# Multiple Markov properties of Gaussian Processes and their control

Win Win Htay
Department of Engineering Mathematics
Yangon Technological University
Yangon, Myanmar

AMS Subject Classification 60H40 White Noise Theory
Key words: Gaussian process, Multiple Markov properties, White Noise, Entropy

**Abstract**

We shall first make a short survey of multiple Markov properties of Gaussian processes, then come to the most general definition of these properties, where we use the white noise theory, in particular recent results on generalized white noise functionals.

Having established the analytic properties of those multiple Markov Gaussian processes, we can observe some basic properties of those processes, then we shall come to some actual procedures to have the innovation as well as the best predictor of the future values based on the past observed data.

We also discuss the entropy loss which is one of the characteristics of multiple Markov Gaussian process expressing the rate of transmission of information.

## §1. Introduction

The multiple Markov properties of Gaussian processes have been given by T. Hida in [1] 1960, having been motivated by P. Lévy's research on Gaussian processes in 1960 at the 3rd Berkeley Symposium. Let $X(t)$, $t \in T$ be a Gaussian process. The definition of its Markov property and multiple Markov properties should be given in such a way that it expresses how those random variables $X(t)$'s are depending on each other. Since the $X(t)$ is Gaussian, their relationship can be described in terms of the correlation function $\Gamma(t, s)$ which exists for every pair $(t, s)$. We may assume that $E(X(t)) = 0$ identically, so that simply we consider

$$\Gamma( t, s ) = E( X(t)X(s) )$$

Thus the multiple Markov properties shall be discussed basically in terms of the covariance function.

Another background to be prepared comes from the main idea of white noise analysis. It is the idea that is called the **reductionism**. Given a general ( not necessarily to be restricted to a Gaussian system) random complex system. To analyze such a system, we first form a system of **independent** random variables that should contain the same information as the given system. Then, the phenomena to be investigated should be expressed as functions of the independent random variables that have been constructed. We are thus ready to analyze the functions, and hence study the random phenomena to be investigated.

Once the given system is restricted to be Gaussian, the main, actually necessary, computations are linear, to our big advantage. Some more details shall be discussed in the next section.

## §2. White noise

Following the reductionism, we first try to find a system of independent random variables that has the same information as the given random phenomena. In general, this problem is too hard to establish. One of the well known direction to approach is in line with the innovation theory. The idea of this theory may be expressed in the form of the infinitesimal equation proposed by P. Lévy. For a given stochastic process $X(t)$ we are interested in the following formal equation

$$\delta X(t) = \Phi(X(s), s \le t, Y_t, t, dt), \qquad (1)$$

where $Y_t$ is the innovation which is an infinitesimal random variable which is independent of the $X(s)$, $s \le t$ and contains the full new information that is gained by the process during the infinitesimal time interval $[ t, t + dt )$. See P. Lévy [5].

This equation illustrates the structure of the process $X(t)$ from the viewpoint of the reductionism.

Although the formula(1) shows clear meaning, the practice to have the innovation is, in general, very difficult, and even so for the computations. There is however, one exceptional case, where this idea can be realized: it is the Gaussian process (as well as fields or systems, etc). We can appeal to the theory of the canonical representations of Gaussian processes, for which we can see how to form the innovation under our idea. We shall first state this fact in somewhat new style in the next section.

This report has another purpose. There may be attempts to give a definition of multiple Markov properties as a generalization of the simple Markov property. We take, among others, the definition of the multiple Markov property as is given in the next section. This definition can be naturally generalized to Gaussian random fields and even to generalized Gaussian processes under quite natural way of generalization. Further, we shall see the processes and fields specified by the definition are quite fitting for discussing their roles in information theory and for the forecasting problems of the systems in question as we shall see some details later.

## §3. Multiple Markov properties, revisited

We shall remind the definition of multiple Markov properties which should be a generalization of the simple Markov property. In the paper [1], Hida has given a definition of the *N*-ple Markov property of a Gaussian process. (The paper[1] is an old literature, however, we refer mainly for the historical interest.) At present we can rediscover the ideas behind the definition that was given more than fifty years ago. The ideas we can state in the following forms:

i) The definition has been given in connection with the aim of forecasting future values. There is involved the causality, namely the time propagation is always taken into account.

ii) Having been guided by the reductionism, we have a representation of a Gaussian process in term of white noise. More precisely, the given Gaussian process $X(t)$ should be expressed as a functional of idealized elemental variables, i.e. as a function of independent identically distributed (i.i.d.) atomic random variables. Since we restrict our attention to Gaussian systems, we are suggested to use a white noise $\{\overset{\circ}{B}(t), t \in R^1\}$, or almost equivalent, we may take Brownian motion $B(t), t \in R^1$. Then we can discuss multiple Markov properties by using the representation of the Gaussian process $X(t)$ in question.

To fix the idea, we remind the definition of an N-ple markov Gaussian process. We refer to [1]. Let $X(t)$ be a Gaussian process with $E(X(t)) = 0$ for every $t$. Let $N$ be a positive integer.

**Definition 1** If $X(t)$ satisfies the following two conditions for conditional expectations for any fixed $t_0$ :

1. The conditional expectations $E(X(t_i)\backslash B_{t_0}), 1 \le i \le N$ are linearly independent for any different $t_i$'s,
2. $E(X(t_i)\backslash B_{t_0}), 1 \le i \le N+1$ are linearly dependent for different $t_i$'s,

then $X(t)$ is called $N$-ple Markov.

Being particularly related to the idea ii) listed above, we then come to a characterization of the $N$-ple Markov property of $X(t)$ and its analytic representation in terms of white noise. See also [1], [2].

**Theorem 1** Suppose an $N$-ple Markov Gaussian process $X(t)$ has the canonical representation of the form

$$X(t) = \int_0^t F(t,u)\overset{\circ}{B}(u)\,du \tag{2}$$

with an $L^2$- kernel $F(t, u)$ of Volterra type. Then, $F(t, u)$ is expressed in the form of Goursat kernel:

$$F(t,u) = \sum_1^N f_i(t)\,g_i(u)\,.$$

For the later discussions, we remind the definition of a Goursat kernel. In the expression above, we state that it satisfies the following two conditions a) and b).

a) For any different $t_i$'s $\det(f_i(t_j)) \ne 0$.

b) $g_i$'s are linearly independent in $L^2([0,u])$ for any $u \ge 0$.

If $X(t), -\infty < t < \infty$ is a stationary Gaussian process, N-ple Markov property can be defined in the same idea. In this case $F(t, u)$ is expressed as $F(t - u)$ and the Fourier transform $\hat{F}$ of $F$ is of the form

$$\hat{F}(\lambda) = \frac{Q(i\lambda)}{P(i\lambda)}\ ,$$

where $P$ and $Q$ are polynomials in $i\lambda$ and the degree of $P$ is greater than $Q$.

There is another particular case. Let $L_t$ be an $N^{th}$ order ordinary differential operator, and let $X(t)$ satisfy the equation

$$L_t\, X(t) = \overset{\circ}{B}(t),\ t \ge 0, \tag{3}$$

with the initial condition $X(0) = 0$. Then we have a unique solution $X(t)$ which is an $N$-ple Markov Gaussian process. Tasitly, we assume some analytic conditions on the operator $L_t$. we often call the $X(t)$ strictly $N$-ple Markov.

With these background, although it seems old fashioned, we are highly motivated to study further directions to have further developments.

## §4. Entropy Loss

This section devotes to discuss the entropy loss which is one of the characteristics of multiple Markov Gaussian process expressing the rate of transmission of information.

We have observed some properties related to entropy for the discrete parameter case [12]. We are interested in the continuous parameter case. For a continuous parameter case, we take a stationary multiple Markov Gaussian process, the canonical representation is expressed in terms of a Goursat kernel and a white noise with parameter α, such that

$$X_{n+1}(t) = c \int_{-\infty}^{t} \sum_{0}^{n} \binom{n}{k} (-1)^k \exp[-\alpha(t-u)(k+1)] \overset{\circ}{B}(u) du$$

We compute the correlation $r(h)$ of $X_{n+1}(t)$ and obtain the entropy loss with the help of the differential operator $L_t$ which is expressed in the previous section since we are concerned with $(n+1)$ -ple strictly Markov process. We can prove the following theorem.

**Theorem 2** The information (entropy) loss through the $(n+1)$ -ple strictly Markov process is given by *log* $\left( \frac{n+1}{2n-1} \alpha^2 h^2 \right)$. *The information loss decreases as $n$ increases, and the loss gets smaller as the parameters $\alpha$ and $h$ increase.*

Suppose $X(t)$ is a $N$-ple Markov with the canonical kernel, in general, expressed in the form

$$\sum_{1}^{N} a_k e^{-\alpha k(t-u)}$$

under a minor assumption ( namely, the Fourier transform of the kernel has no poles in the upper half plane and no multiple pole in the lower half plane).

Then the covariance function $r(h)$ is of the form

$$r(h) = \sum_1^N b_k e^{-\alpha k|h|}$$

where $b_k = b_k(a_1, \ldots\ldots, a_N)$.

Then we have

**Theorem 3** If $r(h)$ satisfies

$$\sum_1^n b_k k^q = 0, \ q = 1, 2, \ldots\ldots, p,$$

*then* $r(h)$ *is* $p$*-times differentiable, and* $r(h)$ *is of order* $h^{p+1}$ *for small* $h$ *and the entropy loss is* nearly

$$\log(1-(1-\sum_1^n \frac{(kh\alpha)^{p+1}}{(p+1)!})b_k))$$

*and is of order* $(p+1)\log(h)$ *for small* $h > 0$.

## 5. Generalized Gaussian processes

Now comes the main part of this report.

What we have discussed so far can be extended to generalized Gaussian processes. There we shall see a best possible class of Gaussian processes where the multiple Markov properties can be introduced.

First we give a definition of a generalized Gaussian process. We provide a nuclear space $E$ which is dense in the Hilbert space $L^2(R^1)$ and a Probability space $(\Omega, P) = \Omega(P)$. Tacitly, the time parameter space is $R^1$.

Assume that i) $X(\xi) = X(\xi, \omega), \xi \in E, \omega \in \Omega(P)$ is a Gaussian random variable and $X = \{X(\xi), \xi \in E\}$ is a Gaussian system such that $E(X(\xi)) = 0$, and

ii) $X(\xi)$ is linear and strongly continuous in $\xi$ in the space $L^2(\Omega, P)$.

Such a system $X$ is called a generalized Gaussian process.
In order to discuss multiple Markov properties, we further introduce notations and assume necessary conditions as follows:

Let $B_t(X)$ be the σ-field with respect to which all the $X(\xi)$'s with $supp(\xi) \subset (-\infty, t]$ are measurable. Set $B(X) = \vee B_t(X)$. Define the spaces $L_t(X) = L^2(\Omega, B_t(X), P)$ and $L(X) = \vee L_t(X)$. The projection from $L(X)$ down to $L_t(X)$ is denoted by $E(t)$.

With these notations let us continue to have further assumptions.

iii) The space $L(X)$ is seperable.

iv) $X(\xi)$ is purely non-deterministic, that is

$$\bigcap L_t(X) = \{0\}$$

We are now ready to appeal to the Hellinger-Hahn theorem to have the direct sum :

$$\mathrm{L(X)} = \bigoplus_{n}^{m} \mathrm{S_n},$$

where

$$S_n = \bigvee_t \{dE(t)Y_n, t \in R^1\},$$

that is a cyclic subspace generated by some vector $Y_n \in L(X)$. Set $dp_n(t) = \|dE(t)Y_n\|^2$. Then the above $S_n$'s are arranged in the decreasing order of $dp_n$, which gurantees the possibility of multiplicity.

It should not depend on the way of decomposition. Thus, the multiplicity is defined to be the maximum number $m$ of subspace $S_n$.

Our final assumptions are

v) $m = 1$, that is $X(\xi)$ has unit multiplicity, and $dp_1$ is equivalent to the Lebesque measure.

vi) $X(\xi)$ is continuous in $\xi$ with respect to the Sobolev norm of order $-k$ with $k \geq 0$.

By the assumption v) we may assume that $X(\xi)$ is a continuous linear functional, or homogeneous polynomial in $\overset{\circ}{B}(t)$. Namely, in terms of the white noise theory, $X(\xi)$ is a linear generalized white noise functional.

Further, by iv), the kernel function of $X(\xi)$ is in the Sobolev space of order $-k$ over $R^1$. In other words, $X(\xi)$ is a linear homogeneous polynomial in the $\overset{\circ}{B}(t)$'s, the coefficients is in the space $K^{(-k)}(R^1)$ depending linearly on $\xi$ . We use the notation $K^p(R^n)$ to express the Sobolev space of order $p$ over $R^n$.

We assume all the conditions mentioned above.

We are now ready to define the multiple Markov properties of $X(\xi)$.

**Definition 2** A generalized Gaussian process $X(\xi)$ is $N$-ple Markov generalized Gaussian process if for any fixed $t_0$ and for any linearly independent $\xi_i$'s with

$$supp(\xi_i) \subset [t_0, \infty)$$

the conditional expectations

1) $E(X(\xi_i) \backslash B_{t_0}(X)), 1 \le i \le N$ are linearly independent, and
2) $E(X(\xi_i) \backslash B_{t_0}(X)), 1 \le i \le N+1$ are linearly dependent.

To fix the idea, we assume that $X(\xi)$ is uniformly $N$-ple Markov, that is , it is $N$-ple Markov for any time interval. Further we assume that $X(\xi)$ is continuous in $\xi\ (\in K^N(R^1))$, where the notation $K^P(R^1)$ is the Sobolev space of order $p$ ($p$ can be any non-zero real number) over $R^1$. The last assumption implies that almost all sample function of $X$ is in $K^{(-N)}(R^1)$.

We can now state a fundamental theorem basically due to Si Si [7] section 7.5 with some modifications.

**Theorem 4** Let $X(\xi)$ be uniformly $N$-ple Markov. Then, there exist two systems of functions $\{f_i, 1 \le i \le N\}$ and $\{g_i, 1 \le i \le N\}$, respectively, such that each system involves $N$ linearly independent functions in the symmetric $K^{(-N)}(R^1)$ and that $\det(< f_i, \xi_j >) \ne 0$ for any linearly independent $\xi_j$'s.

Further, we find a white noise $\overset{\circ}{B}(t)$ such that

$$E(X(\xi) \backslash B_t(X)) = \sum_1^N < f_i, \xi > U_i^t,$$

where $\xi \in E$ and where $U_i^t = < g_i^t, \overset{\circ}{B} >$, $g_i^t$ being the restriction of $g_i$ to $(-\infty, t]$.

Although the expression of the theorem is somewhat complicated, one can, however, easily see that this is in line with the idea of defining the *N-ple* Markov (ordinary) Gaussian process. It can be recognized that this result is best possible in generalization of multiple Markov properties so far as Gaussian is concerned. Thus based on these observations, we can proceed to further investigations in the following section.

## §6. Forecasting and controls of multiple Markov generalized Gaussian processes

Given an *N-ple* Markov generalized Gaussian process $X(\xi)$ satisfying all the conditions in the last theorem. Take $g_1's$ and apply a regularization of generalized function by using a test function $\eta$ in such a way that $\tilde{g}_i(u) = (g_i * \eta)(u)$. Thus, the enough analytic properties of functions involved in what follows are guaranteed, although the $\eta$ is not written explicitly.

The linearly independent property also holds for $\tilde{g}_i's$ so that we can form a Frobenius formula for the $\tilde{g}_i's$. There is defined a linear differential operator $L_u^*$ of the $N$-th order such that

$$L_u^* \tilde{g}_i = 0, i = 1,2,\ldots\ldots, N,\ldots$$

We can define a linear differential operator $L_t$ which is the formal adjoint of $L_u^*$. Associated with this operator $L_u^*$ is a fundamental system $\{\tilde{f}_i, 1 \le i \le N\}$ of solutions of

$$L_t \ \tilde{f}_i = 0, \quad 1 \le i \le N.$$

Then, we can prove that there exist a matrix $A = A(t)$, which is in $GL(N, R^1)$ for every $t$, such that

$$(\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_N) = A(f_1, f_2, \ldots\ldots, f_N).$$

Replacing $f_i's$ in the expression of $X(\xi)$ with $\tilde{f}_i's$, we form a (generalized) Gaussian process $\tilde{X}(\xi)$. By using these facts we can prove

**Theorem 5**

$$L_t \tilde{X}(\xi) = \overset{\circ}{B}(\xi).$$

Proof comes from the fact that the kernel $\sum_{1}^{N} \tilde{f}_i(t) g_i(u)$ of Volterra type is the Riemann's function associated with the linear differential operator $L_t$.

By the actual computations the $\overset{\circ}{B}(\xi)$ which is depending on the test function $\eta$, we can remove it.

Finally we note that we can find $\overset{\circ}{B}(t)$ within the white noise theory. Again, noting the computations used above , we have

**Corollary**

The $\overset{\circ}{B}(t)$ is $B_t(X)$ - measurable.

Thus, we have obtained the so-called innovation of $X$.

In line with the causal calculus where the time developments are always taken into account and where time order-preserving, we can freely discuss forecasting and control of $X$ by using the annihilation operator $\partial_t = \frac{d}{d\overset{\circ}{B}(t)}$ and the creation operator $\partial_t^*$.

**Example** of forecasting (prediction). Let $t$ be the present time. We can form the best forecasting element of $X(\xi)$ base on the observed values up to present in the following element.

$$E(X(\xi) \setminus B_t(X)) = E(X(\xi) \setminus B_t(\overset{\circ}{B})).$$

The right hand side can be computed since we have actually obtained $\overset{\circ}{B}(s),\, s \le t$.

## References

[1] T.Hida, Canonical representations of Gaussian processes and their applications. Mem. Coll. of Sci. Univ. of Kyoto. Ser.A, 1960,109-155.
[2] T. Hida, Brownian motion. Springer-Verlag. 1980.
[3] T. Hida and Si Si, An innovation approach to random fields; Application of White noise theory. World Sci. Pub. Co. 2004.
[4] T. Hida and Si Si, Lectures on White Noise Functionals. World Scientific Pub. Co. 2008.
[5] P. Lévy, Random functions: General theory with special reference to Laplacian random functions, Univ. Calif. Oub. In Statistics 1. 331-388, 1953.
[6] P. Lévy, A special problem of Brownian motion and a general theory of Gaussian random fields. Proc. 3rd Berkeley Symp. On Math. Statist. And Probability. Vol. 2, 1956, 133 – 175.
[7] Si Si, Introduction to Hida Distributions. World Scientific Pub. Co. 2011.
[8] Si Si, Systems of idealized elemental random variables depending on space parameter. to appear
[9] Si Si and Win Win Htay, Entropy in subordination and Filtering. Acta Applicandae Math. Vol 63, pp 433-439, 2000, Kluwer Academic Publishers
[10] N. Wiener, Nonlinear problems in random theory. The MIT Press, 1958.
[11] Win Win Htay, Optimalities for Random functions: Lee-Wiener's Network and Non-canonical representations of stationary Gaussian processes. Nagoya University Math. J. 149, pp 9-17, 1998.
[12] Win Win Htay, Multiple Markov properties for random fields, International Conference on " Quantum Information and Complexity ", Meijo Winter School, 2003, Nagoya, Japan.
[13] Win Win Htay, Linear Process, Workshop on ' White Noise Analysis, Innovation, Complexity and Stochastic Mechanics ', 2003, International Institute for Advanced Studies, Kyoto, Japan.
[14] Win Win Htay, Note on Linear Process, Proceedings of " Quantum Information and Complexity ", pp 449-455, ed. T. Hida et. World Scientific Pub. Co. 2004.