

## TRANSLATING GAUSSIAN PROCESSES<sup>1</sup>

BY BENNETT EISENBERG

Cornell University

**0. Introduction.** Let  $\mu_x$  denote the measure on path space corresponding to a stochastic process  $X$ . A function  $m$  is an admissible translate of  $X$  if  $\mu_x$  and  $\mu_{x+m}$  (alternatively  $d\mu_x(\omega)$  and  $d\mu_x(\omega-m)$ ) are mutually absolutely continuous. The problems of determining conditions for equivalence and of finding the corresponding Radon-Nikodym derivative  $d\mu_{x+m}/d\mu_x$  (alternatively the Jacobian of the path space transformation  $\omega \rightarrow \omega - m$ ) have been widely studied for Gaussian processes.

For the stationary case, Parzen [1] showed that  $m$  is admissible if and only if for  $t$  in the parameter set of the process  $m(t)$  can be written as  $\int e^{i\lambda t} g(\lambda) dF(\lambda)$  for some  $g$  in  $L^2(dF)$ , where  $F$  is the spectral measure of the process. For the Wiener process Segal [2] showed that  $m$  is admissible if and only if  $m(t)$  can be written as  $\int_0^t g(s) ds$  for  $g$  in  $L^2$  and for  $t$  in the parameter set. Completely general conditions for admissibility of translations of arbitrary Gaussian processes are now known. One form of these conditions is that  $m(t)$  must be representable as  $E(X_t, \psi)$ , where  $X_t$  is the random variable evaluation at time  $t$  and where  $\psi$  is some element in the Hilbert space spanned by  $X_t$ ,  $t$  in the parameter set. Another form states that if  $R(s, t)$  is the covariance of  $X$ , then  $m$  must be in the reproducing kernel Hilbert space with kernel  $R$ . Finally, if  $R$  is assumed continuous, the condition is that  $m$  must be in the range of  $\mathbf{R}^{\frac{1}{2}}$  acting on  $L^2[T]$ , where  $\mathbf{R}^{\frac{1}{2}}$  is the square root of the integral operator with kernel  $R$ , and where  $T$  is the parameter interval.

These conditions have been derived by varied methods. To clarify the relation of these results it is worthwhile, without going into the derivations, to show that the different forms of description do indeed describe the same set of translations. This is done in Section 1.

These descriptions have only an indirect probabilistic appeal. In particular, none gives a direct relation between what might be called the innovation structure of the process and the properties of its admissible translations. In Section 2 a new form of admissibility condition is given involving the innovations of the process  $X$ . In Section 3 the structure of the Radon-Nikodym derivative considered as a stochastic process is exposed.

**Section 1.** Let  $X$  be a mean zero Gaussian process with continuous covariance  $R(s, t)$  over a real interval  $T$ . Denote by  $H_T$  the Hilbert space spanned by the functions  $X_t(\omega)$ ,  $t \in T$ . Denote by  $\mathbf{X}$  the integral operator from  $H_T$  to  $L^2(T)$  with kernel  $X(t, \omega)$  ( $=X_t(\omega)$ ), which we may assume to be measurable.

### PROPOSITION 1.

*The following sets are the same:*

1. The range of  $\mathbf{R}^{\frac{1}{2}}$  acting on  $L^2(T)$ .

Received September 10, 1969.

<sup>1</sup> This research was partially supported by NSF Grant GP-7128X.

2. The range of  $\mathbf{X}$  acting on  $H_T$ .
3. The functions in the reproducing kernel Hilbert space with kernel  $R(s, t)$ ,  $s, t \in T$ . (The R.K.H.S.)

PROOF. 1. = 2.

From the Karhunen–Loève expansion it is seen that  $\mathbf{X}$  is a Hilbert–Schmidt operator from  $H_T$  to  $L^2(T)$ . If  $\mathbf{X}^*$  denotes its adjoint then  $\mathbf{X}\mathbf{X}^*$  is the operator from  $L^2(T)$  to  $L^2(T)$  with kernel

$$\int X(t, \omega)X(s, \omega) d\mu(\omega) = R(s, t).$$

$(\mathbf{R}^\pm)^{-1} = \mathbf{R}^{-\pm}$  is an operator on a dense domain in  $L^2(T)/\ker \mathbf{R}$ .  $X^*$  is defined on  $L^2[T]/\ker \mathbf{R}$  since  $\ker X^* = \ker \mathbf{R}$  and  $\mathbf{X}^* \mathbf{R}^{-\pm}$  gives a unitary map from  $L^2(T)/\ker \mathbf{R}$  to  $H_T$ . That is, if  $(\cdot, \cdot)$  and  $(\cdot, \cdot)_T$  denote the  $L^2(d\mu)$  and  $L^2(T)$  inner products, respectively, then for  $f$  in the domain of  $\mathbf{R}^{-\pm}$

$$(\mathbf{X}^* \mathbf{R}^{-\pm} f, \mathbf{X}^* \mathbf{R}^{-\pm} f) = (\mathbf{X}\mathbf{X}^* \mathbf{R}^{-\pm} f, \mathbf{R}^{-\pm} f)_T = (f, f)_T.$$

The map is onto since by the mean-square continuity of the process, as  $f_n$  approaches a delta function at  $t$ ,  $\mathbf{X}^* f_n$  approaches  $X_t(\omega)$  and these functions generate  $H_T$ .

The induced map of operators on  $H_T$  takes  $\mathbf{X}$  into  $\mathbf{X}\mathbf{X}^* \mathbf{R}^{-\pm} = \mathbf{R}^\pm$  acting from  $L^2(T)/\ker \mathbf{R}$  onto the range of  $\mathbf{R}^\pm$  acting on  $L^2(T)$ .

2. = 3.

Direct proofs that 2. = 3. are known, but we give one here for completeness.

For each  $\psi$  in  $H_T$  define  $m_\psi(t) = (X_t, \psi)$  and give this function space the inner product  $\langle m_\psi, m_\varphi \rangle = (\psi, \varphi)$ . These functions are precisely the functions in 2. and we claim they form the reproducing kernel Hilbert space with kernel  $R$ .

Since  $m_{x_t} = R(t, \cdot)$  the space is spanned by linear combinations of the functors  $R(t, \cdot)$ .  $R(t, \cdot)$  is a reproducing kernel since  $\langle R(t, \cdot), m_\psi(\cdot) \rangle = (X_t, \psi) = m_\psi(t)$ .  $\square$

The functions in 1. are really functionoids. To remove the ambiguity we take continuous representatives. This is possible since the elements in 2. are obviously continuous. Also it is important to note that for each admissible  $m$  there is a unique  $\psi$  in  $H_T$  for which  $m(t) = (X_t, \psi)$  for all  $t$  in  $T$ . For if  $(X_t, \psi) = 0$  for all  $t$  in  $T$ , then  $\psi$  is 0. That is, the functions  $X_t$  generate  $H_T$ .

The Parzen and Segal results follow by exploiting special isometries for stationary processes and the Wiener process.

**Section 2.** Assume  $X$  and  $X+m$  are equivalent, where  $m(t) = (X_t, \psi)$ . Since the processes have the same covariance function, the map  $U$  taking  $X_t$  into  $X_t - (X_t, \psi)$  is an isometry from  $L^2(d\mu_x)$  to  $L^2(d\mu_{x+m})$  and extends linearly to a map from  $H_T$  to the closed subspace of  $X_t - (X_t, \psi)$  in  $L^2(d\mu_{x+m})$  by defining  $Ug$  to be  $g - (g, \psi)$  for  $g$  in  $H_T$ . Furthermore, since constants are orthogonal to  $H_T$  in  $L^2(d\mu_x)$  as well as to the span of  $X_t - (X_t, \psi)$  in  $L^2(d\mu_{x+m})$ ,  $U$  can be extended to the constants by defining  $Uc$  to be  $c$ .

The Radon–Nikodym derivative  $\Lambda(\omega) = d\mu_{x+m}(\omega)/d\mu_x$  is given by

$$\exp [\psi(\omega) - \frac{1}{2}(\psi, \psi)]$$

(see Rozanov [3]). According to the measure  $d\mu_x(\omega)$ ,  $\log \Lambda(\omega)$  is Gaussian with mean  $-\frac{1}{2}(\psi, \psi)$  and variance  $(\psi, \psi)$ ; according to  $d\mu_{x+m}$ ,  $\log \Lambda(\omega)$  is Gaussian with mean  $(\psi - \frac{1}{2}(\psi, \psi), 1)_{\mu_{x+m}} = (\psi + \frac{1}{2}(\psi, \psi), 1) = \frac{1}{2}(\psi, \psi)$  and variance

$$(\psi - (\psi, \psi), \psi - (\psi, \psi))_{\mu_{x+m}} = (\psi, \psi).$$

It follows that in the detection problem between  $\mu_x$  and  $\mu_{x+m}$  the greater  $(\psi, \psi)$ , the smaller the error probabilities.

Now consider the problem of deciding between the hypotheses  $X$  and  $X+m$  over the parameter set  $T'$ . The processes are least distinguishable if  $m \equiv 0$ , but if  $m$  must take certain values on  $T \subset T'$  the extension to  $T'$  leaving the processes as indistinguishable as possible is no longer trivial. From the preceding analysis, assuming the processes are equivalent over  $T$ ,  $m(t)$  can be expressed as  $(X_t, \psi)$  for  $t$  in  $T$  with some  $\psi$  in  $H_T$ . Moreover, the processes are least distinguishable the smaller  $(\psi, \psi)$  is. To minimize the additional information in  $T'$  we should therefore keep the same  $\psi$  and define  $m(s)$  as  $(X_s, \psi)$  for  $s$  in  $T'$ . We denote such an extension of  $m$  by  $\tilde{m}_T$  and call it the minimal extension of  $m$ .

Let  $T$  be a finite set  $\{t_i\}$ . Denote by  $T_k$  the elements  $t_1, \dots, t_{k-1}$  so that  $\tilde{m}_{T_k}(t_i)$  is the minimal extension of  $m$  at  $t_i$  assuming the values  $m(t_1), \dots, m(t_{k-1})$  are fixed.

Denote the norm of the innovation (the prediction error) in the process  $X$  at  $t_i$  by  $\sigma_T(t_i) = \|X_{t_i} - P_{T_i} X_{t_i}\|$ . The ratio of the difference of  $m(t_i)$  and the minimal extension value  $\tilde{m}_{T_i}(t_i)$  to the prediction error is the critical factor in admissibility.

PROPOSITION 2. *Let  $X$  be a Gaussian process separable over a denumerable set  $r$ . A continuous  $m$  is an admissible translate of  $X$  if and only if*

$$\sup_{\text{finite } T \subset r} \sum_i \frac{(m(t_i) - \tilde{m}_{T_i}(t_i))^2}{\sigma_T^2(t_i)}$$

is finite where we say  $0/0 = 0, k/0 = \infty$  for  $k > 0$ .

PROOF. Assume  $m(t) = (X_t, \psi)$  for  $\psi$  in  $H_r$ . By definition of  $H_r$ ,  $\sup_{T \subset r} \|P_T \psi\| = \|\psi\| < \infty$ . But since

$$\left\{ \frac{X_{t_i} - P_{T_i} X_{t_i}}{\sigma_T(t_i)} \right\}$$

is an orthonormal basis for  $H_T$ ,

$$\|P_T \psi\|^2 = \sum_i (\psi, X_{t_i} - P_{T_i} X_{t_i})^2 / \sigma_T^2(t_i) = \sum_i (m(t_i) - \tilde{m}_{T_i}(t_i))^2 / \sigma_T^2(t_i).$$

Conversely, assume

$$\sup_{T \subset r} \sum (m(t_i) - \tilde{m}_{T_i}(t_i))^2 / \sigma_T^2(t_i) < \infty.$$

Then for each finite set  $T$  there exists a unique  $\psi_T$  with  $m(t) = (X_t, \psi_T)$ ,  $t$  in  $T$ . (If all  $\sigma_T(t_i) > 0$ , equivalence over  $T$  is trivial. If not, then for some  $i$ ,  $X_{t_i} = P_{T_i} X_{t_i} = \sum_{k=1}^{i-1} c_k X_{t_k}$  and we clearly must have  $m(t_i) = \sum c_k m(t_k)$  for equivalence of  $X$  and  $X+m$ . But  $m(t_k) = (X_{t_k}, \psi)$  for  $k < i$ , so  $\sum c_k m(t_k) = (\sum c_k X_{t_k}, \psi) = (P_{T_i} X_{t_i}, \psi) = \tilde{m}_{T_i}(t_i)$  and  $m(t_i) = \tilde{m}_{T_i}(t_i)$  by assumption.)

If  $T \subset S$  then  $\psi_T = P_T \psi_S$  so that we may choose  $T_n \subset T_{n+1}$  with

$$\lim_n \|\psi_{T_n}\| = \sup_{T \subset r} \|\psi_T\| \quad \text{and} \quad \bigcup T_n = r$$

$\psi_{T_n}$  is Cauchy and converges to some  $\psi$  with

$$(X_t, \psi) = (X_t, \lim P_{T_n} \psi) = m(t), \quad t \in \bigcup T_n = r.$$

But since  $m$  is continuous,  $X$  and  $X+m$  are separable over  $r$  so equivalence over  $r$  implies admissibility of  $m$ .  $\square$

This form is a nice generalization of the Brownian motion result which now takes the form  $\sup_T \sum (\Delta m / \Delta t)^2 \Delta t < \infty$ .

SECTION 3. A similar analysis yields another form for the Radon-Nikodym derivative. If  $X$  and  $X+m$  are equivalent over  $[0, T']$  then

$$\frac{d\mu_{X+m}}{d\mu_X}(\omega) =_{\text{def}} \Lambda(T', \omega) = \exp[\psi - \frac{1}{2}(\psi, \psi)],$$

where  $m(t) = (X_t, \psi)$ . It follows that over a smaller observation period  $T \subset T'$

$$\Lambda(T, \omega) = \exp[P_T \psi - \frac{1}{2} \|P_T \psi\|^2]$$

as in the Brownian motion case where

$$\Lambda(t, \omega) = \exp[\int_0^t f db - \frac{1}{2} \int_0^t f^2 ds].$$

In this section the differential structure of  $\log \Lambda(t, \omega)$  or from the above point of view the relation of  $\Delta_t P \psi$  to the innovation structure of the processes is studied.

Let  $T$  be a finite set  $\{t_i\}$ . Then

$$\begin{aligned} (P_{T_{k+1}} - P_{T_k})\psi &= \frac{(\psi, X_{t_k} - P_{T_k} X_{t_k})(X_{t_k} - P_{T_k} X_{t_k})}{\sigma_{T^2}(t_k)} \\ &= \frac{(m(t_k) - \tilde{m}_{T_k}(t_k))(X_{t_k} - P_{T_k} X_{t_k})}{\sigma_{T^2}(t_k)} \end{aligned}$$

and

$$\log \Lambda(T, \omega) = \sum_k \frac{(m(t_k) - \tilde{m}_{T_k}(t_k))(X_{t_k} - P_{T_k} X_{t_k})}{\sigma_{T^2}(t_k)} - \frac{1}{2} \sum_k \frac{(m(t_k) - \tilde{m}_{T_k}(t_k))^2}{\sigma_{T^2}(t_k)}.$$

Taking limits we have

$$\log \Lambda(S, \omega) = \lim_{T_\alpha \rightarrow S} \log \Lambda(T_\alpha, \omega).$$

Again we have a direct generalization of the Brownian motion result, which could be written

$$\log \Lambda(S, \omega) \lim_{T_\alpha \rightarrow S} \sum \left( \frac{(\Delta b) \Delta m}{\Delta t} - \frac{1}{2} \frac{\Delta m^2}{\Delta t} \right)$$

an interesting observation, but everything is hidden in the formula  $\Lambda(\omega) = \exp(\psi - \frac{1}{2}(\psi, \psi))$ . A more direct approach gives insight and involves some novel manipulations with Gaussian Hilbert spaces.

Assume  $\Lambda(T, \omega)$  is known for  $T$  a finite set  $\{t_i\}, i = 1 \cdots k-1$ . Denote  $T \cup t_k$  by  $T'$  and write  $\Delta_k X$  for  $X_{t_k} - X_{t_{k-1}}$ .  $\Delta_k X$  has a conditional Gaussian distribution with respect to  $\mu_x$  and  $\mu_{x+m}$ . For the process  $X, E_{\mu_x}(\Delta_k X | \mathcal{F}_T) = P_T(\Delta_k X)$ . For the process  $X+m$ , the conditional mean can be expressed in several ways. Intuitively, we have

$$\begin{aligned} E_{\mu_{x+m}}(\Delta_k X | \mathcal{F}_T)(\omega) &= E_{\mu_x}(\Delta_k X | \mathcal{F}_T)(\omega - m) + \Delta_k m \\ &= P_T(\Delta_k X)(\omega - m) + \Delta_k m. \end{aligned}$$

But what is  $P_T(\Delta_k X)(\omega - m)$ ?

Another approach in effect gives an explicit expression for  $P_T(\Delta_k X)(\omega - m)$ . Let  $\hat{P}_T$  denote the projection on the span of  $X_t - m_t, t \in T$  in  $L^2(d\mu_{x+m})$ . All elements here have mean zero so that

$$E_{\mu_{x+m}}(\Delta_k X | \mathcal{F}_T) = \hat{P}_T(\Delta_k X - \Delta_k m) + \Delta_k m.$$

But the map  $U$  taking  $g$  to  $g - (g, \psi)$  is unitary from  $L^2(d\mu_x)$  to  $L^2(d\mu_{x+m})$  taking  $H_T$  to the span of  $X_t - m_t$  so that

$$\begin{aligned} \hat{P}_T(\Delta_k X - \Delta_k m) &= UP_T(\Delta_k X) = P_T(\Delta_k X) - (P_T \Delta_k X, \psi) \\ &= P_T(\Delta_k X) - \tilde{m}_{T_k}(t_k) + m(t_{k-1}). \end{aligned}$$

Hence

$$\begin{aligned} E_{\mu_{x+m}}(\Delta_k X | \mathcal{F}_T) &= P_T(\Delta_k X) - \tilde{m}_{T_k}(t_k) + m(t_{k-1}) + m(t_k) - m(t_{k-1}) \\ &= P_T(\Delta_k X) - \tilde{m}_{T_k}(t_k) + m(t_k). \end{aligned}$$

The conditional variance of  $\Delta_k X$  with respect to  $\mu_x$  is

$$E_{\mu_x}[(\Delta_k X - P_T(\Delta_k X))^2 | \mathcal{F}_T] = E_{\mu_x}((X_{t_k} - P_T X_{t_k})^2 | \mathcal{F}_T) = \sigma_T^2(t_k)$$

by the independence of  $(X_{t_k} - P_T X_{t_k})^2$  and  $\mathcal{F}_T$ . The conditional variance of  $\Delta_k X$  with respect to  $\mu_{x+m}$  is the same. For the finite dimensional case the Radon-Nikodym derivative is

$$\Lambda(T', \omega) = \Lambda(T, \omega) \frac{d\mu_{x+m}(\Delta_k X | \mathcal{F}_T)}{d\mu_x(\Delta_k X | \mathcal{F}_T)}$$

so that

$$\begin{aligned} \log \frac{\Lambda(T', \omega)}{\Lambda(T, \omega)} &= \frac{-1}{2\sigma_T^2(t_k)} (\Delta_k X - P_T(\Delta_k X) - (m(t_k) - \tilde{m}_{T_k}(t_k)))^2 \\ &\quad + \frac{1}{2\sigma_T^2(t_k)} (\Delta_k X - P_T(\Delta_k X))^2 \\ &= \frac{(X_{t_k} - P_T X_{t_k})(m(t_k) - \tilde{m}_{T_k}(t_k))}{\sigma_T^2(t_k)} - \frac{(m(t_k) - \tilde{m}_{T_k}(t_k))^2}{2\sigma_T^2(t_k)}. \end{aligned}$$

In summary, it seems that the differential approach used in this paper complements the Fourier analytic method usually used in studying Gaussian processes and is in line with the more recent work in stochastic processes.

## REFERENCES

- [1] PARZEN, E. (1961). Regression analysis of continuous parameter time series. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 469–489. Univ. of California Press.
- [2] SEGAL, I. E. (1958). Distributions in Hilbert space and canonical systems of operators. *Trans. Amer. Math. Soc.* **88** 12–41.
- [3] ROZANOV, YU A. (1966). Some remarks on the paper “On Gaussian Distribution Derivatives”. *Theor. Probability Appl.* 483–485.