Innovation Theory の歴史 II

Si Si
愛知県立大学
情報科学部

Abstract

The concept of innovation can be seen in P. Lévy's book although he did not express the word innovation. Following Lévy, many authors including N. Wiener, T. Hida, T. Kailath have done a lot of contributions in different aspects. In this note some of them are discussed from historical view point.

1 Initial Construction of Innovation

P. Lévy initiated the necessary and sufficient concept to be able to construct the innovation in his literture :

Théorie de l'addition des variables aléatoires. Gauthier-Villars. **1937**. *2ème éd. 1954.*

Although it is a strong condition, it seems being simple for the case of discrete parameter case but it is difficult to handle without continuity condition for the parameter.

For continuous parameter case, the infinitesimal equation

$$\delta X(t) = \Phi(X(s), s \leq t; Y(t), t, dt).$$

is proposed in his paper :

Random functions: general theory with special reference to Laplacian random functions.I, (1953), 331-388. Univ. of Calif. Pub.

There $Y(t)$ is innovation, which express the new information obtained by $X(t)$ in a small time interval $[t, t + \delta t)$.

Stationary second order processes

Kailath [14] mentioned that Innovation approach was firstly used by A. N. Kolmogorov dealing with Linear least-squares prediction problem for stationary stochastic processes in the following paper :
A. N. Kolmogorov, "Stationary sequences in Hilbert spaces", *Bull. Math. Univ. Moscow* (in Russian), vol 2, no.6, p 40, 1941.

For continuous time parameter, Bode and Shannon (1950) and Zadeh and Ragazzinni (1950) rediscovered the technique independently, known as *Whitening-filter method.*

Detection Theory

Whitening-filter technique was proposed by Kotel'nikov (1959, Ph.D desertation presented in 1947) as a device for colored-noise problems by reduction to the white-noise case.

Non-stationary second order processes

Kalman made first really successful application to non-stationary second order processes in his 1960 paper:
A new approach to linear filtering and prediction problems, Trans. ASME.J. Basic Egrg. vol. 82. pp 33-34, 1960,
but there were some dificulties in extending to continuous time Gauss-Markov process.

2 Wiener's construction of the innovation

Wiener used the idea of innovation to define a non-linear analogue of the Wold decomposition for a purely non-deterministic process.

Wold's decomposition

A weakly stationary process $X(= (X_n))$ can be uniquely expressed as the sum of two weakly stationary processes X' and X'' so that X' and X'' are orthogonal (i.e. $X'_m \perp X''_n$ for all m and n, X' is purely non-deterministic and X'' is deterministic.

For nonlinear prediction, it is assumed that X is strictly stationary. Let

$$\mathbf{B}_n(X) = \sigma(X_j, j \leq n).$$

[W.1] ξ_0 is $\mathbf{B}_0(X)$ measurable, uniformly distributed in $[0, 1]$ and independent of X_k, $k < 0$.

[W.2] $\mathbf{B}_0(X) = \sigma(\xi_0, X_{-1}, X_{-2}, \cdots)$

[W.3] $\xi_n = T^n \xi_0$, where T is the shift operator.

If [W.1], [W.2], [W.3] holds, $\{\xi_n\}$ is a sequence of independent random variables, uniformly distributed in [0,1] so that ξ_n is $\mathbf{B}_n(X)$ measurable for each n.

In other words, for some Borel function g not depending on n we can write

[W.4] $\xi_n = g(X_n, X_{n-1}, \cdots)$.

If for each n, we are able to invert W.4, i.e.

[W.5] $X_n = h(\xi_n, \xi_{n-1}, \cdots)$, h is a Borel function.

then ξ_n pocessing the properties [W.1] to [W.3] are called *innovations* of the given process (X_n).

Information theoretically speaking, *coding* and *decoding* can be expressed as follows.

Coding

Let X_n be a descrete set of messages and $\{\xi_n\}$ be a coded set of completely independent messages. Then we can write

$$X_n \quad \longrightarrow \quad g \quad \longrightarrow \quad \xi_n$$

Decoding

The recovery of X_n from the channel [W.5] is a decoding problem.

$$\xi_n \quad \longrightarrow \quad h \quad \longrightarrow \quad X_n$$

Wiener's Construction

Wiener started with the following two hypotheses. (The same thing was done by Lévy [18], in 1937.)

Hypotheses

(1) $\{X_n\}$ is purely non-deterministic. Consider the conditional distribution

$$\begin{aligned} G(a_n|a_{n-1},a_{n-2},...) &= P[X_n(\omega)\le a_n|X_j(\omega)=a_j, j\le n-1] \\ &= P[X_n(\omega)\le a_n|B_{n-1}(X)](a_{n-1},a_{n-2},...) \end{aligned}$$

(2) Assume that for almost all $(a_{n-1},a_{n-2},...), G(a_n|(a_{n-1},a_{n-2},...)$ is a strictly increasing continuous function of a_n.

Define

[W.6] $\xi_n = G(X_n|X_{n-1},X_{n-2},\cdots)$.

Then ξ_n is uniformly distributed over [0,1] and independent of the past $\mathbf{B}_{n-1}(X)$. The random variables ξ_n are independent and

[W.7] $\mathbf{B}_n(X)=\sigma\{\xi_n,X_{n-1},X_{n-2},\cdots\}$.

Then it follows that for every integer $N\ge 1$,

[W.8] $\mathbf{B}_n(X)=\sigma\{\xi_n,\xi_{n-1},\cdots,\xi_{n-N},X_{n-N-1},\cdots\}$.

Consequently, by taking $N\to\infty$ in W.8

[W.9] $\mathbf{B}_n(X)=\mathbf{B}_n(\xi)$

in Kallianpur and Wiener [16]. Thus [W.5] follows from [W.9].

3 Kailath's Innovation Processes

In Kailath paper *innovation process* of a given stochastic process $y(\cdot)$ is defined as a Gaussian white noise $v(\cdot)$, which is related to $y(\cdot)$ by a causal and causally invertible transformation (i.e. canonical representation)

$$y(t)=\int_0^t F(t,u)v(u)du \tag{3.1}$$

exists then $v(\cdot)$is called Innovation process.

(Lévy 1950's papers)

3.1 Innovations for a class of Gaussian processes and some applications

The given observation is of the form

$$y(t) = z(t) + w(t), \qquad 0 \le t \le T < \infty$$

where $w(\cdot)$ is a sample function of Gaussian white noise such that

$$E[w(t)] = 0, \quad E[w(t)w(s)] = \delta(t-s)$$

and $z(\cdot)$, which is often called the signal process, is a Gaussian and such that

$$E[z(s)w(t)] \equiv 0, \text{for } s < t.$$

Note that $z(t)$ may depend upon past $\{w(s), 0 \le s < t\}$, but future $w(\cdot)$ must be independent of past $w(\cdot)$.

Define

$$E[z(t)z'(s) + z(t)w'(s) + w(t)z'(s)] = K(t,s), \quad 0 \le t, s \le T$$

Assume that

i) $K(t,s)$ is continuous in (t,s),

ii) $\int_0^T tr[K(t,t)]dt < \infty$. and

iii) $E[z(t)] = 0, \quad 0 \le t \le T.$

Define $\hat{z}(t)$ be the least-squares estimate of $z(t)$ given by $\{y(s),\ 0 \le t \le T\}$. That is, $\hat{z}(t)$ be the functional of past $y(\cdot)$ which yields $\text{tr}\{E[z(t) - \hat{z}(t)][z(t) - \hat{z}(t)]'\}$ is minimum.

$$\hat{z}(t) - z(t) \perp \{y(s),\ 0 \le t \le T.\}$$

If we can write

$$\hat{z}(t) = \int_0^t h(t,s)y(s)ds$$

then the Volterra kernel $h(\cdot,\cdot)$ must satisfy

$$h(t,s) + \int_0^t h(t,\tau)K(\tau,s)d\tau = K(t,s) 0 \le s, t \le T,$$

which is called Wiener-Hopf equation for $h(t,s)$.

Theorem K.1 Let $y(\cdot), w(\cdot), z(\cdot)$ and $\hat{z}(t)$ be defined above. Then the process

$$v(t) = y(t) - \hat{z}(t) = \tilde{z}(t) + w(t), \quad 0 \leq t \leq T$$

is the innovation process of $y(t)$. i.e. it is a *white noise Gaussian process* with

$$E[v(t)] = o, \quad E[v(t)v(s)] = \delta(t-s)$$

and $v(\cdot)$ is related to $y(\cdot)$ by a causal and causally invertible transformation. In fact, in operator notation

$$\begin{aligned} v &= (I-h)y, \\ y &= (I-h)^{-1}v. \end{aligned}$$

Application I. The Kalman-Bucy Formulas

In this problem, the signal process $z(\cdot)$ is modeled as the output of a dynamical system driven by white noise.

Kalman and Bucy derived a differential equation for the estimate $\hat{x}(\cdot)$, of the states for the model

$$\dot{x}(t) = F(t)x(t) + G(t)u(t), \ E[u(t)u'(s)] = Q(t)\delta(t-s),$$

$$z(t) = H(t)x(t), E[x(0)] = 0, \ E[x(0)x'(0)] = P_0,$$

$$E[u(t)x'(0)] \equiv 0, \ t \geq 0.$$

The observations are

$$y(t) = z(t) + w(t), \ E[w(t)w'(s)] = I\delta(t-s),$$

$$E[w(t)u'(s)] = 0.$$

The state vector $x(\cdot)$ may represent the position and velocity (after linearization) of a satellite and the forcing term $u(\cdot)$ may represent random drag.

The estimate of $z(\cdot)$ can then be obtained by linearity as

$$\hat{z}(t) = H(t)\hat{x}(t).$$

By innovation approach, Kalman and Bucy Formulas can be obtained. First, replace the given observation process $y(\cdot)$ by its innovation process $v(\cdot)$ which is given by the above theorem such that

$$v(t) = y(t) - \hat{z}(t) = y(t) - H(t)\hat{x}(t)$$

then the desired estimate $\hat{x}(t)$ is calculated as a linear functional of the innovations

$$\hat{x}(t) = \int_0^t g(t,s)v(s)ds,$$

where $g(t,s)$ is obtained by the orthogonality condition such that

$$g(t,s) = E[x(t)v'(s)], \quad 0 \le s < t.$$

Thus we have

$$\hat{x}(t) = \int_0^t E[x(t)v'(s)][y(s) - H(s)\hat{x}(s)]ds.$$

Differentiate $\hat{x}(t)$ with respect to t, and use the given facts of $u(\cdot), x(0)$ and $v(\cdot)$, Kalman and Bucy's differential equation is obtained as

$$\frac{d}{dt}\hat{x}(t) = F(t)\hat{x}(t) + K(t)[y(t) - H(t)\hat{x}(t)], \quad \hat{x}() = 0,$$

where

$$K(t) = E[x(t)v'(t)].$$

Application II. Prediction in linear difference system

Apply the innovation method to solve the following difference equation:

$$\begin{aligned} y(i) &= a_1(i)y(i-1) + \cdots + a_N(i)y(i-N) + w(i), \ i \ge N \\ y(i) &= w(i), \quad 0 \le i \le N-1, \end{aligned}$$

where $w(i)$ is a colored-noise sequence with $E[w(i)] = 0, E[w(i)w'(j)] = r_{ij}$.

The problem is to calculate $\hat{y}(i|i-1)$, the least squares linear estimate of $y(i)$, given $\{y(j),\ 0 \le j \le i-1\}$.

First determine the innovation for $\{y(\cdot)\}$ which is

$$v(i) = \frac{\epsilon(i)}{\sqrt{\epsilon^2(i)}}, \quad \epsilon(i) = y(i) - \hat{y}(i|i-1).$$

Application III. Detection of Gaussian signals

Consider a detection problem with hypotheses of the form

$$\begin{aligned} h_i &: \ y(t) = z(t) + w(t), \ 0 \le t \le T \\ h_0 &: \ y(t) = w(t), \ 0 \le t \le T \end{aligned}$$

where $w(\cdot)$ is white noise with mean zero , covariance $E[w(t)w(s)] = \delta(t-s)$ and $\dot{z}(t)$ is a Gaussian process satisfying with mean $m(t)$ and finite variance, $E[w(t)w(s)] \equiv 0, t > s$.

Theorem K.2 The above detection problem is nonsingular and the likelihood ratio can be written

$$LR = \exp \int_0^T \hat{z}_1(t)y(t)dt - \frac{1}{2}\int_0^T \hat{z}_1^2(t)dt$$

where $\hat{z}_1(t) = E[z(t)|\{y(s),\ 0 \le s \le t\}, h_1]$, the causal least squares estimate of $z(t)$ given past $y(\cdot)$ assuming h_1 is true.

Theorem K.3 A Gaussian process $y(\cdot)$ is equivalent to a white noise.

i.e. the problem of deciding whether an observation is a sample function of $y(\cdot)$ or of the white noise is nonsingular, iff it can be expressed as

$$y(t) = z(t) + w(t)$$

where $z(\cdot)$ and $w(\cdot)$ obey the conditions defined in Theorem K.2.

3.2 Innovations for a class of non-Gaussian processes and some applications

Consider the process

$$y(t) = z(t) + w(t)$$

where $w(\cdot)$ is white noise of unit intensity and $\dot{z}(\cdot)$ is not a necessarily Gaussian processn satisfying

i) $\int_0^T E[z^2(t)dt < \infty$.

ii) The white noise $w(t)$ is independent of past $\{w(s), z(s), 0 \le s < t\}$.

For example $z(\cdot)$ could be a Poisson process or a telegraph wave.

Let $\hat{z}(t)$ be the causal least-squares estimate of $z(t)$, not necessarily linear, of $z(t)$ given past $\{y(s),\ 0 \le s < t.]$

As in the Gaussian case, the new information or innovation process is

$$v(t) = y(t) - \hat{z}(t) = \tilde{z}(t) + w(t).$$

It could be proved that $w(\cdot)$ is a white noise process

$$E[w(t)w(s)] = \delta(t-s)$$

Theorem K.4 Under the above assumptions, the process

$$v(t) = y(t) - \hat{z}(t)$$

is a white noise with the same statistics as $w(\cdot)$.

Application IV. Non-Gaussian signals in white noise

Start with the hypotheses

$$\begin{aligned} h_1 &: \ y(t) = z(t) + w(t), \\ h_0 &: \ y(t) = w(t), \ 0 \leq t \leq T \end{aligned}$$

where $w(\cdot)$ is white noise with mean zero , covariance $E[w(t)w(s)] = \delta(t-s)$ and $\dot{z}(t)$ is a Gaussian process satisfying with mean $m(t)$ and finite variance, $E[w(t)z(s)] \equiv 0, t > s$.

Theorem K.5 The likelihood ratio (LR) for the above detection problem can be written as

$$LR = \exp \int_0^T \hat{z}_1(t)y(t)dt - \frac{1}{2}\int_0^T \hat{z}_1^2(t)dt$$

where $\hat{z}_1 t)$ = the causal least squares estimate of $z(t)$ given by $y(s), 0 \leq s < t$ and assuming h_1 is true.

Theorem K.6 A process $y(\cdot)$ is absolutely continuous (as defined above) w.r.t. a white noise, iff it can be expressed as

$$y(t) = z(t, \{y(s), 0 \leq s < t\}) + w(t)$$

where $w(\cdot)$ is white noise and $z(\cdot)$ is a functional of past $y(\cdot)$ that is square integrable on $[0, T]$.

3.3 Examples

Example 1. Stationary Gaussian processes with rational power Spectral densities

This class was treated by Bode and Shannon admitting a simple solution. Let

$$y(t) = \int_{-\infty}^{\infty} e^{it\lambda} dz(\lambda) \tag{3.2}$$

where

$$E[|dz(\lambda)|^2] = \frac{\lambda^2+4}{(\lambda^2+1)(\lambda^2+9)}d\lambda, \tag{3.3}$$

then the factorization gives the fourier transform of kernel function F, for the canonical representation, as

$$\hat{F}(\lambda) = \frac{i\lambda+2}{(i\lambda+1)(i\lambda+3)}$$

The problem of finding the canonical (innovation) representation is often refered as a spectral factorizing representation problem.

Example 2. Non-stationary Gaussian processes

Let $y(t)$ be a Wiener process i.e. Gaussian with mean zero and covariance

$$Ey(t)y(s) = t \wedge s, 0 \leq t, s, \leq T.$$

The canonical representation

$$y(t) = \int_0^t v(\tau)d\tau,$$

but there are many causal representations but not canonical, for instance (ref Lévy),

$$y(t) = \int_0^t \left(3 - \frac{12\tau}{t} + \frac{10\tau^2}{t^2}\right) v(\tau)d\tau.$$

Example 3. Gaussian Markov processes

Let $y(t)$ be a Gaussian Markov process with covariance $a(t \vee s)b(t \wedge s), s \leq t, s \leq T.b(t) > 0$, where $0 \leq t \leq T$. According to Doob, the representation is obtained as

$$y(t) = b(t)\int_0^t p(\tau)\dot{v}(\tau)d\tau,$$

where $p(\cdot)$ is defined by the relation

$$\frac{a(t)}{b(t)} = \int_0^t p^2(\tau)d\tau + \frac{a(0)}{b(0)}.$$

Example 4.

Let $B_i(t), i = 1, 2$, be two Brownian motions which are mutually independent.

$$X(t)=\begin{cases} B_1(t), & 0\le t\le 1\\ B_1(t), & t\ge 1 \text{ and rational}\\ B_2(t), & t\ge 1 \text{ and irrational}\end{cases}$$

$$X(t)=\int_0^t F_1(t,u)\dot{B}_1(u)du+\int_0^t F_2(t,u)\dot{B}_2(u)du$$

where for $t\le 1$

$$\begin{aligned} F_1(t,u) &= 1\\ F_2(t,u) &= 0, \end{aligned}$$

and for $t>1$

$$F_1(t,u)=\begin{cases}\chi_{[0,t]}, & t: \text{ rational}\\ 0, & t: \text{irrational}\end{cases}$$

and

$$F_2(t,u)=1-F_1(t,u).$$

It is possible to form a single Brownian motion as

$$Y(t)=\int_0^t F(t,u)\dot{B}(u)du.$$

$Y(t)$ has the same distribution with $X(t)$ but the representation is not canonical.

Concrete construction of a noncanonical representation of $X(t)$

$$\dot{B}(t)=\begin{cases}\frac{1}{\sqrt{2}}\dot{B}_1(2t-n), & t\in[n,n+\frac{1}{2}],\\ \frac{1}{\sqrt{2}}\dot{B}_2(2t-n), & t\in[n+\frac{1}{2},n+1],\end{cases}$$

where $n\ge 0$.
The kernel $F(t,u)$ is

$$F(t,u)=\begin{cases} 1, & t\in[0,\frac{1}{2}],\\ 0, & t\in[n+\frac{1}{2},n],\ t \text{ is rational},\ n\ge 0\\ 1, & t\in[n+\frac{1}{2},n],\ t \text{ is irrational},\ n\ge 0\\ 1, & t\in[n,n+\frac{1}{2}],\ t \text{ is rational},\ n\ge 1\\ 0, & t\in[n,n+\frac{1}{2}],\ t \text{ is irrational},\ n\ge 1.\end{cases}$$

We can see that

$$M_t(X) = M_t(B_1) \oplus \Big(M_t(B_2) \cap M_1(B_2)^{\perp}\Big)$$

while

$$M_t(X) \subset M_t(B).$$

4 Weak sense Innovation

This section is devoted to the work of H. Crammér (1961). In the proceeding of the fourth Berkeley symposium , H. Crammer gave the analogue of the Wold decomposition for the vector valued stochastic processes with finite second order moments .

He considered a q-dimensional stochastic vector processes

$$X(t) = (x^{(1)}(t), \cdots, x^{(q)}(t))$$

where $t \in T$.

Denote that

$$\begin{aligned} \mathbf{H}(X) &= span(x^{(j)}(t),\ 1 \le j \le q,\ t < \infty), && (4.1)\\ \mathbf{H}(X, t_0) &= span(x^{(j)}(t),\ 1 \le j \le q, \tau \le t_0). && (4.2)\end{aligned}$$

Discrete parameter case

Here the parameter space T is taken to be the set of all integers $n = 0. \pm 1, \pm 2, \cdots$. And $X(t)$ is to be written X_n.

Let $X_n = (x_n^{(1)}, \cdots, x_n^{(q)})$ be vector process with each component $x_n^{(j)}$ is a complex-valued stochastic process with discrete time parameter n, zero means and finite second order moments.

$$P_{n-1}X_n = (P_{n-1}x_n^{(1)}, \cdots, P_{n-1}x_n^{(q)})$$

Set

$$\xi_n = X_n - P_{n-1}X_n, \quad \xi_n^{(j)} = x_n^{(j)} - P_{n-1}x_n^{(j)}.$$

Thus the vector valued random variables

$$\xi_n = (\xi_n^{(1)}, \cdots, \xi_n^{(q)})$$

defines a vector valued stochastic processes with discrete time parameter n. The ξ_n process is called the innovation process to the given process X_n.

The set of values of n, for which the innovation does not reduce to zero, may be said to form the *innovation spectrum* of the X_n process. That is, *innovation spectrum* is the set containing precisely all those time points where a new impulse, or an innovation, enters into the process.

Remark. For a deterministic process, the innovation spectrum is evidently empty while for any nondeterministic process it has to contain atleast one value of n. For a nondeterministic stationary process, the innovation spectrum includes all integers.

Theorem (H.1) There is a unique decomposition of the X_n

$$X_n = U_n + V_n$$

satisfying

i) $U_n = (u_n^{(1)}, \cdots, u_n^{(q)})$ and $V_n = (v_n^{(1)}, \cdots, v_n^{(q)})$ where all $u_n^{(j)}$ and $v_n^{(j)}$ belong to $\mathbf{H}(X, n)$.

ii) The processes U_n and V_n are orthogonal and U_n is purely deterministic and V_n is nondeterministic.

In addition the processU_n can be expressed as a linear combination of those innovations ξ_i, of X_n, which have entered into the process before or at the inatant t,

$$U_n = \sum_{I=infty}^{n} A_{ni}\xi_i,$$

where the $A_{ni} = \{a_{ni}^{(jk)}\}$ are $q \times q$ matrices.

Continuous parameter case

In the paper [5], Crammer expressed in the following way.

I am indebted to Professor K. Ito for the observation that there are interesting points of contact of this section and a work by T. Hida on

"Canonical representations of Gaussian processes" which will appear in the Memoirs of the college of Science, University of Kyoto.

Here the parameter space T is taken to be the set of real numbers.

Theorem (H.2) There is a unique decomposition of the $X(t)$ process

$$X(t) = U(t) + V(t)$$

where $U(t) = (u^{(1)}(t), \cdots, u^{(q)}(t))$ and $V(t) = (v^{(1)}(t), \cdots, v^{(q)}(t))$ where all $u^{(j)}(t)$ and $v^{(j)}(t)$ belong to $\mathbf{H}(X)$ and $U(t)$ and $V(t)$ processes are orthogonal and $U(t)$ is purely deterministic and $V(t)$ is nondeterministic.

Then he dealt with vector process X(t) satifying

[C.1] $X(t)$ is purely nondeterministic, i.e. $\mathbf{H}(X, -\infty) = 0$,

[C.2] The limits $x^{(j)}(t-0)$ and $x^{(j)}(t+0)$ exist (as always in the $\mathbf{H}$ topology) for $1 \leq j \leq q$.

From [C.2], it follows that the space $\mathbf{H}(X)$ is sepaprable and the set of points of discontinuity of $X(t)$ is almost enumerable.

The innovation spectrum is defined as

$$\{t \mid \mathbf{H}(X, t+h) - \mathbf{H}(X, t-h) \neq 0, \text{ for any } h > 0\}.$$

The following assertion is ,as mentioned above, almost the same as is obtained by T. Hida independently at the same time.

Theorem (H.3) The vector variable $X(t)$ of any stochastic process satisfying conditions [C.1] and [C.2] can be expressed in the form

$$X(t) = \int_{-\infty}^{t} G(t, u) dZ(u)$$

where $Z(u)$ is an $N-$dimensional vector process with orthogonal increments, while $G(t, u)$ is a $q \times N$ matrix and where N is the multiplicity of $X(t)$.

References

[1] L. Accardi et al, Selected papers of Takeyuki Hida. World Scientific Pub. Co.Ltd. 2001

[2] L. Accardi, T. Hida and Si Si, Innovation approach to some stochastic processes and random fields. Volterra Center Notes. 2002.

[3] I. Antoniou,T. Hida, Si Siand Z. Suchanecki), Innovations and time operators ofstationary processes. Solvay Institute preprint

[4] H. Cramér,On some classes of nonstationary stochastic processes, Proc. of the fourth Berkeley Vol ". 1961, 57-78.

[5] H. Cramér, Half a century with probability theory: some personal recollections. The Ann. of Probability. 4 (1976) 509-546.

[6] T. Hida, Canonical representations of Gaussian processes and their applications. Memoires College of Sci. Univ. Kyoto, A. 33 (1960), 109-155.

[7] 飛田武幸，櫃田倍之、ガウス過程、紀伊国屋書店、1976.

[8] T. Hida and Si Si, An innovation approach to random fields. Application of white noise theory. World Scientific Pub. Co. Ltd. 2003.

[9] T. Hida and Si Si, Lévy field as a generator of elemental noises. Proc. Levico Conference on classical and quantum Lévy process. 2003, to appear.

[10] M. Hitsuda, Representation of Gaussian process equivalent to Wiener-Processes, Osaka J. Math. 5, 299-312.

[11] M. Hitsuda, Multiplicity of some classes of Gaussian processes, Nagoya-Math. J. 52, 39-46.

[12] I.M. Gel'fand and N. Ya. Vilenkin, Generalizede functions. vol.4.

[13] T. Kailath, An innovation approach to least-square estimation Part I, II, IEEE Transactions on Automatic Control Vol. AC-13, 1968, 646-654, 665-660.

[14] T. Kailath, "The innovation approach to detection and estimation theory" , *Pros. of the IEEE.* vol. 58. No.5 1970.

[15] G. Kallianpur, Some ramifications of Wiener's ideas on nonlinear prediction. Norbert Wiener: Collected Works, vol.III, The MIT Press, 1981, 402-424.

[16] G. Kallianpur and N. Wiener, Nonlinear prediction. Technical Report (CNR) 1956.

[17] K. Karhunen, Ueber die Struktur stationarer zufallige Funktionen. Arkiv f. Mat. 1(1950), 141-160.

[18] H. Kunita and S. Watanabe, On square integrable martingales, Nagoya Math. J. 30 (1967), 209-245.

[19] P. Lévy, Théorie de l'addition des variables aléatoires. Gauthier-Villars. 1937. 2ème éd.1954.

[20] P. Lévy, Processus stochastiques et mouvement brownien. Gauthier-Villars, 1948; 2ème ed. 1965.

[21] P. Lévy, Random functions: general theory with special reference to Laplacian random functions.I, (1953), 331-388. Univ. of Calif. Pub.

[22] P. Lévy, A Special problem of Brownian motion, and a general theory of Gaussian random functions. Proc. of the 3rd Berkeley Symp. on Math. Statist. and Probability, vol. II, 1955. 133-174.

[23] P. Lévy, Fonction aléatoire à corrélation linéaire. Illinois J. Math. 1 (1957), 217-258.

[24] R.S. Liptzer and A.N. Shiraev, Statistics of stochastic processes, Springer, Part I 1977,Part II 1978.

[25] P. Masani and N. Wiener, Non-linear prediction. Probability and Satatistics, The Harald Cramér volume. ed. U.Grenander, 1959, 190-212. Almqvist & Wiksell, John Wiley & Sons.

[26] P. Masani, Isomorphie entre les domains temporal et spectral d'un processus vectoriel, re'gulier, C.R. Acad Sci Paris496-498.

[27] M. Rosenblatt, Stationary processes as shifts of functions of independent random variables, J. Math. Mech. 8 (1959), 665-681.

[28] A. N. Shirjaev, Nonlinear filtering of Markov diffusion processes, Trudy Mat. Inst. Steklov, 104(1968), 135-180.

[29] N. Wiener, (with R.E.A.C. Paley), Fourier transforms in the complex domain. Amer. Math. Soc. 1934.

[30] N. Wiener, Extrapolation, interporation, and smoothing of stationary time series. The MIT Press, 1949

[31] N. Wiener, Nonlinear problems in random theory. The MIT Press, 1958.

[32] Nobert Wiener, Special issue of Bull. American Math. Soc. 72 (1966).

[33] Win Win Htay, Note on linear processes, Proc. Meijo Winter School, Jan 2003.