y(t; \omega), \psi)\,, \quad \forall \psi \in R^n, \text{a.s. } \omega\,. \tag{3.6}$$

Of course, we cannot say a priori such a thing for the observation, since the space $L^2(0, T; R^p)$ is an infinite dimensional space. On the contrary, in the following, the observation will not be a random variable, with values in $L^2(0, T; R^p)$.

We now shall precise the properties of μ:

$$\mathsf{E}\{\mu\} = 0\,, \tag{3.7}$$

$$\operatorname{Cov}\{\mu\} = \begin{pmatrix} P_0, & 0, & 0 \\ 0, & Q(t), & 0 \\ 0, & 0, & R(t) \end{pmatrix} \tag{3.8}$$

with the assumption $R(t)$ invertible, $\forall t$.

The filtering problem is to estimate $Y(T)$ by $z(.)$. It is not difficult to convince oneself that $z(.)$ cannot be a true random variable in $L^2(0, T; R^p)$, and therefore we shall use our generalized linear estimation theory. Since $R(t)$ is invertible, the assumptions of Proposition 2.1 are satisfied, and therefore the best estimate exists and is given through (2.9).

It is possible to compute explicitly the operator Λ_0 of (2.9) (cf. [1]), but we shall restrict ourselves to a subclass of estimates*, denoted by the "Filter" Class. Let

$$K(.) \in L^\infty(0, T; \mathscr{L}(R^p; R^n))\,,$$

we associate to $K(.)$ the solution of the differential equation

$$\frac{\mathrm{d}e}{\mathrm{d}t} + A(t)\, e = f(t) + K(t)\,(z(t) - C(t)\, e(t))\,, \qquad e(0) = y_0 \tag{3.9}$$

where

$$z(.) \in L^2(0, T; R^p)\,.$$

We shall say that $e(t)$ is a filter and K is the gain of the filter.

The fact that the filters are recursive, make them the most interesting among all estimators. Although it is possible to show that the best estimator is indeed a filter, it is of interest to restrict oneself to that class. Let us set

$$K_0(t) = P(t)\, C^*(t)\, R^{-1}(t) \tag{3.10}$$

where $P(t)$ is the solution of the Riccati equation

$$\frac{\mathrm{d}P}{\mathrm{d}t} + AP + PA^* + PC^*R^{-1}\,CP = BQB^*\,, \qquad P(0) = P_0\,, \tag{3.11}$$

* To which the best estimate will belong anyway.

 we have

Theorem 3.1. *There exists a unique solution* $P(t)$ *of* (3.11), *on* $[0, T]$, $P(.) \in L^\infty(0, T; \mathscr{L}(R^n; R^n))$, *self adjoint positive. If* $\Lambda_K(t)$ *denotes the covariance of the estimation error* $e_K(t) - y(t)$, *at time* t, *we have*

$$\Lambda_K(t) \geqq \Lambda_{K_0}(t), \quad \forall K, \forall t, \tag{3.12}$$

and if $K \neq K_0$ *on* $]0, t[$,

$$\Lambda_K(t) > \Lambda_{K_0}(t), \quad \forall t.$$

Proof. The proof of the existence and uniqueness of the solution of the Riccati equation can be found in [2]. It is useful to remark that if one considers the system

$$\begin{aligned} &\frac{d\beta}{dt} + A(t)\,\beta + B(t)\,Q(t)\,B^*(t)\,\gamma = 0, \\ &-\frac{d\gamma}{dt} + A^*(t)\,\gamma - C^*(t)\,R^{-1}(t)\,C(t)\,\beta = 0, \\ &\beta(0) = -P_0\,\gamma(0), \\ &\gamma(T) = h, \end{aligned} \tag{3.13}$$

then, (3.13) has a unique solution and

$$\beta(t) = -P(t)\,\gamma(t), \quad \forall t. \tag{3.14}$$

Let us set

$$\varepsilon_K(t) = e_K(t) - y(t). \tag{3.15}$$

We have

$$\begin{aligned} &\frac{d\varepsilon_K}{dt} + A(t)\,\varepsilon_K + K(t)\,C(t)\,\varepsilon_K = -B(t)\,\xi(t) + K(t)\,\eta(t), \\ &\varepsilon_K(0) = -\zeta. \end{aligned} \tag{3.16}$$

The covariance operator $\Lambda_K(t)$ of $\varepsilon_K(t)$, is the solution of

$$\begin{aligned} &\frac{d\Lambda_K}{dt} + A(t)\,\Lambda_K + \Lambda_K\,A^*(t) + K(t)\,C(t)\,\Lambda_K + \Lambda_K\,C^*(t)\,K^*(t) = \\ &= B(t)\,Q(t)\,B^*(t) + K(t)\,R(t)\,K^*(t), \\ &\Lambda_K(0) = P_0 \end{aligned} \tag{3.17}$$

which can be rewritten as

$$\frac{d\Lambda_K}{dt} + A\Lambda_K + \Lambda_K A^* + \Lambda_K C^* R^{-1} C \Lambda_K = BQB^* + \Sigma_K, \qquad (3.18)$$

$$\Lambda_K(0) = P_0$$

where

$$\Sigma_K(t) = (K(t) - \Lambda_K(t)\, C^*(t)\, R^{-1}(t))\, R(t)\, (K^*(t) - R^{-1}(t)\, C(t)\, \Lambda_K(t))\,. \qquad (3.19)$$

The form of (3.18), analogous to (3.11), allows us to write

$$\frac{d\beta_1}{dt} + A(t)\, \beta_1 + (B(t)\, Q(t)\, B^*(t) + \Sigma_K(t))\, \gamma_1 = 0\,, \qquad (3.20)$$

$$-\frac{d\gamma_1}{dt} + A^*(t)\, \gamma_1 - C^*(t)\, R^{-1}(t)\, C(t)\, \beta_1 = 0\,,$$

$$\beta_1(0) = -P_0\, \gamma_1(0)\,,$$

$$\gamma_1(T) = h\,,$$

and

$$\beta_1(t) = -\Lambda_K(t)\, \gamma_1(t)\,. \qquad (3.21)$$

We then have

$$(P(T)\, h, h) = (P_0\, \gamma(0), \gamma(0)) + \int_0^T (B(t)\, Q(t)\, B^*(t)\, \gamma(t), \gamma(t))\, dt + \qquad (3.22)$$

$$+ \int_0^T (C^*(t)\, R^{-1}(t)\, C(t)\, P(t)\, \gamma(t), P(t)\, \gamma(t))\, dt =$$

$$= (P_0\, \gamma(0), \gamma_1(0)) + \int_0^T (BQB^*\gamma, \gamma_1)\, dt + \int_0^T (C^* R^{-1} C P\gamma, \Lambda_K \gamma_1)\, dt$$

and

$$(\Lambda_K(T)\, h, h) = (P_0\, \gamma_1(0), \gamma_1(0)) + \int_0^T (BQB^*\gamma_1, \gamma_1)\, dt + \qquad (3.23)$$

$$+ \int_0^T (C^* R^{-1} C \Lambda_K \gamma_1, \Lambda_K \gamma_1)\, dt + \int_0^T (\Sigma_K \gamma_1, \gamma_1)\, dt\,.$$

Therefore if ϱ is defined by

$$\varrho = (P_0(\gamma(0) - \gamma_1(0)), \gamma(0) - \gamma_1(0)) + \int_0^T (BQB^*(\gamma - \gamma_1), \gamma - \gamma_1)\, dt + \qquad (3.24)$$

$$+ \int_0^T (C^* R^{-1}\, C(\Lambda_K \gamma_1 - P\gamma), \Lambda_K \gamma_1 - P\gamma)\, dt + \int_0^T (\Sigma_K \gamma_1, \gamma_1)\, dt$$

then

$$\varrho = ((\Lambda_K(T) - P(T))\, h, h) \geqq 0\,.$$

Thus

$$\Lambda_K(T) \geqq P(T) = \Lambda_{K_0}(T)\,.$$

Now, if

$$\Lambda_K(T) = P(T)$$

we have $\varrho = 0 \Rightarrow BQ\, B^*(\gamma - \gamma_1) = 0$, $P_0(\gamma(0) - \gamma_1(0)) = 0$, $C^*R^{-1}\, C(\beta_1 - \beta) = 0$. Therefore

$$\gamma = \gamma_1$$

and $\Sigma_K \gamma = 0$. Therefore $\beta = \beta_1$, and (3.20) and (3.13) are the same systems.

Let then be $t < T$, we have

$$\gamma_1(t) = \gamma(t)\,,$$

whatever be $h \in H$. Therefore since $\beta_1(t) = \beta(t)$, we get

$$\Lambda_K(t)\, \gamma_1(t) = P(t)\, \gamma_1(t)\,. \tag{3.25}$$

Now, it is clear that $\gamma_1(t)$ satisfies the equation

$$-\frac{d\gamma}{dt} + (A^*(t) + \Lambda_K(t)\, C^*(t)\, R^{-1}(t)\, C(t))\, \gamma_1 = 0\,, \tag{3.26}$$

$$\gamma_1(T) = h\,.$$

But, when h varies in H, $\gamma_1(t)$ varies in a dense subspace of H (because γ_1 is the solution of a linear parabolic equation). Therefore, according to (3.25), $\Lambda_K(t)$ and $P(t)$ are identical on a dense subspace of H, and consequently are identical everywhere. Thus, we get

$$\Lambda_K(t) = P(t)\,, \quad \forall t \in [0, T]\,. \tag{3.27}$$

If one then compares the Riccati equation (3.18) and (3.11), one gets

$$\Sigma_K(t) = 0 \quad \text{a.e.} \quad t \in]0, T[\tag{3.28}$$

and therefore

$$K(t) = K_0(t) \quad \text{a.e. on} \quad]0, T[\,. \tag{3.29}$$

It follows that if $K \neq K_0$ on a non zero Lebesgue measure subset of $]0, T[$, then

$$\Lambda_K(T) > P(T)\,. \tag{3.30}$$

In fact, the argument does not depend on the particular value of T, which proves the last part of the theorem.

(Received December 10, 1969.)

REFERENCES

[1] A. Bensoussan: Filtre de Wiener et Filtre de Kalman Bucy pour les équations aux dérivées partielles. I. R. I. A. report, to appear.

[2] R. E. Kalman, R. S. Bucy: New results in linear filtering and prediction theory. Journal of Basic Engineering (1961), 95—107.

VÝTAH

Statistické problémy v Hilbertových prostorech Aplikace na teorii filtrace

ALAIN BENSOUSSAN

Nejprve je zaveden pojem „lineárního náhodného funkcionálu", což je zobecnění pojmu „náhodné veličiny" s hodnotami v Hilbertově prostoru. Účelem tohoto pojmu je obsáhnout nejen obyčejné stochastické procesy, nýbrž také procesy typu bílého šumu, a to jak vzhledem k časovým tak i fázovým proměnným. Dále je pro tuto třídu pojmů zobecněna obyčejná teorie odhadu. Ukazuje se, že pro konečně rozměrné prostory není tohoto zobecnění třeba, jeho nutnost se projevuje v případě nekonečně rozměrných prostorů.

Tato teorie je poté aplikována na Kalman-Bucyho filtrační problém pro obyčejné diferenciální rovnice. Teorie je pro tento problém velmi vhodná a poskytuje prostředek k odvození známého rekurzivního filtru a to způsobem, který lze snadno rozšířit na parciální diferenciální rovnice.

Prof. Alain Bensoussan, Institut de recherche d'informatique et d'automatique, Domaine de Voluceau, 78 Rocquencourt, France.