ON RADON-NIKODYM DERIVATIVES OF FINITELY-ADDITIVE MEASURES INDUCED BY NONLINEAR TRANSFORMATIONS ON HILBERT SPACE

ARUNABHA BAGCHI† and RAVI R. MAZUMDAR‡
† Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; and
‡ INRS-Telecommunications, Université du Québec, 3 Place du Commerce, Île des Soeurs, P.Q. H3E 1H6, Canada

(Received 27 June 1992; received in revised form 24 December 1992; received for publication 18 May 1993)

Key words and phrases: Hilbert spaces, cylindrical measures, Hilbert-Schmidt operators, Radon-Nikodym derivatives, physical random variables, Gauss measure.

1. INTRODUCTION

The study of finitely-additive probability measures arises in problems of modelling and estimation of stochastic signals with bounded variation paths in \( L_2[0, T] \). This is because physical signals are of bounded variation and of finite energy for which \( L_2[0, T] \) is the natural setting and results obtained via Ito theory hold only on a set of Wiener measure 1. It is well known that integrals of paths in \( L_2[0, T] \) lie in a set of Wiener measure zero. Hence, there is a need to construct a theory of white noise to model large bandwidth noise arising in signal analysis with the usual properties associated with it. In a series of papers [1-3], in which he advocated the use of such a framework, Balakrishnan showed that it could indeed be done and, thus, laid the basis for a such a theory. He considered white noise to be the identity map on a Hilbert space \( H \) with standard Gauss measure thereon. It is well known that such a measure is only finitely additive on the algebra of cylinder sets and cannot be extended to the Borel sets of \( H \). This line of work culminated in the excellent treatise by Kallianpur and Karandikar [4] devoted to filtering and smoothing problems. The key assumption in the development of the theory in the nonlinear context so far, has been that the signal process is assumed to be defined on a countably-additive probability space with paths in \( H \) while the measurement noise process is defined on a cylindrical probability space associated with the Gauss measure. The formulation does not allow for signal-noise dependence and one is forced to work with a quasi-cylindrical probability space (a product space). Despite the fact that the mathematical difficulties are daunting, there nevertheless arises the need to develop a complete theory of white noise in order to study modelling issues as well as signal-noise dependence, since these issues arise quite naturally in physical problems. This paper presents one step in such a direction and is motivated by the issue of likelihood ratio evaluation for signals arising in differential systems driven by white noise.

The specific problem addressed in this paper is the study of Radon-Nikodym derivatives (and their evaluation) of cylindrical (or finitely-additive) measures induced by nonlinear transformations on \( H \) with standard Gauss measure thereon. In the linear case this problem was solved by Balakrishnan [5]. Balakrishnan [2] also obtained some results in the nonlinear case when the transformation is given by \( I + K \) where \( K \) defines a homogeneous, finite “Volterra” polynomial by exploiting the connection with the pioneering work of Cameron and Martin [6].
Cameron and Martin obtained the Radon–Nikodym derivative (R–N derivative) for nonlinear transformations of the type above on classical Wiener space. In a situation similar to the one considered in this paper, Gross [7], formulated the problem on $H$ with standard Gauss measure thereon and showed that under suitable conditions on $K$, the R–N derivative is always defined on a representation or lifted space. These ideas were extended to abstract Wiener spaces $(i, H, E)$, where $i$ is the canonical injection from $H$ into a Banach space $E$ with Wiener measure thereon, by Kuo [8] and Ramer [9]. The most comprehensive results in the abstract Wiener space framework were obtained by Kusuoka [10, 11]. The advantage of an abstract Wiener space is that the finitely-additive standard Gauss measure on $H$ extends to a countably-additive measure on $E$ and standard results from measure theory may be directly applied. The difficulty in practical application is that $H$ has zero measure in the abstract Wiener space $(i, H, E)$ while physically realistic signals have realizations in $H$. Thus, probabilistic statements made in an abstract Wiener space cannot often be applied to real data. There are two ways out of this dilemma. One is to convert results into the so-called robust form (see Davis [12]) and the other is to develop results by remaining in $H$ itself. The first approach is somewhat ad hoc and has to be worked out on a case by case basis. The second approach, on the other hand, forces one to work with finitely-additive or cylindrical measures but always yields robust pathwise results eliminating the need for notions such as adaptedness. This is the motivation for developing results in $H$ in line with the pioneering works of Gross, instead of using the elegant and comprehensive results available in an abstract Wiener space. A detailed discussion for the necessity of remaining on $H$ can be found in [13].

In this paper we work with standard Gauss measure on a separable Hilbert space $H$ (typically $L^2[0, T]$) and show that the R–N derivative is given by a Jacobi type of transformation under suitable conditions on the nonlinear map $K$ which generalizes the result in [2]. We also obtain the explicit characterization of the derivative for the finitely-additive measure induced by the process which satisfies a nonlinear differential equation driven by additive white noise. This application is of importance in likelihood ratio evaluation in parameter estimation, as well as the filtering problem in the signal–noise dependence case by the reference probability method. An important observation in this context is the appearance of an additional term which when interpreted for paths in $C[0, T]$ corresponds to the Wong–Zakai correction term [14]. This is a direct consequence of the white noise formulation and obtained without any limiting pathwise approximation.

The organization of the rest of the paper is as follows: in Section 2 we present salient facts from the white noise theory of importance to the problem at hand. In Section 3 we present the main results on the existence of R–N derivatives in such a set-up. In Section 4 we give sufficient conditions for nonlinear maps to define physical random variables (see Section 2). In Section 5 we give a proof of the R–N derivative formula. In Section 6 we specialize the results to the case of differential equations driven by white noise.

2. FINITELY-ADDITIVE WHITE NOISE THEORY—AN OVERVIEW

Let $H$ be a real, separable Hilbert space. Let $\mathcal{P}$ be the class of all finite-dimensional projections on $H$. For any $P \in \mathcal{P}$, a set of the form $P^{-1}(B)$, $B$ a Borel set in range $P$, is called a cylinder set with $B$ as its base. For fixed $P$, we denote the collection of all such cylinder sets by $\mathcal{C}_P$ which is clearly a $\sigma$-algebra. The collection $\mathcal{C} = \{C \in \mathcal{C}_P \mid P \in \mathcal{P}\}$ of all cylinder sets is an algebra.
A finitely-additive measure $\mu$ on $(H, \mathcal{C})$ is called a cylindrical measure if the restriction of $\mu$ on $\mathcal{C}_P$ is a countably-additive measure. If $\mu(H) = 1$, we call $(H, \mathcal{C}, \mu)$ a cylindrical probability space. For any finite number of elements $h_i \in H$, consider the map $\phi: H \rightarrow \mathbb{R}^n$ defined by $\phi(h) = ([h, h_1], \ldots, [h, h_n])$. The mapping induces a countably-additive probability measure $v$ on $\mathbb{R}^n$ defined by $v(B) = \mu(h \in H \mid \phi(h) \in B)$, $B$ Borel in $\mathbb{R}^n$. That $v$ is countably-additive follows from the definition of a cylindrical measure. In particular, for $n = 1$, we can define the characteristic functional corresponding to a cylindrical measure by

$$C_\mu(h') = \int_H e^{i[h, h']} \, d\mu(h) = \int_{-\infty}^{\infty} e^{iy} \, dv_h(y)$$

(2.1)

where $v_h$ is the measure on $[0]$ defined by

$$v_h((-\infty, y]) = \mu(h \in H \mid [h, h'] \leq y).$$

Let $\mathfrak{B}$ denote the Borel $\sigma$-algebra of $H$; it is the smallest $\sigma$-algebra containing $\mathcal{C}$. Since $\mu$ cannot, in general, be extended to be countably additive on $\mathfrak{B}$, most results of measure theory cannot be directly used in this framework. Let $f: H \rightarrow H'$ (possibly another Hilbert space) be a Borel measurable map. This does not qualify $f$ to be a random variable, since inverse images $f^{-1}(B')$, $B'$ a Borel set in $H'$, are not necessarily in $\mathcal{C}$ and probabilities cannot be assigned to these "events". Note, however, that $\{f^{-1}(B') \mid B'$ Borel set in $H')$, for fixed $f$, is a $\sigma$-algebra and it may well be possible to extend $\mu$ by limiting procedure to this sub-$\sigma$-algebra of $\mathfrak{B}$. Those Borel maps $f$ for which this is possible will qualify to be random variables. This is formalized in the following definition.

\textbf{Definition 2.1.} A measurable map of the form $f \circ P$, $P$ a finite-dimensional projection operator on $H$ and $f: PH \rightarrow H'$, is called a tame function. Inverse images of tame functions can always be assigned probabilities. Let $\mathcal{L}(H, \mathcal{C}, \mu; H')$ be the class of Borel measurable functions $f: H \rightarrow H'$ such that for all $\varepsilon > 0$, $\delta > 0$, $\exists P_0 \in \mathfrak{B}$ such that $P_1, P_2 \in \mathfrak{B}$, $P_0 \leq P_i$, $i = 1, 2$, $(P_0 \leq P_i$ if Range $P_0 \subseteq$ Range $P_i$) implies that

$$\mu\{h \in H \mid \|f \circ P_1(h) - f \circ P_2(h)\|' > \delta\} < \varepsilon$$

(2.2)

where $\|\cdot\|'$ denotes the norm in $H'$. An element in $\mathcal{L}(H, \mathcal{C}, \mu; H')$ is called a $H'$-valued $\mu$-random variable. When $H' = \mathfrak{B}$, we denote the space simply by $\mathcal{L}(H, \mathcal{C}, \mu)$. A $H'$-valued $\mu$-random variable $f$ induces a countably-additive measure $\mu'$ on $H'$ which we denote by $\mu \circ f^{-1}$.

\textbf{Definition 2.2.} A mapping $f: (H, \mathcal{C}, \mu) \rightarrow (H', \mathcal{C}')$ is called a $H'$-valued $\mu$-weak random variable if, for all $h' \in H'$

$$[h', f'] \in \mathcal{L}(H, \mathcal{C}, \mu)$$

where $[\cdot, \cdot]'$ is the inner product in $H'$.

\textbf{Remark.} A $H'$-valued $\mu$-weak random variable $f$ induces a cylindrical measure $\mu'$ on $H'$.

\textbf{Definition 2.3.} Let $f: H \rightarrow H'$ be Borel measurable and $\{P_n\}$ be a sequence of monotonically increasing finite-dimensional projections on $H$ converging strongly to the identity, i.e. $P_n \uparrow I$. 


Suppose that \(|f(P_n, h)|\) is Cauchy in \(\mu\) probability and

\[
C(h') = \lim_{n \to \infty} \int_H e^{i[f(P_n, h), h']} \, d\mu(h)
\]

is independent of the sequence \(\{P_n\}\) chosen. Then we call \(f(\cdot)\) a \(\mu\)-p.r.v.

**Lemma 2.4.** Let \(f: H \to H'\) define a p.r.v. Then we can always extend \(\mu\) to the events \(B' \in \mathcal{B}'\) where \(\mathcal{B}'\) denotes the Borel \(\sigma\)-field in \(H'\).

**Proof.** To prove this it is sufficient to show that

\[
C(h') = \lim_{P_n \to I} \int_H e^{i[f(P_n, h), h']} \, d\mu(h)
\]

defines the characteristic functional of a countably-additive measure on \(H\).

Indeed by the convergence of the characteristic functional for all \(P_n \in \mathcal{I}\) we have \(\exists m', \varepsilon > 0\) such that for all \(n, m > m'\)

\[
|C_n(h') - C_n(h')|^2 < \frac{\varepsilon}{2}.
\]

Now by the definition of a p.r.v. as a sequence of tame functions which converges in probability there exists \(N_\varepsilon\) such that on the set

\[
E_m = \{h: \|f(P_n, h)\| \leq N_\varepsilon\}
\]

we have

\[
\mu(E_m) \geq 1 - \frac{\varepsilon}{2}.
\]

Then,

\[
|1 - C_n(h')|^2 \leq 2|1 - C_m(h')|^2 + 2|C_n(h') - C_m(h')|^2
\]

\[\leq 4 \int_H |1 - \cos[f(P_m h), h']| \, d\mu(h) + \varepsilon.
\]

Using the fact that \(1 - \cos(x) \leq x^2/2\) we obtain

\[
|1 - C_n(h')|^2 \leq \int_{E_m} |f(P_m h), h'|^2 \, d\mu(h) + 3\varepsilon.
\]

Defining the operator \(R_m\) as

\[
[R_m h', h'] = 8 \int_{E_m} |f(P_m h), h'|^2 \, d\mu(h)
\]

we see that

\[
\text{Trace } R_m = 8 \int_{E_m} \|f(P_m h)\|^2 \, d\mu(h) < \infty
\]

implying that \(R_m\) defines a nuclear operator.
Therefore, since
\[ |1 - C(h')|^2 \leq 2|1 - C_n(h')|^2 + 2|C_n(h') - C(h)|^2 \]
for \( n \) sufficiently large we obtain the following
\[ |1 - C(h')|^2 \leq [R_n, h', h'] + 4\epsilon \]
implying that \( C(h') \) is uniformly continuous in the \( S \)-topology at the origin and, hence, also everywhere (uniformly) (see definition 2.8). Hence, by the Prohorov theorem [14] it follows that \( C(h') \) is the characteristic functional of a countably-additive measure on \( H \) denoted by \( \mu \).

We denote the class of \( H' \)-valued p.r.v.s by \( \mathcal{L}^*(H, \mathcal{C}, \mu; H') \). When \( H' = \mathfrak{d} \), we denote this class by \( \mathcal{L}^*(H, \mathcal{C}, \mu) \). The definition of \( \mu \)-weak p.r.v.s is similar.

From the point of view of applications we isolate a sub-class of p.r.v.s which are extremely useful.

**Definition 2.5.** Let \( \mathcal{L}^3(H, \mathcal{C}, \mu; H') \) denote the sub-class of \( \mathcal{L}^*(H, \mathcal{C}, \mu; H') \) consisting of measurable functions \( f: H \rightarrow H' \) such that for any sequence \( \{P_n\} \) of finite-dimensional projection operators on \( H \) converging strongly to the identity, the sequence \( \{f(P_n h)\} \) is \( L^2 \) Cauchy in \( \mu \) measure. We simply refer to them as \( L^2(\mu) \) p.r.v.s or as belonging to the class \( L^2(\mu) \).

**Definition 2.6.** The cylindrical measure \( \mu_G \) corresponding to the characteristic functional
\[ C(h) = \int_H e^{ih \cdot h'} \, d\mu_G(h') = \exp(-\frac{1}{2}||h||^2) \]
is called the Gauss (standard) measure on \( H \). \( \mu_G \) can also be defined directly by
\[ \mu_G(C) = \int_{\tilde{B}} G(x) \, dx \]
where \( \tilde{B} \) is the Borel set in \( \mathbb{R}^n \) isomorphic to \( B \), the base of the cylinder set \( C \) and \( G(\cdot) \) denotes the \( n \)-dimensional Gaussian density with zero mean and identity covariance.

When we do not specifically mention the cylindrical measure, we always mean it to be the Gauss measure \( \mu_G \). We refer to \( \mu_G \)- (physical) random variables simply as (physical) random variables.

**Definition 2.7.** The identity map \( I: (H, \mathcal{C}, \mu_G) \rightarrow (H, \mathcal{C}) \) is called Gaussian white noise or white noise for short.

A useful sufficient condition for mapping \( f: H \rightarrow B \) to define a p.r.v. is obtained by considering a locally convex topology called the \( S \)-topology induced by seminorms \( ||L h|| \) where \( L \) is a Hilbert–Schmidt operator mapping \( H \) into \( H \). See [1, 15].

**Definition 2.8.** Let \( H \) be a separable Hilbert space. A map \( f: H \rightarrow B, B \) a Banach space, is said to be continuous w.r.t. the \( S \)-topology as \( x \in H \) if for any \( \epsilon > 0 \) there exists a Hilbert–Schmidt operator \( L_\epsilon(x): H \rightarrow H \) such that
\[ \|L_\epsilon(x)(x - y)\| < 1 \Rightarrow \|f(x) - f(y)\| < \epsilon. \]
If this result holds for \( x, y \in U \subset H \) with \( I_x(x) \) independent of \( x \), \( f \) is said to be uniformly \( S \)-continuous on \( U \). If \( U = H \), we simply say that \( f \) is uniformly \( S \)-continuous.

A sufficient condition for maps to define p.r.v.s is guaranteed by a weaker notion of uniform \( S \)-continuity around the origin. This is introduced in the following definition.

**Definition 2.9.** A map \( f: H \to B \) is said to be uniformly \( S \)-continuous around the origin (u.s.c.a.o.) if \( f(\cdot) \) is uniformly \( S \)-continuous on sets

\[
U_n = \{ x \in H : \|A_n x\| \leq 1 \}
\]

where \( \{A_n\}_1^\infty \) is a sequence of Hilbert–Schmidt operators such that

\[
\|A_n\|_{H,S} \to 0 \quad \text{as } n \to \infty \quad \text{and} \quad \bigcup_{n=1}^\infty U_n = H.
\]

A map which is uniformly \( S \)-continuous is obviously u.s.c.a.o. since defining \( A_n = (1/n)I \) suffices. The following result has been shown by Kuo [8, p. 109].

**Theorem 2.10.** A necessary and sufficient condition for a mapping \( f \) to be u.s.c.a.o. is that there exists a Hilbert–Schmidt operator \( L \) such that \( f \) is uniformly \( S \)-continuous on sets

\[
U_n = \{ x : \|L x\| \leq n \}.
\]

Recently, a characterization of maps which are u.s.c.a.o. was given by De Santis et al. [16] which is given below.

**Theorem 2.11.** A map \( f: H \to B \) is u.s.c.a.o. if and only if there exists a Hilbert–Schmidt operator \( \Lambda: H \to H \) and a continuous map \( g: II \to B \) such that

\[
f = g \circ \Lambda.
\]

We now give the following sufficient condition for a map to define a p.r.v. For a proof see Gandolfi and Germani [17].

**Theorem 2.12.** A sufficient condition for a map \( f: H \to H' \) to be p.r.v., is that it is u.s.c.a.o.

Having defined random variables in the finitely-additive context we can define integration w.r.t. a cylindrical measure in an obvious manner.

We conclude this section with some well-known examples in the white noise context.

**Examples** (see [1]). (1) Let \( L: H \to H' \) be linear and bounded. Then \( Lh \) is a p.r.v. if and only if \( L \) is Hilbert–Schmidt.

(2) \( f: H \to \mathbb{R} \) defined by \( f(h) = [Lh, h] \) where \( L: H \to H \) is linear and bounded is a p.r.v. if and only if \( (L + L^*) \) is nuclear (trace class).

(3) Gaussian white noise is a weak p.r.v. since \( \mu_G \) is countably-additive on finite-dimensional subspaces. It is never a p.r.v. (example (1)).
(4) From example (2) it follows that $f(h) = \|h\|^2$ is never a p.r.v.

(5) If $f(\cdot)$ is uniformly continuous in $S$-topology and $g$ is a strongly continuous map defined on the range of $f$ then $g \circ f(\cdot)$ defines a p.r.v.

3. RADON–NIKODYM DERIVATIVES FOR CYLINDRICAL MEASURES

In this section we discuss the absolute continuity and the existence of Radon–Nikodym derivatives in the cylindrical probability space $(H, \mathcal{C}, \mu)$.

Definition 3.1. A cylindrical measure $\lambda$ is said to be absolutely continuous with respect to another cylindrical measure $\mu$ if, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for any cylinder set $C \in \mathcal{C}$, $\mu(C) < \delta \Rightarrow \lambda(C) < \varepsilon$.

It is important to note that this does not imply the existence of the Radon–Nikodym derivative as in the countably-additive framework. The following definition gives a natural way of defining the Radon–Nikodym derivative in the cylindrical measure framework.

Definition 3.2. Suppose that $\lambda$ and $\mu$ are cylindrical measures on $(H, \mathcal{C})$ and $\lambda$ is absolutely continuous w.r.t. $\mu$. We say that $f$ is the Radon–Nikodym derivative of $\lambda$ w.r.t. $\mu$ if $f$ is a $\mu$-p.r.v. and, for any sequence of projections $P_n \in \mathcal{C}$ increasing strongly to the identity

$$\int_C f(P_n h) \, d\mu(h)$$

uniformly in $C \in \mathcal{C}$ and written as

$$f(h) = \frac{d\lambda}{d\mu}(h)$$

for $h \in H$.

The following two theorems deal with the existence of R–N derivatives in the cylindrical measure context. The proofs also appear in [18] but are repeated for the paper to be self-contained.

Theorem 3.3. Let $\lambda$ and $\mu$ be two cylindrical probability measures on $H$ and let $\lambda$ be absolutely continuous w.r.t. $\mu$. Then there exists a sequence $\{\phi_n\}$ of tame functions with $\int_H |\phi_n(h)| \, d\mu(h) < \infty$ such that

$$\lambda(C) = \lim_{n \to \infty} \int_C \phi_n(h) \, d\mu(h)$$

uniformly in $C \in \mathcal{C}$.

Proof. Our proofs of this and the following theorem are based on the method used by Fefferman [19].

Let $\nu = \lambda + \mu$ so that $\nu$ is a finite cylindrical measure.

Let $T(\nu)$ denote the class of tame functions on $H$ with $\int_H |f(h)|^2 \, d\nu(h) < \infty$. $T(\nu)$ denotes the completion of $T(\nu)$ with respect to the norm $\|f\| = (\int_H |f|^2 \, d\nu)^{1/2}$. 
Clearly, for \( f \in T(v) \),
\[
\int_H f \, dv = \int_H f \, d\mu + \int_H f \, d\lambda.
\]

Define a linear functional \( L \) on \( T(v) \) by
\[
L f = \int_H f \, d\lambda.
\]

Then
\[
|L f| \leq \int_H |f| \, dv \leq \sqrt{2} \left( \int_H |f|^2 \, dv \right)^{1/2}
\]
so that \( L \) is continuous on \( T(v) \) and has a continuous extension \( \bar{L} \) to all of \( \bar{T}(v) \). Therefore, for all \( g \in \bar{T}(v) \), \( \bar{L} g = (f, g) \) for some \( f \in \bar{T}(v) \) where \((\cdot, \cdot)\) is the inner product in \( \bar{T}(v) \). Since \( T(v) \) is dense in \( \bar{T}(v) \), there exists a sequence \( \{f_n\} \) in \( T(v) \) such that
\[
\lim_{n \to \infty} f_n = f \in \bar{T}(v)
\]
and for every \( M > 0 \)
\[
\lim_{n \to \infty} (f_n, g) = \bar{L} g
\]
uniformly for \( g \) in an \( M \)-ball of \( \bar{T}(v) \).

For \( C \subseteq \mathcal{C} \), the indicator function \( 1_C \) is in \( L(v) \) and
\[
\|1_C\| = \sqrt{\nu(C)} \leq \sqrt{\nu(H)} = \sqrt{2}.
\]
Therefore, \( \{1_C, C \subseteq \mathcal{C}\} \) consists of elements of an \( M \)-ball \((M = \sqrt{2})\) of \( \bar{T}(v) \) and it follows that
\[
\lim_{n \to \infty} (f_n, 1_C) = \bar{L} 1_C
\]
uniformly in \( C \subseteq \mathcal{C} \); or equivalently
\[
\lim_{n \to \infty} \int_C f_n(h) \, dv(h) = \lambda(C)
\]
uniformly in \( C \subseteq \mathcal{C} \).

Now \( f \in T(v) \) implies that \( f \in T(\mu) \). Let \( H_n \) be any finite-dimensional sub-space of \( H \). \( H_n \) is isomorphic to a finite-dimensional Euclidean space which we denote by \( \bar{H}_n \) and let \( \bar{h} \) be the isomorphic image of \( h \in H_n \). \( \mu \) restricted to \( H_n \) is a countably-additive probability measure \( \mu_n \) which by means of isomorphism, induces a probability measure on \( \bar{H}_n \) which we also denote by \( \mu_n \). Let \( f_n(h) = \xi_n(P_n h) \) where \( P_n \) is a finite-dimensional projection on \( H \) and \( \xi_n \) is a measurable function on \( P_n H \). Then
\[
\int_C f_n(h) \, dv(h) = \int_C f_n(h) \, d\mu(h) + \int_C f_n(h) \, d\lambda(h)
\]
\[
= \int_C f_n(h) \, d\mu(h) + \int_{P_n C} \xi_n(x) \, d\lambda_n(x)
\]
and since by assumption $\lambda_n \ll \mu_n$ let

$$\eta_n(x) = \frac{d\lambda_n}{d\mu_n}(x) \in L^1(P_n H, \mu_n).$$

Then

$$\int_{\mathcal{P}_n C} \xi_n(x) d\lambda_n(x) = \int_{\mathcal{P}_n C} \xi_n(x) \eta_n(x) d\mu_n(x)$$

$$= \int_{C} f_n(h) g_n(h) d\mu(h)$$

where $g_n(h) = \eta_n(P_n h)$ is a tame function with $\int_{H} |g_n| d\mu < \infty$. Letting $\phi_n = f_n(1 + g_n)$ we see that $\phi_n$ is a tame function with $\int_{H} |\phi_n| d\mu < \infty$ and

$$\int_{C} f_n(h) d\nu(h) = \int_{C} \phi_n(h) d\mu(h)$$

so that

$$\lim_{n \to \infty} \int_{C} \phi_n(h) d\mu(h) = \lambda(C)$$

uniformly in $C \in \mathcal{C}$.

**Theorem 3.4.** Let $\{P_n\}$ be an increasing sequence of projections in $\mathcal{P}$ converging strongly to the identity. Let $\lambda$ and $\mu$ be as in theorem 3.3 and define $\lambda_n = \lambda|_{C_{P_n}}$ and $\mu_n = \mu|_{C_{P_n}}$. Suppose there exists a function $f(\cdot)$ such that

$$f(P_n h) = \frac{d\lambda_n}{d\mu_n}(P_n h).$$

Then

$$\lambda(C) = \lim_{n \to \infty} \int_{C} f(P_n h) d\mu(h)$$

uniformly in $C \in \mathcal{C}$.

**Proof.** Let $X$ denote the vector space of all bounded $\mu$-continuous cylindrical set functions on $\mathcal{C}$. For any $\lambda \in X$, let $\lambda = \lambda^+ - \lambda^-$ be its decomposition into cylindrical measure. Let $\lambda^+_n$, $\lambda^-_n$, and $\mu_n$ be the corresponding Borel measures on the Euclidean space $\mathcal{P}_n H$. Let

$$\xi_n^\lambda = \frac{d\lambda^+_n}{d\mu_n}(x) - \frac{d\lambda^-_n}{d\mu_n}(x), \quad x \in \mathcal{P}_n H$$

and

$$f^\lambda(P_n h) = \xi_n^\lambda(\mathcal{P}_n h).$$

Define continuous linear operators $T_n$ on $X$ by

$$(T_n \lambda)(C) = \int_{C} f^\lambda(P_n h) d\mu, \quad C \in \mathcal{C}, \lambda \in X.$$
The theorem will be proved if we can show that, for each \( \lambda \in X \), \( \lim_{n \to \infty} T_n \lambda = \lambda \). By the Banach–Steinhaus theorem, the above assertion will be true if we can show that

1. \( \lim_{n \to \infty} T_n \lambda = \lambda \) for all \( \lambda \) in a dense subset of \( X \);
2. the norms \( \| T_n \| \) of the operators \( T_n \) are uniformly bounded.

To prove (1), we select as our dense subset, the set of all \( \lambda \in X \) of the form

\[
\lambda(C) = \int_C f(h) \, d\mu(h)
\]

for all tame functions \( f \) with \( \int_H |f(h)| \, d\mu(h) < \infty \). The decomposition of \( \lambda \) and the conclusion of theorem 3.3 implies that this is indeed a dense subset of \( X \). We now show that, if \( \lambda \) is of this form, \( \lim_{n \to \infty} T_n \lambda = \lambda \). For this, let

\[
\lambda(C) = \int_C f(h) \, d\mu(h)
\]

where \( f(h) = \zeta(\tilde{P}h) \), \( P \) being a finite-dimensional projection operator on \( H \) and \( \zeta \) a measurable function on \( \tilde{P}H \). Then

\[
\lambda(C) = \int_{\tilde{P}H} \zeta(x) \, d\mu_P(x)
\]

where \( \mu_P \) is the measure \( \mu|_{C_P} \). Then

\[
(T_n \lambda)(C) = \int_C f^\lambda(P_n h) \, d\mu(h)
\]

\[
= \int_{\tilde{P}_n C} \frac{d\lambda_n^+}{d\mu_n}(x) - \int_{\tilde{P}_n C} \frac{d\lambda_n^-}{d\mu_n}(x), \quad x \in \tilde{P}_n H
\]

\[
= \lambda_n(P_n C) = \int_{\tilde{P}_n C} \zeta(x) \, d\mu_P(x) = \int_{\tilde{P}C} \zeta(x) \, d\mu_P(x) = \lambda(C)
\]

whenever \( P_n > P \), yielding the desired result.

To establish (2), we show that \( \| T_n \| < 1 \) for all \( n \), i.e. \( |T_n \lambda| < |\lambda| \) for all \( n \) and \( \lambda \in X \). Since \( \lambda \) can be decomposed into the difference of two finite cylindrical measures it is enough to show that \( (T_n \lambda)(H) < \lambda(H) \). But

\[
(T_n \lambda)(H) = \int_H f^\lambda(P_n h) \, d\mu(h)
\]

\[
= \int_{\mathbb{R}^n} \frac{d\lambda_n}{d\mu_n}(x) \, d\mu_n(x) = \lambda_n(\mathbb{R}^n) = 1
\]

and this proves the theorem.

4. A CLASS OF NONLINEAR p.r.v.s

In this and the following sections \( \mu \) will denote Gauss measure on \( H \). We give a sufficient condition for a nonlinear mapping \( f: H \to \mathbb{R} \) defined by \( f(h) = [M(h), h] \), \( h \in H \) to define a p.r.v. This is a significant generalization of the result in the linear case (Section 2, example (2)), which we recall below.
THEOREM 4.1. Let $L$ be a linear operator mapping $H \to H$. Then $[Lh, h]$ defines a $L^2(\mu)$-p.r.v. if and only if $L + L^*$ is nuclear.

We begin by first introducing some notation: let $\|A\|_2 = (\text{Trace}(A^*A))^{1/2}$ denote the Hilbert–Schmidt norm where $A$ is a Hilbert–Schmidt operator. Let $\|B\|_1 = \text{Trace}(BB^*)^{1/2}$ where $B$ is a nuclear operator.

Let $M(\cdot)$ be a nonlinear mapping from $H \to H$ and $M_x$ denote the Fréchet derivative at $x \in H$. Then we can state the following theorem which is the analog of theorem 4.1 above for nonlinear mappings.

THEOREM 4.2. Let $M: H \to H$ be a nonlinear mapping which satisfies the following hypotheses:

1. $\|M_x y\| \leq \|Ay\|$; $x, y \in H$, where $A$ is a Hilbert–Schmidt operator from $H$ into $H$, independent of $x$.
2. Let $M_x + M_x^*$ be trace class and $\sup_x \|M_x + M_x^*\|_1 < \infty$.

Then $[M(h), h], h \in H$ is a $L^2(\mu)$ p.r.v.

Before we prove the above result we prove some intermediate results which are necessary.

LEMMA 4.3. Let $T(\cdot): H \to H$ be a nonlinear mapping satisfying the following conditions.

(i) $T(\cdot)$ is u.s.c.a.0.
(ii) $\int_H \|T(x)\|^2 \, d\mu(x) < \infty$.

Then

$$\int_H \|T(P_n h) - P_n T(P_n h)^2 \, d\mu(h) \to 0$$

as $P_n \to I$ strongly.

Proof. Since $T(\cdot)$ is u.s.c.a.0. by theorem 2.11 we have $T(h) = f(I, h)$ where $f$ is continuous and $L$ is Hilbert–Schmidt.

Define the cylindrical measure $\mu * L^{-1}$ on $H$ by

$$\mu * L^{-1}(C) = \mu(h: Lh \in C)$$

for cylindrical sets $C \subset H$. Then $\mu * L^{-1}$ defines a countably additive measure on $H$ since

$$C(x) = \int_H e^{i[h, x]} \, d(\mu * L^{-1}) = \int_H e^{i[Ly, x]} \, d\mu(h) = e^{-1/2\|Lx\|^2}$$

which is continuous in $S$-topology since $L$ is Hilbert–Schmidt and the conclusion follows from the Sazonov theorem [14].

Furthermore for any projection operator $P: H \to H$

$$d\mu * L^{-1}(h) = d\mu * L^{-1}(Ph) \times d\mu * L^{-1}((I - P)h)$$

since

$$e^{-1/2\|Lx\|^2} = e^{-1/2\|PLx\|^2 + \|(I - P)Lx\|^2} = \int_{PH} e^{i[h_1, x]} \, d\mu * L^{-1}(h_1) \int_{(I - P)H} e^{i[h_2, x]} \, d\mu(h_2)$$

implying the orthogonality of the measure $\mu * L^{-1}$ on $PH \oplus (I - P)H$. 
Hence, noting that \( \| T(P_n h) - P_n T(P_n h) \| = \| (I - P_n) f(LP_n h) \| \)

\[
\int_H \| (I - P_n) f(LP_n h) \|^2 d\mu(h) = \int_H \| (I - P_n) f(LP_n h) \|^2 d\mu(P_n h)
\]

\[
= \int_{P_n H} \| (I - P_n) f(h') \|^2 d\mu \ast L^{-1}(h')
\]

\[
= \int_H \| (I - P_n) f(P_n h) \|^2 d\mu \ast L^{-1}(h)
\]

by the use of Fubini's theorem for the orthogonal decomposition. Finally noting that by assumption

\[
\int_H \| (I - P_n) T(P_n h) \|^2 d\mu(h) \leq 2 \int_H \| T(P_n h) \|^2 d\mu(h) = 2 \int_{P_n H} \| T(h) \|^2 d\mu(h)
\]

\[
\leq 2 \int_H \| T(h) \|^2 d\mu(h) < \infty.
\]

Hence, by the dominated convergence theorem, since \( \mu \ast L^{-1} \) defines a countably-additive measure on \( H \) and \( f(\cdot) \) is continuous we have

\[
\int_H \| (I - P_n) f(P_n h) \|^2 d\mu \ast L^{-1}(h) \to 0 \quad \text{as } P_n \to I \text{ strongly}
\]

and the proof is complete.

**Lemma 4.4.** For the class of nonlinear mappings \( M(h) \) satisfying the assumptions above

(i) \( M(h) \) defines an \( L_2 \) p.r.v.

(ii) \( \text{Trace}(M_x + M_x^*) \) is also an \( L_2 \) p.r.v.

**Proof.** (i) Assumption (1) implies, using the mean value theorem, that \( M(\cdot) \) is uniformly \( S \)-continuous and, hence, is also u.s.c.a.o. with

\[
\int_H \| M(x) \|^2 d\mu(x) \leq 2 \| M(0) \|^2 + \text{Trace}(AA^*) < \infty.
\]

This establishes the fact that \( M(\cdot) \) is an \( L_2 \) p.r.v.

(ii) The proof of the second statement is identical to the one of lemma 4.3 as \( \text{Trace}(M_x + M_x^*) \) is u.s.c.a.o. as a mapping from \( H \rightarrow R \) since

\[
M_x = f_{Lx} \circ L.
\]

Noting that \( |\text{Trace}(M_x + M_x^*)| \leq \| M_x + M_x^* \|_1 \) and \( \sup \| M_x + M_x^* \|_1 < \infty \) we obtain that

\[
\int_H |\text{Trace}(M_x + M_x^*)|^2 d\mu(x) < \infty.
\]

Finally, we need the following result shown for standard Gauss measures on finite-dimensional spaces by Ramer [9].
Lemma 4.5. Let $f: \mathbb{R}^n \to \mathbb{R}^n$ be a $C^1$ map such that $|f(x)|$ and $\|f_x\|_2 \in L^2(\mathbb{R}^n, \mu^n)$ here $\mu^n$ is standard Gauss measure in $\mathbb{R}^n$ with variance parameter 1. Then

$$\int_{\mathbb{R}^n} (|f(x)|^2 + \|f_x\|^2) \, d\mu^n(x) \leq \int_{\mathbb{R}^n} (|f(x)|^2 + \|f_x\|^2) \, d\mu^n(x).$$

Proof of theorem 4.2. Let $f: H \to H$ be defined by

$$f(x) = [M(x), x] - \frac{1}{2} \operatorname{Tr}(M_x + M^*_x).$$

Then (taking $n > m$)

$$f(P_n x) - f(P_m x) = [M(P_n x), P_n x] - [M(P_m x), P_m x] + \frac{1}{2} \operatorname{Tr}(M_{P_n x} + M^*_{P_n x})$$

$$= \frac{1}{2} \operatorname{Tr}(M_{P_n x} + M^*_{P_n x})$$

$$= \Phi_1 + \Phi_2$$

where

$$\Phi_1 = [(P_n - P_m)M(P_n x), P_n x]$$

$$\Phi_2 = [P_m(M(P_n x) - M(P_m x)), P_n x].$$

Note $P_n M P_n \to P_n$ defines an operator from $P_n H \to P_n H$ which we can identify with a finite-dimensional mapping from $0^a \to 0^a$. Similarly, defining $L_{P_n x} = M_{P_n x} + M^*_{P_n x}$

$$\psi = \frac{1}{2} \operatorname{Tr}(L_{P_n x} - L_{P_m x})$$

$$= \frac{1}{2} \operatorname{Tr}[(I - P_n)L_{P_n x} + P_n L_{P_n x} - (I - P_m)L_{P_m x} - P_m L_{P_m x}]$$

$$= \Psi_1 + \Psi_2$$

where

$$\Psi_1 = \frac{1}{2} \operatorname{Tr}[(P_n - P_m)L_{P_n x}]$$

$$\Psi_2 = \frac{1}{2} \operatorname{Tr}[(I - P_n)L_{P_n x} - (I - P_m)L_{P_m x} + P_m(L_{P_n x} - L_{P_m x})].$$

Hence,

$$\Phi_1 - \Psi_1 - [(P_n - P_m)M(P_n x), P_n x] - \frac{1}{2} \operatorname{Tr}((P_n - P_m)(M_{P_n x} + M^*_{P_n x}))$$

and noting that in finite dimensions

$$\frac{1}{2} \operatorname{Tr}((P_n - P_m)(M_{P_n x} + M^*_{P_n x})) = \operatorname{Tr}((P_n - P_m)M_{P_n x})$$
and by the use of Fubini's theorem for Gauss measure for the decomposition $H = P_n H \oplus (I - P_n)H$ (by orthogonality) and using lemma 4.4 we obtain

$$\int_H (\Phi_1 - \Psi_1)^2 \, d\mu(x) \leq \int_H \left( \|(P_n - P_m)M(P_n x)\|^2 + \frac{1}{4} \|(P_n - P_m)L_{P_n x}\|^2 \right) \, d\mu(x) \rightarrow 0 \text{ as } n, m \rightarrow \infty$$

where the first term goes to zero by lemma 4.3 since

$$\|(P_n - P_m)M(P_n x)\|^2 \leq 3\|(I - P_n)M(P_n x)\|^2 + 3\|M(P_n x) - M(P_m x)\|^2 + 3\|(I - P_m)M(P_m x)\|^2$$

and the second term is bounded by

$$\int_H \|(P_n - P_m)L_{P_n x}\|^2 \, d\mu(x) \leq \|(P_n - P_m)A\|^2$$

which goes to 0 as $n, m \rightarrow \infty$ since $A$ is Hilbert-Schmidt.

Similarly noting that $(M \circ P_m)_{P_n x} = M_{P_n x} P_m$ and $L_x$ is self-adjoint

$$\Psi_2 = \frac{1}{2} \text{Tr}[(I - P_n)L_{P_n x}(I - P_n) - (I - P_m)L_{P_m x}(I - P_m) + P_m(L_{P_m x} - (L \circ P_m)_{P_n x})]$$

$$= x_1 + x_2$$

where

$$x_1 = \frac{1}{2} \text{Tr}[(I - P_n)(M_{P_n x} + M^*_{P_n x})(I - P_n)] - \frac{1}{2} \text{Tr}[(I - P_m)(M_{P_m x} + M^*_{P_m x})(I - P_m)]$$

$$x_2 = \frac{1}{2} \text{Tr}[P_m(M_{P_n x} + M^*_{P_n x}) - P_m((M \circ P_m)_{P_n x} + (M \circ P_m^*_{P_n x})].$$

Using lemma 4.4 we have

$$\int_H (\Phi_2 - x_2)^2 \, d\mu(x) \leq \int_H \left( \|P_m(M(P_n x) - M(P_m x))\|^2 + \frac{1}{4} \|P_m(L_{P_n x} - L_{P_m x})\|^2 \right) \, d\mu(x) \rightarrow 0 \text{ as } n, m \rightarrow \infty$$

since $[M(P_n x)]$ and $[\text{Trace } L_{P_n x}]$ are $L^2(\mu)$ Cauchy.

Finally

$$\int_H |x_1|^2 \, d\mu(x) \leq 2 \int_H [(\text{Tr}(I - P_n)L_{P_n x})^2 + (\text{Tr}(I - P_m)L_{P_m x})^2] \, d\mu(x) \rightarrow 0 \text{ as } n, m \rightarrow \infty$$

using the fact that $\text{Trace } L_x$ is u.s.c.a.o. and uniformly bounded as in lemma 4.3.

Therefore, if we have shown that $f(P_n x)$ is Cauchy in $L^2(\mu)$ and hence is a p.r.v. and the proof is complete.

Theorem 4.2 provides us with one of the main results needed to prove the R–N theorem for nonlinear maps in the white noise context. This is done in the next section.
5. RADON-NIKODYM THEOREM

Consider the nonlinear map \((I - M): H \to H\) with \(M\) a nonlinear map. Let \(M_x\) denote the Fréchet derivative of \(M\) at \(x \in H\) and suppose it is defined for all \(x \in H\) and is Hilbert-Schmidt, Volterra (quasi-nilpotent). Suppose further that \(M\) satisfies the hypotheses of Section 4, i.e.

(A) the Fréchet derivative \(M_x\) satisfies
\[
\|M_x h\| \leq \|Ah\|; \quad x, h \in H
\]

where \(A\) is a Hilbert-Schmidt operator, independent of \(x\);

(B) the Fréchet derivative \(M_x\) is such that \(M_x + M_x^*\) is trace class and \(\sup_x \|M_x + M_x^*\|_1 < \infty\).

Remark 1. Note that the Fréchet derivative of \(Z - M(\cdot)\) is given by \(Z - hM(\cdot)\) and \(M(\cdot)\) is Hilbert-Schmidt and Volterra. It follows that \((Z - M(\cdot))^{-1}\) exists and by the result in Dunford and Schwartz [20, p. 1039] we have
\[
\|(I - M(\cdot))^{-1}\| \leq \exp\left(\frac{1 + \|M\|_2^2}{2}\right) \leq \exp\left(\frac{1 + \|A\|_2^2}{2}\right).
\]

Hence, by the Hadamard theorem (see Berger [21, p. 222]) it follows that \((I - M)^{-1} = I + K\) exists.

Remark 2. Condition (A) above implies that the Fréchet derivative of \(K\) denoted \(K_x\) satisfies condition (A) for the Hilbert-Schmidt operator \(B\) defined by
\[
B = \exp\left(\frac{1 + \|A\|_2^2}{2}\right)A.
\]

This is because (see Balakrishnan [1])
\[
K_x = (I - M_x)^{-1}M_x,
\]
where \(z(x) = (I - M)^{-1}(x)\) and \(M_x\) is Volterra for all \(x \in H\).

Furthermore, \(M_x\) being Hilbert Schmidt and Volterra, by the mean value theorem we see that \(K(x)\) is uniformly \(S\)-continuous and
\[
\int_H \|K(x)\|^2 \, d\mu(x) < \infty.
\]

Let \(T = (I - M)^{-1}\) and for any cylinder set \(C \in \mathcal{E}\) define the cylindrical measure induced on \(H\) by \(T\) defined by
\[
\mu_T(C) = \mu(h: T(h) \in C) = \mu \circ T^{-1}(C).
\]

Then we can state the following theorem.

**Theorem 5.1.** Let \(M: H \to H\) be a nonlinear operator. Let \(M_x\), the Fréchet derivative of \(M\) at \(x\) in \(H\) be defined for all \(x \in H\) and let \(M_x\) satisfy the hypotheses (A) and (B) above.
Then $T = (I - M)^{-1} : H \to H$, induces a cylindrical measure $\mu_T$ on $H$ which is absolutely continuous w.r.t. $\mu$, the Gauss measure on $H$ and

$$\frac{d\mu_T}{d\mu}(h) = g(h)$$  \hspace{1cm} (5.1)

$$g(h) = \exp\left([M(h), h] - \frac{1}{2}\|M(h)\|^2 - \frac{1}{2} \text{Tr}(M_n + M_n^*)\right)$$  \hspace{1cm} (5.2)

for $h \in H$.

Or equivalently, we have for any cylinder set $C \in \mathcal{C}$

$$\mu_T(C) = \lim_{n \to \infty} \int_C g(P_n h) \, d\mu(h)$$

uniformly in $C \in \mathcal{C}$ as $P_n$ converges to $I$ strongly.

We prove the theorem via the following steps.

Consider the sequence of finite-dimensional maps

$$S_n = (I - P_n MP_n) : P_n H \to P_n H.$$  \hspace{1cm} (5.3)

(1) We show that $S_n^{-1} = T_n$ exists and $T_n = I + P_n NP_n$ with $\|P_n N P_n y\| \leq k \|A P_n y\|$ and denoting the induced cylindrical measure of $T_n$ by $\mu_{T_n}$ we show that the corresponding characteristic function converges to the characteristic function of $\mu_T$ as $P_n \to I$ strongly.

(2) Then using the Jacobian formula for finite dimensions we show that

$$\frac{d\mu_{T_n}}{d\mu}(P_n h) = f_n(P_n h).$$

(3) We then show that $f_n(P_n h)$ converges to $g(h)$ in $L_1(\mu)$ and noting that $g(h)$ is a p.r.v. by theorem 4.2 we complete the proof via theorem 3.4.

**Lemma 5.2.** Let $M(x)$ satisfy the assumptions above and $S_n$ be given by (5.3).

Then:

(i) $S_n^{-1} = T_n = I + P_n NP_n$ exists and $(P_n NP_n)_x = P_n N P_n x$ satisfies $\|(P_n NP_n)_x y\| \leq K \|A P_n y\|$ for $y \in H$ where $K$ is a constant which does not depend on $n$.

(ii) Let $C_n(h)$ denote the characteristic functional of $\mu_{T_n}$ defined by

$$C_n(h) = \int_H e^{i[h, T_n(x)]} \, d\mu(x) = \int_H e^{i[h, y]} \, d\mu_{T_n}(y); \quad y \in P_n H.$$  \hspace{1cm} \hspace{1cm} (5.4)

Then $C_n(h)$ converges to $C(h)$ the characteristic function of $T = (I - M)^{-1}$.

**Proof.**

$$S_n(x) = (I - P_n MP_n)(x).$$

Hence,

$$(S_n)_x = I - P_n M P_n x.$$  \hspace{1cm} (5.5)

Now since $M$ is Hilbert–Schmidt, Volterra it is easy to see that 1 cannot be an eigenvalue of $P_n M P_n x$ implying that 1 is in the resolvent set. Noting that

$$\text{Trace}(P_n M P_n x) = \frac{1}{2} \text{Trace}(P_n (M P_n + M^* P_n)x P_n).$$
Using the fact that $M_n$ is Hilbert–Schmidt and Volterra, [20, theorem 15, p. 1023, lemma 22, p. 1029] and the assumption $\|M_n\|_2^2 \leq \|A\|_2^2$, can be used to show that
\[
\|(I - P_n M_{P_n x} P_n)\| \leq \exp \left( \frac{1}{2} \frac{\|P_n M_{P_n x} P_n\|_2^2}{2} + \frac{1}{2} \|M_{P_n x}(I - P_n)\|_2^2 \right)
\]
\[
\leq \exp \{1 + \|A\|_2^2\} - K.
\]
Hence, by Hadamard’s theorem it follows that $S_n^{-1}$ exists.

Furthermore denoting $S_n^{-1} = T_n = I + P_n N P_n$ and $P_n N P_n = P_n M P_n (I - P_n M P_n)^{-1}$,
\[
(P_n N P_n)_x = (I - P_m M T_n(x) P_n)^{-1} M T_n(x) P_n.
\]
Hence, we obtain
\[
\|(P_n N P_n)_x y\| \leq K \|A P_n y\|.
\]
To prove (ii) it is sufficient to show that $C_n(h)$ converges as $P_n \rightarrow I$ strongly since $T_n \rightarrow T$ as $P_n \rightarrow I$ strongly.

Hence, without loss of generality choosing $n > m$,
\[
\int_H |e^{[h, T_n x]} - e^{[h, T_m x]}| \, d\mu(x)
\]
\[
\leq \int_H |[h, (T_n - T_m)x]| \, d\mu(x)
\]
\[
\leq \int_H |[h, (P_n - P_m)x]| \, d\mu(x) + \int_H |[h, P_n N(P_n x) - P_m N(P_m x)]| \, d\mu(x)
\]
\[
\leq \|(P_n - P_m)h\| (1 + K\|A\|_2) + K\|h\| \|A(P_n - P_m)\|_2
\]
where we have used the fact that $|e^x - e^y| \leq |x - y| \frac{1}{2} (e^x + e^y)$ and the fact that
\[
\|P_n N(P_n x) - N(P_m x)\| = \left\| \int_0^1 P_n N P_n x + (1 - t)(P_n - P_m)x \right\| \, d\mu(x)
\]
\[
\leq K \|A(P_n - P_m)\|
\]
and
\[
\int_H \|A(P_n - P_m)x\| \, d\mu(x) \leq \|A(P_n - P_m)\|_2.
\]
Hence, the terms on the r.h.s. above go to 0 as $P_n, P_m \rightarrow I$ strongly implying that $C_n(h)$ converges and the proof is complete.

**Lemma 5.3.** Let $T_n = S_n^{-1}$ with $S_n$ defined by (4.3) and $M$ satisfy the assumptions above. Then $\mu_{T_n}$ is absolutely continuous w.r.t. $\mu_n - \mu_{P_n h}$ and
\[
\frac{d\mu_{T_n}}{d\mu_n}(P_n h) = f_n(P_n h)
\]
\[
f_n(P_n h) = \exp \{[P_n M(P_n h), P_n h] - \frac{1}{2} \|P_n M(P_n h)\|^2
\]
\[
+ \frac{1}{2} \text{Tr} \log(I - P_n M_{P_n h} P_n)(I - P_n M_{P_n h}^* P_n)\}.
\]
Proof. The proof of this lemma follows from the Jacobian formula for Gaussian measures on finite-dimensional spaces.

Note that for any finite-dimensional invertible map defined by

\[ L = I - P_n Q P_n : P_n H \to P_n H \]

define \( \mu_L \cdot (B) = \mu_n(L(B)) \) where \( B \) is a Borel set in \( P_n H \). Then from the Jacobi formula in finite dimensions we have for any bounded measurable function on \( P_n H \)

\[
\int_{P_n H} F(y) \, d\mu(y) = \int_{P_n H} F(Lx) J^L(x) \, d\mu_n(x)
\]

or equivalently \( \mu_{L^{-1}} \ll \mu_n \) and \( (d\mu_{L^{-1}}/d\mu_n)(x) = J^L(x) \) and furthermore

\[ J^L(x) = \exp\left[\frac{1}{2} \| P_n Q(x) \|_2^2 \right] \left| \det(I_n - P_n Q P_n)(I_n - P_n Q^* P_n) \right|^{1/2} \]

for \( x \in P_n H \). By lemma 5.2, \( T_n \) is invertible and, hence, identifying \( T_n^{-1} \) with \( L \) and from the assumptions on \( M \) (i.e. \( M \) is Volterra, H.S.) we have (see Gross [7, lemma 4.1])

\[ \log \det(I_n - (P_n M P_n)_x)(I_n - (P_n M P_n)_x^*) = \text{Tr} \log((I_n - (P_n M P_n)_x)(I_n - (P_n M P_n)_x^*)). \]

Noting that \( (P_n M P_n)_x = P_n M P_n P_n \) we obtain the required result.

Proof of theorem 5.1. From lemma 5.3 we have

\[ \frac{d\mu_{T_n}}{d\mu_n}(P_n h) = f_n(P_n h). \]

Let

\[ g(P_n h) = \exp\left[M(P_n h), P_n h\right] - \frac{1}{2}\|M(P_n h)\|^2 - \frac{1}{2} \text{Tr}(M_{P_n h} + M_{P_n h}^*). \]

Then in order to complete the proof we need to show that

\[ \int_H |f_n(P_n h) - g(P_n h)| \, d\mu(h) \to 0 \]

as \( P_n \to I \) strongly. First note that since \( M \) is Hilbert–Schmidt, Volterra we have (see [5, Appendix])

\[ \det(I - M_x)(I - M_x^*) = \text{Trace} \log(I - M_x)(I - M_x^*) = -\text{Trace}(M_x + M_x^*). \]

Hence,

\[ \int_H |f_n(P_n h) - g(P_n h)| \, d\mu(h) = \int_H f_n(P_n h)|1 - \phi(P_n h)| \, d\mu(h) \]

where

\[ \phi(P_n h) = \exp\left[-\frac{1}{2}\|I - P_n M(P_n h)\|^2 - \frac{1}{2} \text{Trace} \log((I - M_{P_n h})(I - M_{P_n h}^*)) \right. \]

\[ \left. - \text{Trace} \log(I - P_n M_{P_n h} P_n)(I - P_n M_{P_n h}^* P_n)\right]. \]

Define

\[ C_n = M_{P_n h} M_{P_n h}^* - (M_{P_n h} + M_{P_n h}^*) \]
and

\[ B_n = P_n M_{P_n h} P_n M_{P_n h}^* P_n - P_n (M_{P_n h} + M_{P_n h}^*) P_n. \]

Then both \( C_n \) and \( B_n \) are nuclear and the third term in the exponent of \( \phi(P_n h) \) can be written as

\[
\text{Trace } \log(I + C_n) - \text{Trace } \log(I + B_n).
\]

Now

\[
|\text{Trace } \log(I + C_n) - \text{Trace } \log(I + B_n)| \leq \| \log(I + C_n) - \log(I + B_n) \|_1.
\]

Furthermore by the definition, \( \log(I + C) = \int_0^1 C(I + sC)^{-1} \, ds \), it can be seen that

\[
\| \log(I + C) \|_1 \leq \| (I + C)^{-1} \| \| C \|,
\]

provided \( (I + C)^{-1} \) exists and is bounded. Noting that since \( M_{P_n h} \) is Hilbert–Schmidt and Volterra, and both \( (I - M_{P_n h}) \) and \( (I - P_n M_{P_n h} P_n) \) are invertible,

\[
\| (I + C_n)^{-1} \| = \| (I - M_{P_n h}^*)^{-1} (I - M_{P_n h})^{-1} \|
\]

\[
\leq \| (I - M_{P_n h})^{-1} \| \| (I - M_{P_n h}^*)^{-1} \|
\]

\[
\leq \exp((1 + \| A \|_2^2)).
\]

This follows from [20, p. 1039] since if \( N \) is a Hilbert–Schmidt and quasi-nilpotent operator

\[
\| (I - N)^{-1} \| \leq \exp \frac{1}{2} \| N \|_2^2
\]

and noting that \( \| M_{P_n h} \|_2 = \| M_{P_n h}^* \|_2 \leq \| A \|_2 \). Similarly by using the result obtained in lemma 5.2 we have

\[
\| (I + B_n)^{-1} \| \leq K^2
\]

where \( K \) is defined in lemma 5.2.

From the definition of \( \log(I + C) \) it can be shown that (see Gross [7, lemma 4.1])

\[
\| \log(I + B_n) - \log(I + C_n) \|_1 \leq \| (I + C_n)^{-1} \| \| C_n - B_n \|_1
\]

\[
+ \| (I + B_n)^{-1} \| \| (I + C_n)^{-1} \| \| C_n - B_n \|_1
\]

and, hence, we can write

\[
\| \log(I + C_n) - \log(I + B_n) \|_1 \leq K' \| C_n - B_n \|_1
\]

where \( K' = 2e^{1+\| A \|_2^2} \).

Now the assumption on the Fréchet derivative of \( M \) implies that \( M \) is u.s.c.a.o. and continuity of the Fréchet derivative as a mapping from \( x \) to the Hilbert space of operators implies that

\[ M_n = f_{Lh} \circ L \]

where \( M(h) = f(Lh) \) for some \( f: H \to H \). Hence, it follows that (from the definition of the Fréchet derivative)

\[ M_{P_n h} = f_{Lh} \circ LP_n \]

and

\[ (P_n M(P_n h) P_n)_x = P_n f_{Lh} \circ LP_n \]

implying that both \( f_n(P_n h) \) and \( \phi(P_n) \) can be written as

\[ f_n(P_n h) = a_n(LP_n h) \]
and 

\[ \phi(P_n h) = b(LP_n h) \]

where \( a(\cdot) \) and \( b(\cdot) \) are continuous mappings from \( H \to \mathfrak{R} \). In view of the fact that 

\[ \|B_n - C_n\|_1 \leq 2(K_1 + \|A\|_2) \]

it is easy to see that \( \phi(P_n h) \) is uniformly bounded.

From the definition of \( f_n(P_n h) \) and using the fact that \( |1 - e^t| \leq |x|((1 + e^t)/2) \)

\[ \int_{H} |1 - \phi(P_n h)| f_n(P_n h) \, d\mu(h) = \int_{H} |1 - \phi(P_n h)| f_n(P_n h) \, d\mu(P_n h) \]

\[ = \int_{H} |1 - \phi(P_n h)| \, d\mu_{\tau_n}(P_n h) \]

\[ \leq K_2 \int_{H} (\|(B_n - C_n)\|_1 + \|(I - P_n) M(P_n h)\|^2) \, d\mu_{\tau_n}(P_n h) \]

where \( K_2 = (1 + K_3)K \) and \( K_3 \) is the bound on \( \phi(\cdot) \). The result follows provided 

\[ \|(B_n - C_n)\|_1 \to 0 \text{ as } P_n \to I \text{ strongly under } \mu_{\tau_n}. \]

Now

\[ \|(B_n - C_n)\|_1 \leq \|(M_{P_n h} + M_{P_n h}^* - P_n(M_{P_n h} + M_{P_n h}^*)P_n)\|_1 \]

\[ + \|(M_{P_n h} M_{P_n h}^* - P_n M_{P_n h} P_n M_{P_n h}^*)P_n)\|_1 \]

and since \( B_n \) and \( C_n \) are self-adjoint we have for any CONS \( \{e_i\} \)

\[ \|(M_{P_n h} + M_{P_n h}^* - P_n(M_{P_n h} + M_{P_n h}^*)P_n)\|_1 \]

\[ = \sum_{i=1}^{\infty} \|([M_{P_n h} + M_{P_n h}^*]e_i, e_i] - [P_n(M_{P_n h} + M_{P_n h}^*)P_n e_i, e_i]) \|

\[ = \sum_{i=1}^{\infty} \|([M_{P_n h} + M_{P_n h}^*(I - P_n)e_i, P_n e_i] + [M_{P_n h} + M_{P_n h}^*]e_i, (I - P_n)e_i] \|

\[ \leq 2\|\sqrt{(M_{P_n h} + M_{P_n h}^*)(I - P_n)}\|_2 \|P_n \sqrt{(M_{P_n h} + M_{P_n h}^*)}\|_2 \]

\[ = 2\|M_{P_n h} + M_{P_n h}^*\|_1 \|M_{P_n h} + M_{P_n h}^*\|_1 \]

which goes to 0 as \( P_n \to I \) strongly. Note in the above we have used \( \|AB\|_1 \leq \|A\|_2\|B\|_2 \) for \( A, B \) Hilbert–Schmidt. Similarly,

\[ \|(M_{P_n h} M_{P_n h}^* - P_n M_{P_n h} P_n M_{P_n h}^*)P_n)\|_1 \leq \sum_{i=1}^{\infty} \|([M_{P_n h} M_{P_n h}^* e_i, e_i] - [P_n M_{P_n h} P_n e_i, P_n M_{P_n h} P_n e_i]) \|

\[ \leq 3\|M_{P_n h}(I - P_n)\|_2 \|M_{P_n h}^*\|_2 \]

\[ \leq 3\|\|A\|_2\|A(I - P_n)\|_2 \]

which also goes to zero as \( P_n \to I \) strongly. Let us also note that

\[ \int_{H} \|(B_n - C_n)\|_1 \, d\mu_{\tau_n}(h) \leq K_1 \int_{H} \|(M_{P_n h} + M_{P_n h}^*)(I - P_n)\|_1 \, d\mu_{\tau_n}(h) + 3\|\|A\|_2\|A(I - P_n)\|_2 \]

\[ \leq 2(K_1)^2 + 3\|\|A\|_2 \| < \infty. \]
Using the fact that $\int_H G(T_n h) \, d\mu_T(h, T_n h) = \int_H G(T_n h) \, d\mu(P_n h)$ and $f_{LT_n h} = f_{L(P_n^* h)}$ and $N(\cdot \cdot)$ is also uniformly $S$-continuous with respect to the semi-norm corresponding to $A^* A$ as is $M(\cdot \cdot)$

with appropriately defined $\bar{f}$. Hence, by passing over to the countably additive measure $\mu \circ L^{-1}$ the proof can be completed by using dominated convergence as in lemma 4.3. The second term $\int_H \| (I - P_n) M(T_n h) \|^2 \, d\mu(h)$ can also be shown to go to zero since by the mean value theorem

$$
\| M(T_n h) - M(0) \|^2 \leq \| A(I + P_n H P_n)(P_n h) \|^2 \\
\leq 2 \| A P_n h \|^2 + 2 \| A \|_2^2 \| D P_n h \|^2 \\
\leq 2(1 + c \| A \|_2^2) \| A P_n h \|^2
$$

where $c = \exp(1 + \| A \|_2^2)$ and $c^{1/2} A = B$ the bound on the Fréchet derivative of $K$. Hence, $M(T_n h)$ is uniformly $S$-continuous and the rest follows from lemma 4.3 since

$$
\int_H \| (I - P_n) M(T_n h) \|^2 \, d\mu(h) \leq 4(1 + c) \text{Trace}(A A^*) + 2 \| M(0) \|^2 < \infty.
$$

Also note that $g(P_n h)$ converges to $g(h)$ in $L_1(\mu)$ by noting that

$$
|e^x - e^y| \leq \frac{|x - y|}{2} (e^x + e^y)
$$

(and the exponential terms are bounded), by theorem 4.2.

6. DIFFERENTIAL EQUATIONS WITH WHITE NOISE INPUTS

We now apply the results in Section 5 to obtain the Radon–Nikodym formula for the cylindrical measure induced by the process which satisfies a nonlinear differential equation driven by white noise. This is the white noise version of the Prohorov formula [22] and plays an important role in parameter estimation by the maximum likelihood method.

Consider the following nonlinear differential equation

$$
\dot{x}_t = f(x_t) + n_t \quad \text{a.e. in } t; \ 0 \leq t \leq T
$$

with $x_0$ given and $n_t$ is white noise in $H = L_2[0, T]$.

We will now study the cylindrical measure induced by the process $\dot{x}$ and then obtain the corresponding Radon–Nikodym formula. First we impose the following assumption on $f(\cdot)$

(A1) Let $f(\cdot)$ be $C^1$ and $f''(\cdot)$ be bounded.

We can then state the following theorem.

**Theorem 6.1.** Under the condition (A1) the process $\dot{x}$ induces a cylindrical measure, denoted $\mu_\dot{x}$ on $H$, which is absolutely continuous w.r.t. Gauss measure on $H$ and the Radon–Nikodym derivative is given by

$$
\frac{d\mu_\dot{x}}{d\mu}(h) = \exp \left( \left[ f(L h), h \right] - \frac{1}{2} \| f(L h) \|^2 - \frac{1}{2} \int_0^T f'( \left( \int_0^s h_s ds \right) ) \, ds \right) dt \right) \right)
$$
where \([\cdot, \cdot]\) denotes the inner product on \(L^2[0, T]\) with \(\|\cdot\|\) as the corresponding norm and \(L\) is a linear Volterra operator from \(H\) into \(H\) given by

\[
(Lh)(t) = g_t; \quad g_t = \int_0^t h_s \, ds.
\]

**Proof.** First note that the equation (6.1) can be written as

\[
\dot{x}_t = f((Lx)(t)) + n_t
\]

where

\[
(Lh)(t) = g_t; \quad g_t = \int_0^t h_s \, ds
\]

and \(L\) is a Hilbert–Schmidt, Volterra operator mapping \(L^2[0, T]\) into itself. Hence, \(Lh\) is a p.r.v. For simplicity we use the notation \(L_t h\) to denote \((Lh)(t)\). Define the mapping \(M(h) = f(Lh)\). Since \(f(\cdot)\) is \(C^1\) simple calculation shows that the Fréchet derivative of \(M\) is given by

\[
(M_h g)(t) = f'(L_t h) g; \quad f, g \in H
\]

where \(f'(x) = df(x)/dx\). Since \(f'\) is assumed to be bounded it implies that

\[
\|M_h g\| \leq C \|Lg\|
\]

and \(L\) is Hilbert–Schmidt. Moreover, \(M_h + M_h^*\) is just given by

\[
(M_h + M_h^*)g = \int_0^T f'(L_t h) g_t \, dt
\]

and, hence,

\[
\text{Trace}(M_h + M_h^*) = \int_0^T f'(L_t h) \, dt
\]

implying that \(M_h + M_h^*\) is trace class with \(\|(M_h + M_h^*)\|_1 \leq CT\). This also implies that \(M_h\) defines a Volterra, H.S. operator from \(L^2[0, T] \rightarrow L^2[0, T]\). Hence, we can write

\[
\dot{x}_t = [(I - f \circ L)^{-1} n](t) = [(I - M)^{-1} n](t)
\]

where

\[
(Mh) = g; \quad g_t = f(L_t h).
\]

Therefore, by theorem 5.1 \(\mu_\dot{x}\) is absolutely continuous w.r.t. \(\mu\) on \(H\) and substituting \(f(Lh)\) for \(M(\cdot)\) in (5.1)

\[
\frac{d\mu_\dot{x}}{d\mu}(h) = \exp \left\{ [f(Lh), h] - \frac{1}{2} \|f(Lh)\|^2 - \frac{1}{2} \int_0^T f'(L_t h) \, dt \right\}
\]

and the proof is complete.
From the preceding theorem we obtain the likelihood functional of the process \( \{x_t, 0 \leq t \leq T\} \) to be

\[
\frac{d\mu_x}{d\mu}(x) = \exp \left\{ \frac{1}{2} \left[ \int_0^T f(x_t)^2 \, dt - 2 \int_0^T f(x_t) \dot{x}_t \, dt + \int_0^T f'(x_t) \, dt \right] \right\}.
\] (6.3)

It is illustrative, at this point, to compare the above expression for the likelihood functional with the corresponding result for the usual Wiener process model. The latter formula obtained for the first time by Prohorov [22], is given by

\[
\frac{d\mu_x}{d\mu_W}(x) = \exp \left\{ -\frac{1}{2} \left[ -2 \int_0^T f(x_t) \, dx_t + \int_0^T |f(x_t)|^2 \, ds \right] \right\}
\] (6.4)

where \( \mu_x \) is the measure induced on \( C[0, T] \) by the process \( \{x_t, 0 \leq t \leq T\} \) satisfying the Ito stochastic differential equation

\[
dx_t = f(x_t) + dW_t
\] (6.5)

where the solution exists in the strong sense. \( \mu_W \) denotes the Wiener measure on \( C[0, T] \). The first integral in the exponential expression on the right-hand side of (6.4) is an Ito stochastic integral.

Let us analyse the difference between the expressions (6.3) and (6.4). In the white noise formulation, the likelihood functional is the exact limit of the corresponding result for band-limited processes, as the bandwidth increases without bound. On the other hand, as Wong and Zakai [23] showed in the late 1960s, the limit of integrals for band-limited processes converge to the corresponding stochastic integral only when corrected by a suitable term. In fact, from the result of Wong and Zakai, one gets

\[
\left[ \int_0^T f(x_t) \, dx_t \right] = \lim_{k \to \infty} \left[ \int_0^T f(x_t^k) \dot{x}_t^k \, dt - \frac{1}{2} \int_0^T f'(x_t^k) \, dt \right]
\] (6.6)

where \( \{x_t^k, 0 \leq t \leq T\} \) satisfies

\[
\frac{d\dot{x}_t^k}{dt} = f(x_t^k) + n_t^k
\]

with \( \{n_t^k, 0 \leq t \leq T\} \) approaching a Gaussian white noise in the limit (see Wong and Hajek [24] for details). This explains the difference between the expressions.

The advantage of working with the direct model of white noise becomes apparent when one has to handle real data. The process \( \{x_t, 0 \leq t \leq T\} \) which is the solution of (6.5) is only an idealization. The real data that one observes in practice cannot be nondifferentiable everywhere, as the model (6.5) implies. Every measurement apparatus acts as a filter on the idealized data \( \{\dot{x}_t, 0 \leq t \leq T\} \). When one has to evaluate the stochastic integral \( \int_0^T f(x_t) \, dx_t \), \( dx_t \) is replaced by \( \dot{x}_t \, dr \), where \( \dot{x}_t \) is calculated from the real data. In fact, for real data, \( \dot{x}_t \) is band-limited, albeit with large bandwidth. But replacing \( dx_t \) by \( \dot{x}_t \, dr \) will give us the wrong result and will have to be corrected by the Wong–Zakai correction term \( -\frac{1}{2} \int_0^T f'(x_t) \, dt \) in our case. In the white noise model this term already appears in the expression for the likelihood ratio and no such correction is necessary when handling real data. Thus, although the result can be obtained by using other methods such as an Ito to Stratonovich conversion the robustness is inherent in the white noise formulation.
Acknowledgement—The authors wish to thank Professor A. V. Balakrishnan of UCLA for his constant encouragement and fruitful discussions on white noise theory which led to these results. The research of RM was supported in part by a grant from NSERC OGP 0042024.

REFERENCES