Open Access

SCHOLARENA

Journal of Mathematical and Statistical Analysis

RESEARCH ARTICLE

ISSN: 2642-018X

Conditioning Gaussian Measure on Hilbert Space

Owhadi H and Scovel C*

Computing and Mathematical Sciences, California Institute of Technology, Pasadena, United States

***Corresponding author:** Scovel C, Computing and Mathematical Sciences, California Institute of Technology, Pasadena, United States, Tel: 5054557267, E-mail: clintscovel@gmail.com

Citation: Owhadi H, Scovel C (2018) Conditioning Gaussian Measure on Hilbert Space. J Math Stat Anal 1: 109

Article history: Received: 19 June 2018, **Accepted:** 19 September 2018, **Published:** 21 September 2018

Abstract

For a Gaussian measure on a separable Hilbert space with covariance operator C, we show that the family of conditional measures associated with conditioning on a closed subspace $S^{\perp}$ are Gaussian with covariance operator the short $S(C)$ of the operator C to S. Although the shorted operator is a well-known generalization of the Schur complement, this fundamental generalization to infinite dimensions of the well-known relationship between the Schur complement and the covariance operator of the conditioned Gaussian measure is new. Moreover, the conditioning of infinite dimensional Gaussian measures appears in many fields so that this simply expressed result appears to unify and simplify these efforts.

We provide two proofs. The first uses the theory of Gaussian Hilbert spaces and a characterization of the shorted operator by Andersen and Trapp. The second uses recent developments by Corach, Maestripieri and Stojanoff on the relationship between the shorted operator and C-symmetric projections onto $S^{\perp}$. To obtain the assertion when such projections do not exist, we develop an approximation result for the shorted operator by showing, for any positive operator A, how to construct a sequence of approximating operators A^n which possess A^n-symmetric oblique projections onto $S^{\perp}$ such that the sequence of shorted operators $S(A^n)$ converges to $S(A)$ in the weak operator topology. This result combined with the martingale convergence of random variables associated with the corresponding approximations C^n establishes the main assertion in general. Moreover, it in turn strengthens the approximation theorem for shorted operator when the operator is trace class; then the sequence of shorted operators $S(A^n)$ converges to $S(A)$ in trace norm.

Keywords: Conditioning; Gaussian Measure; Hilbert Space; Shorted Operator; Schur; Oblique Projection; Infinite Dimensions

AMS subject classifications: 60B05, 65D15

Introduction

For a Gaussian measure μ with injective covariance operator C on a direct sum of finite dimensional Hilbert spaces $H = H_1 \oplus H_2$, the conditional measure associated with conditioning on the value of H_2 can be computed in terms of the Schur complement corresponding to the partitioning of the covariance matrix C. Evidently, the natural extension to infinite dimensions of the Schur complement is the shorted operator, first discovered by Krein and developed in Anderson and Trapp [1,2]. However, the connection between the shorted operator and the covariance operator of the conditional Gaussian measure on an infinite dimensional Hilbert space appears yet to be established. Indeed, Lemma 4.3 of Hairer, Stuart, Voss, and Wiber, see also Stuart, characterizes the conditional measure through a measurable extension result of Dalecky and Fomin of an operator defined on the CameronMartin reproducing kernel Hilbert space [3-5]. For other representations, see Mandelbaum, and Tarieladze and Vakhania's extension of the optimal linear approximation results of Lee and Wasilkowski from finite to infinite rank, extending results in the Information-Based Complexity of Traub, Wasilkowski and Wozniakowski [6-10].

The primary purpose of this paper is to demonstrate that, for a Gaussian measure with covariance operator C, the covariance operator of the Gaussian measure obtained by conditioning on a subspace is the short of C to the orthogonal complement of that subspace. We provide two distinct proofs. The first uses the theory of Gaussian Hilbert spaces and a characterization of the shorted operator by Andersen and Trapp. The second proof, corresponding to the secondary purpose of this paper, uses recent developments by Corach, Maestripieri and Stojanoff on the relationship between the shorted operator and A-symmetric oblique projections. This latter approach has the advantage that it facilitates a general approximation technique that not only can be used to approximate the covariance operator but the conditional expectation operator. This is accomplished through the development of an approximation theory for the shorted operator in terms of oblique projections followed by an application of the martingale convergence theorem. Although the proofs are not fundamentally difficult, the result (which appears to have been missed in the literature) provides a simple characterization of the conditional measure, leading to significant approximation results. For instance, the attainment of the main result through the martingale approach feeds back a strengthening of the approximation theorem for the shorted operator that was developed for that purpose: when the operator is trace class the approximation improves from weak convergence to convergence in trace norm.

Conditioning Gaussian measures has applications in Information-Based Complexity and, beginning with Poincare, publications by e.g. Diaconis, Sul'din, Larkin, Sard, Kimeldorf and Wahba, Shaw, and Hagan they have been useful in the development of statistical approaches to numerical analysis [10-18]. Although they received little attention in the past, the possibilities offered by combining numerical uncertainties/errors with model uncertainties/errors are stimulating the reemergence of such methods and, as discussed in Briol et al. and Owhadi and Scovel, the process of conditioning on closed subspaces is of direct interest to the reemerging field of Probabilistic Numerics where solutions of PDEs and ODEs are randomized and numerical errors are interpreted in a Bayesian framework as posterior distributions [19-31] Furthermore, as shown in Gaussian measures are a class of optimal measures for minmax recovery problems emerging in Numerical Analysis (when quadratic norms are used to define relative errors) and conditioning such measures on finite-dimensional linear projections lead to the identification of scalable algorithms for a wide range of operators[20,32]. Representing the process of conditioning Gaussian measures on closed (possibility infinite dimensional) subspaces via converging sequences of shorted operators, could be used as a tool for reducing/compressing infinite-dimensional operators and identifying reduced models. In particular, it is shown in that the underlying connection with Schur complements can be exploited to invert and compress dense kernel matrices appearing in Machine Learning and Probabilistic Numerics in near linear complexity, thereby opening the complexity bottleneck of kernel methods [31].

Let us review the basic results on Gaussian measures on Hilbert space. A measure μ on a Hilbert space H is said to be Gaussian if, for each $h \in H$ considered as a continuous linear function $h: H \to R$ by $h(x) := \langle h, x \rangle, x \in H$, we have that the pushforward measure $h_*\mu$ is Gaussian, where we say that a Dirac measure is Gaussian. For a Gaussian measure μ, its mean m is defined by

$$(h,m) := \int_H \langle h, x \rangle d\mu(x), \qquad h \in H$$

and its covariance operator $C : H \to H$ is defined by

$$\langle Ch_1, h_2 \rangle := \int_H \langle h_1, x \rangle \langle h_2, x \rangle d\mu(x) - \langle h_1, m \rangle \langle h_2, m \rangle, \qquad h_1 h_2 \in H.$$

A Gaussian measure has a well defined mean and a continuous covariance operator, see e.g. Da Prato and Zabczyk [33]. Mourier's Theorem, see Vakhania, Tarieladze and Chobanyan, asserts, for any $m \in H$ and any positive symmetric trace class operator C, that there exists a Gaussian measure with mean m and covariance operator C, and that all Gaussian measures have a well defined mean and positive trace class covariance operator. This characterization also follows from Sazonov's Theorem [34].

Since separable Hilbert spaces are Polish, it follows from the product space version, see e.g. Dudley, of the theorem on the existence and uniqueness of regular conditional probabilities on Polish spaces, that any Gaussian measure μ on a direct sum $H = H_1 \oplus H_2$, of separable Hilbert spaces has a regular conditional probability, that is there is a family μ_t, $t \in H_2$ of conditional measures corresponding to conditioning on H_2. Moreover, Tarieladze and Vakhania demonstrate that the corresponding family of conditional measures are Gaussian[8,35]. Bogachev's theorem of normal correlation of Hilbert space valued Gaussian random variables shows that if two Gaussian random vectors ξ and η on a separable Hilbert space H are jointly Gaussian in the product space, then E[ξ|η] is a Gaussian random vector and $\xi = E[\xi|\eta] + \zeta$ where ζ is Gaussian random vector which is independent of η. Consequently, for any two vectors $h_1, h_2 \in H$ we have

$$\begin{aligned} \mathrm{E}\big[\langle \xi - \mathrm{E}[\xi|\eta], h_1 \rangle \langle \xi - \mathrm{E}[\xi|\eta], h_2 |\eta\big] &= \mathrm{E}\big[\langle \varsigma, h_1 \rangle \langle \varsigma, h_2 \rangle |\eta\big] \\ &= \mathrm{E}\big[\langle \varsigma, h_1 \rangle \langle \varsigma, h_2 \rangle\big] \end{aligned}$$

and so we conclude that, just as in the finite dimensional case, the conditional covariance operators are independent of the values of the conditioning variables [36].

Since both proof techniques will utilize the characterization of conditional expectation as orthogonal projection, we introduce these notions now. Consider the Lebesgue Bochner space $L^2(H, \mu, \mathcal{B}(H))$ space of (equivalence classes) of H-valued Borel measurable functions on H whose squared norm

$$\|f\|^2_{L^2(H,\mu,\mathcal{B}(H))} := \int_H \|f(x)\|^2_H \, d\mu(x)$$

is integrable. For a sub σ-algebra $\Sigma \subset \mathcal{B}(H)$ of the Borel σ-algebra, consider the corresponding Lebesgue-Bochner space $L^2(H,\mu,\Sigma)$. As in the scalar case, one can show that $L^2(H,\mu,\mathcal{B}(H))$ and $L^2(H,\mu,\Sigma)$ are Hilbert spaces and that $L^2(H,\mu,\Sigma) \subset L^2(H,\mu,\mathcal{B}(H))$ is a closed subspace. Then, if we note that contractive projections on Hilbert space are orthogonal, it follows from Diestel and Uhl that conditional expectation amounts to orthogonal projection [37].

Shorted Operators

A symmetric operator $A: H \to H$ is called positive if $\langle Ax,x\rangle \geq 0$ for all $x \in H$. We denote by $L_+(H)$ the set of positive operators and we denote such positivity by $A \geqq 0$. Positivity induces the (*L*oewner) partial order $\geqq$ on $L_+(H)$. For a closed subspace $S \subset H$ and a positive operator $A \in L_+(H)$ consider the set

$$\mathrm{H}(A,S) := \left\{ X \in L+(H) : X \leq A \text{ and } R(X) \subset S \right\}.$$

Then Krein and later Anderson and Trapp showed that *H(A,S)* contains a maximal element, which we denote by *S(A)* and call the short of *A* to *S*. For another closed subspace $T \subset H$, we denote the short of *A* to *T* by *T (A)*. In the proof, Anderson and Trapp [1] demonstrate that when *A* is invertible, that in terms of its $(S,S^{\perp})$ partition representation [1,2].

$$A = \begin{pmatrix} A_{SS} & A_{SS^{\perp}} \\ A_{S^{\perp}S} & A_{S^{\perp}S^{\perp}} \end{pmatrix}$$

that is $A_{S^{\perp}S^{\perp}}$ invertible and

$$S(A) = \begin{pmatrix} A_{SS} - A_{SS^{\perp}} A_{S^{\perp}S^{\perp}}^{-1} A_{S^{\perp}S} & 0 \\ 0 & 0 \end{pmatrix}.$$

It is easy to show that the assertion holds under the weaker assumption that $A_{S^{\perp}S^{\perp}}$ be invertible. Moreover, Anderson and Trapp asserts for $A,B \in L+(H)$, that

$$A \leq B \Rightarrow S(A) \leq S(B)$$

that is, *S* is monotone in the *L*oewner ordering, and for two closed subspaces *S* and *T*, we have

$$(S \cap T)(A) = S(T(A)).$$

Finally, Theorem 6 of Anderson and Trapp asserts that if $A: H \to H$ is a positive operator and $S \subset H$ is a closed linear subspace, then

$$\langle S(A)s,s\rangle = \inf\left\{\left\langle A\binom{s}{t},\binom{s}{t}\right\rangle, t \in S^{\perp}\right\}, \qquad \forall_s \in S. \tag{2.1}$$

In Section 4.1 we demonstrate how the characterization (2.1) of the shorted operator combined with the theory of Gaussian Hilbert spaces provides a natural proof of our main result, the following theorem. Here we consider direct sum split $H = H_1 \oplus H_2$, and let $S = H_1$ and $S^{\perp} = H_2$, so that the short *S(A)* of an operator to the subspace $S = H_1$ will be written as $H_1(A)$.

Theorem 2.1. Consider a Gaussian measure μ on an orthogonal direct sum $H = H_1 \oplus H_2$ of separable Hilbert spaces with mean *m* and covariance operator *C*. Then for all $t \in H_2$, the conditional measure μt is a Gaussian measure with covariance operator $H_1(C)$.

Oblique Projections

In this section, we will prepare for an alternative proof of Theorem 2.1 using oblique projections along with the development of approximations of the covariance operator and the conditional expectation operator generated by natural sequences of oblique projections. To that end, let us introduce some notations. For a separable Hilbert space *H*, we denote the usual, or strong, convergence of sequences by $h_n \to h$ and the weak convergence by $h_n \xrightarrow{\omega} h$ Let *L(H)* denote the Banach algebra of bounded linear operators on *H*. For an operator $A \in L(H)$, we let *R(A)* denote its range and ker*(A)* denote its nullspace. Recall the uniform operator topology on *L(H)* defined by the metric $\|A\| := \sup_{\|h\|\leq 1}\|Ah\|$. We say that a sequence of operators $A^n \in L(H)$ converges strongly to $A \in L(H)$, that is

$$A = s - \lim_{n\to\infty} A^n$$

if $A^n h \to Ah$ for all $h \in H$, and we say that $A^n \to A$ weakly or

$$A = w - \lim_{n\to\infty} A^n$$

if $A^n h \xrightarrow{\omega} Ah$ for all $h \in H$. Recall that an operator $A \in L(H)$ is called trace class if the trace norm

$$\|A\|_1 := \sum_{i=1}^{\infty} \langle |A| e_i, e_i \rangle$$

is finite for some orthonormal basis, where $|A| := \sqrt{A^*A}$ is the absolute value. When it is finite, then $tr(A) := \sum_{i=1}^{\infty} \langle Ae_i, e_i \rangle$ is well defined, and for all positive trace class operators A we have $tr(A) = \|A\|_1$. The trace norm $\|.\|_1$ makes the subspace $L_1(H) \subset L(H)$ of trace class operators into a Banach space. It is well known that the sequence of operator topologies weak→strong→ uniform operator→trace norm

increases from left to right in strength.

For a positive operator A: $H \to H$, let us define the set of (A-symmetric) oblique projections

$$P(A,S^{\perp}) := \{Q \in L(H) : Q^2 = Q, R(Q) = S^{\perp}, AQ = Q^* A\}$$

onto $S^{\perp}$, where Q^* is the adjoint of Q with respect to the scalar product $\langle\cdot,\cdot\rangle$ on H. The pair $(A, S^{\perp})$ is said to be compatible, or $S^{\perp}$ is said to be compatible with A, if $P(A, S^{\perp})$ is nonempty. For any oblique projection $Q \in P(A, S^{\perp})$, Corach, Maestripieri and Stojanoff asserts that for $E := 1-Q$, we have

$$S(A) = AE = E^* AE. \tag{3.1}$$

Moreover,when $(A,S^{\perp})$ is compatible, according to Corach, Maestripieri and Stojanoff,there is a special element $Q_{A,S^{\perp}} \in P(A,S^{\perp})$ defined in the following way: by their Proposition 3.3 and their factorization Theorem 2.2 there is a unique operator $\hat{Q}: S \to S^{\perp}$ which satisfies $A_{S^{\perp}S^{\perp}}\hat{Q} = A_{S^{\perp}S}$ such that ker $\hat{Q} = \ker(A_{S^{\perp}S})$ and $R(\hat{Q}) \subset \overline{R(A_{S^{\perp}S^{\perp}})}$ needs overbar Defining

$$Q_{A,S^{\perp}} = \begin{pmatrix} 0 & 0 \\ \hat{Q} & 1 \end{pmatrix}, \tag{3.2}$$

their Theorem 3.5 asserts that $Q_{A,S^{\perp}} \in P(A,S^{\perp})$.

When the pair $(A,S^{\perp})$ is not compatible, we seek an approximating sequence A^n to A which is compatible with $S^{\perp}$, such that the limit of $S(A^n)$ is $S(A)$. Although Anderson and Trapp show that if A^n is a monotone decreasing sequence of positive operators which converge strongly to A, that the decreasing sequence of positive operators $S(A^n)$ strongly converges to $S(A)$, the approximation from above by $A^n := A + \frac{1}{n} I$ determines operators which are not trace class, so is not useful for the approximation problem for the covariance operators for Gaussian measures. Since the trace class operators are well approximated from below by finite rank operators one might hope to approximate A by an increasing sequence of finite rank operators. However, it is easy to see that, in general, the same convergence result does not hold for increasing sequences. The following theorem demonstrates, for any positive operator A, how to produce a sequence of positive operators A^n which are compatible with $S^{\perp}$ such that $S(A^n)$ weakly converges to $S(A)$[2,38].

Henceforth we consider a direct sum split $H = H_1 \oplus H_2$, and let $S = H_1$ and $S^{\perp} = H_2$, so that the short $S(A)$ of an operator to the subspace $S = H_1$ will be written as $H_1(A)$. Let us also denote by P_i: $H \to H$ the orthogonal projections onto H_i, for i= 1,2, and let $\Pi_i : H \to H_i$ denote the corresponding projections and $\Pi_i^* : H_i \to H$ the corresponding injections. For any operator A: $H \to H$, consider the decomposition

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{12} \end{pmatrix}$$

where the components are defined by $A_{ij} := \Pi_i A \Pi_i^*, i, j = 1,2.$

Theorem 3.1. *Consider a positive operator A: $H \to H$ on a separable Hilbert space. Then for any orthogonal split $H = H_1 \oplus H_2$, and any ordered orthonormal basis of H_2, we let H_2^n denote the span of the first n basis elements and let* $P \quad P \quad P$ *denote the orthogonal projection onto $H_1 \oplus H_2^n$. Then the sequence of positive operators*

$$A^n := P^n A P^n, \ n = 1,...$$

is compatible with H_2 and

$$H_1(A) = \omega - \lim_{n\to\infty} H_1(A^n),$$

Remark 3.2. For an increasing sequence A^n of positive operators converging strongly to A, the monotonicity of the shorting operation implies that the sequence $H_1(A^n)$ is increasing, and therefore Vigier's Theorem implies that the sequence $H_1(A^n)$ converges strongly. Although the sequence $A^n := P^n A P^n$ defined in Theorem 3.1 is positive and converges strongly to A, in general, it is not increasing in the Loewner order, so that Vigier's Theorem does not apply, possibly suggesting why we only obtain convergence in the weak operator topology. With stronger assumptions on the operator A and a well chosen selection of an ordered orthonormal basis of H_2, we conjecture that convergence in a stronger topology may be available. In particular, as a corollary to our main result, when A is trace class, we establish in Corollary 3.4 that

$H_1(A^n) \to H_1(A)$ in trace norm.

For any $m \in H$, we let $m = (m_1, m_2)$ denote its decomposition in $H = H_1 \oplus H_2$. Moreover, for any projection Q: $H \to H$ with $R(Q) = H_2$ we let $\hat{Q}^* : H_2 \to H_1$ denote the unique operator such that

$$Q=\begin{pmatrix} 0 & 0 \\ \hat{Q} & 1 \end{pmatrix},$$

and denote by $\hat{Q}^*: H_2 \to H_1$ the adjoint of $\hat{Q}$ defined by the relation $\langle \hat{Q}^* h_2, h_1 \rangle_{H_1} = \langle h_2, \hat{Q}^* h_1 \rangle_{H_2}$ for all $h_1 \in H_1$, $h_2 \in H_2$.

The following theorem constitutes an expansion of our main result, Theorem 2.1, to include natural approximations for the conditional covariance operator and the conditional expectation operator.

Theorem 3.3. *Consider a Gaussian measure μ on an orthogonal direct sum $H = H_1 \oplus H_2$ of separable Hilbert spaces with mean m and covariance operator C. Then for all $t \in H_2$, the conditional measure μ_t is a Gaussian measure with covariance operator $H_1(C)$.*

If the covariance operator C is compatible with H_2, then for any oblique projection Q in $P(C, H_2) \neq \emptyset$, the mean mt of the conditional measure μ_t is

$$m_t = \begin{pmatrix} m_1 + \hat{Q}^*(t-m_2) \\ t \end{pmatrix}.$$

In the general case, for any ordered orthonormal basis for H_2, let H_2^n denote the span of the first n basis elements, let $P^n : P_{H_1} + P_{H_2^N}$ denote the orthogonal projection onto $H_1 \oplus H_2^n$. and define the approximate $C^n := P^n C P^n$. Then C^n is compatible with H_2 for all n, and for any sequence $Q_n \in P(C^n, H_2) \neq \emptyset$ of oblique projections, we have

$$m_t = \begin{pmatrix} m_1 + \lim_{n\to\infty} \hat{Q}_n^*(t - P_{H_2^n} m_2) \\ t \end{pmatrix}$$

for μ-almost every t. If the sequence Q_n eventually becomes the special element $Q_n = Q_{C^n, H_2}$ defined near (3.2), then we have

$$m_t = \begin{pmatrix} m_1 + \lim_{n\to\infty} \hat{Q}_n^*(t - m_2) \\ t \end{pmatrix}$$

for μ-almost every t.

As a corollary to Theorem 3.3, we obtain a strengthening of the assertion of Theorem 3.1 when the operator A is trace class.

Corollary 3.4. *Consider the situation of Theorem 3.1 with A trace class. Then*

$$H_1(A^n) \to H_1(A) \text{ in trace norm}$$

Proofs

First proof of Theorem 2.1 Consider the Lebesgue-Bochner space $L^2(H,\mu,\mathcal{B}(H))$ space of (equivalence classes) of H-valued Borel measurable functions on whose squared norm

$$\|f\|^2_{L^2(H,\mu,\mathcal{B}(H))} := \int_{H'} \|f(x)\|_H^2 \, d\mu(x)$$

is integrable. For any square Bochner integrable function $f \in L^2(H,\mu,\mathcal{B}(H))$ and any $h \in H$, we have that $\langle f, h\rangle$ is square integrable, that is $\langle f,h\rangle \in L^2(R,\mu,\mathcal{B}(H))$. Moreover, it is easy to see that if f is Bochner integrable, then for all $h \in H$, we have $\langle f, h\rangle$ is Bochner integrable and $\int \langle f,h\rangle d\mu = \left\langle \int f d\mu, h\right\rangle$.

Now consider the orthogonal decomposition $H = H_1 \oplus H_2$ and the Borel σ-algebra $\mathcal{B}(H_2)$. Let us denote the shorthand notation

$$\mathcal{B} := \mathcal{B}(\mathrm{H}),\ \mathcal{B}_2 := \{(H_1, T) : T \in \mathcal{B}(\mathrm{H}_2)\}.$$

The definition of conditional expectation in Lebesgue-Bochner space, that is that $E[f|\mathcal{B}_2]$ is the unique μ-almost everywhere $\mathcal{B}_2$ -measurable function such that

$$\int_B E[f|\mathcal{B}_2] d\mu = \int_B f d\mu, \quad B \in \mathcal{B}$$

combined with Hille's theorem [13,Thm. II.6], that for each $h \in H$ we have

$$\left\langle h, \int_B E[f|\mathcal{B}_2] d\mu \right\rangle = \int_B \langle h, E[f|\mathcal{B}_2] \rangle d\mu, \qquad B \in \mathcal{B}_2$$

and

$$\langle h, \int_B f d\mu, = \int_B \langle h, f\rangle d\mu, \qquad B \in \mathcal{B}_2$$

implies that

$$E[\langle h, f\rangle | \mathcal{B}_2] = \langle h, E[f | \mathcal{B}_2]\rangle, \qquad h \in H$$

thus implying the following commutative diagram for all h$\in$ H:

$$\begin{tikzcd} L^2(H,\mu,\mathcal{B}) \arrow[r, "\mathrm{E}[.|\mathcal{B}_2]"] \arrow[d, "\langle h"'] & L^2(H,\mu,\mathcal{B}_2) \arrow[d, "\langle h"'] \\ L^2(R,\mu,\mathcal{B}) \arrow[r, "\mathrm{E}[.|\mathcal{B}_2]"] & L^2(R,\mu,\mathcal{B}_2) \end{tikzcd} \tag{4.1}$$

When μ is a Gaussian measure, the theory of Gaussian Hilbert spaces, see e.g. Jansen, provides a stronger characterization of conditional expectation of the canonical random variable $X(h):=h,h\in H$ when conditioning on a subspace and captures the full linear nature of Gaussian conditioning [39]. Let us assume henceforth that μ is a centered Gaussian measure. Then Fernique's Theorem, see Theorem 2.6 in Da Prato, implies that the random variable X is square Bochner integrable [33]. For any element $h\in H$, let us denote the corresponding function $\xi_h : H \to R$ defined by $\xi_h(h') = \langle h,h'\rangle, h'\in H$. Then the the discussion above shows that for any h$\in$H, that the real-valued random variable ξ_h is square integrable, that is $\xi_h \in L^2(R,\mu,\mathcal{B})$, for all $h\in H$. Let

$$\xi : H \to L^2(R,\mu,\mathcal{B})$$

denote the resulting linear mapping defined by

$$h \to \xi_h \in L^2(R,\mu,\mathcal{B}), h \in H.$$

It is straightforward to show that ξ is injective if and only if the covariance operator C of the Gaussian measure μ is injective. By the definition of a centered Gaussian vector X, it follows that the law $(\xi_h)_*\mu$ in R is a univariate centered Gaussian measure, that is ξ_h is a centered Gaussian real-valued random variable. Consequently, let us consider the closed linear subspace

$$H^\mu := \overline{\xi(H)} \subset L^2(R,\mu,\mathcal{B})$$

generated by the elements $\xi_h \in L^2(R,\mu,\mathcal{B}), h\in H.$ By Theorem I.1.3 of Jansen, this closure $H^\mu \subset L^2(R,\mu,\mathcal{B})$ also consists of centered Gaussian random variables, and since it is a closed subspace of a Hilbert space, it is a Hilbert space and therefore a Gaussian Hilbert space as defined in Jansen [39]. Moreover, by Theorem 8.15 of Jansen, H^μ is a feature space for the Cameron-Martin reproducing kernel Hilbert space with feature map $\xi: H \to H^\mu$ and reproducing kernel the covariance operator. For a closed Hilbert subspace, $H_2 \subset H$, we can consider the closed linear subspace

$$H_2^\mu := \overline{\xi(H_2)} \subset L^2(R,\mu,\mathcal{B}_2)$$

generated by the elements $\xi_{h_2}, h_2 \in H_2$ in the same way. H_2^μ is also a Gaussian Hilbert space and we have the natural subspace identification $H_2^\mu \subset H^\mu$. Since separable Hilbert spaces are Polish, and an orthonormal basis is a separating set, it follows, see e.g. Vakhania, Tarieladze and Chobanyan, that for an orthonormal basis e_i, $i\in I$ of a separable Hilbert space, that the σ-algebra generated by the corresponding real-valued functions $\sigma(\{\xi_{e_i}, i\in I\})$ is the Borel *σ-algebra* of the Hilbert space. Consequently, we obtain from Janson that for any $h\in H$, that

$$E[\xi_h | \mathcal{B}_2] = E\left[\xi_h | \sigma\left(\cup \xi_{h_2}, h_2 \in H_2\right)\right] = P_{H_2^\mu}\xi_h$$

where $P_{H_2^\mu} : H^\mu \to H_2^\mu$ is orthogonal projection. That is, if we let $E[\bullet|\mathcal{B}_2] : L2(R,\mu,\mathcal{B}) \to L2(R,\mu,\mathcal{B}_2)$ be the conditional expectation represented as orthogonal projection and $E[\cdot|\mathcal{B}_2]: H^\mu \to H_2^\mu$ be the conditional expectation represented as orthogonal projection from the linear subspace $H\mu \subset L2(R,\mu,\mathcal{B})$ onto the closed subspace $H_2^\mu \subset H^\mu$, we have the following commutative diagram, where and $i_{H_2^\mu} : H_2^\mu \to L^2(R,\mu,\mathcal{B}_2)$ denote the closed subspace injections [34,39].

$$\begin{tikzcd} L^2(R,\mu,\mathcal{B}) \arrow[r, "\mathrm{E}[.|\mathcal{B}_2]"] \arrow[d, "i_H^{\mu}"'] & L^2(R,\mu,\mathcal{B}_2) \arrow[d, "i_{H2}^{\mu}"'] \\ H^\mu \arrow[r, "\mathrm{E}[.|\mathcal{B}_2]"] & H_2^\mu \end{tikzcd} \tag{4.2}$$

which when combined with Figure 4.1, representing the commutativity of vector projection and conditional expectation, produce the following commutative diagram for all $h \in H$:

$$\begin{tikzcd}
L^2(H,\mu,\mathcal{B}) \arrow[r, "{E[.|\mathcal{B}_2]}"] \arrow[d, "\langle h"'] & L^2(H,\mu,\mathcal{B}_2) \arrow[d, "\langle h"'] \\
L^2(R,\mu,\mathcal{B}) \arrow[r, "{E[.|\mathcal{B}_2]}"] & L^2(R,\mu,\mathcal{B}_2) \\
H^{\mu} \arrow[u, "i_H{}^{\mu}"] \arrow[r, "{E[.|\mathcal{B}_2]}"] & H_2^{\mu} \arrow[u, "i_{H2}^{\mu}"'] \\
H \arrow[u, "\xi"] & H_2 \arrow[u, "\xi"']
\end{tikzcd} \tag{4.3}$$

Although there is a natural projection map $P_{H_2}: H \to H_2$ for the bottom of this diagram, in general it cannot be inserted here and maintain the commutativity of the diagram. This comes from the fact that there may exist an $h \in H$ such that $\xi_h = 0$. However, this does not imply that $\xi_{P_{H_2} h} = 0$.

We are now prepared to obtain the main assertion. The covariance operator of the random variable X is defined by

$$\begin{aligned}\langle Ch, h'\rangle &= E_{\mu}\left[\langle X,h\rangle\langle X,h'\rangle\right] \\ &= E_{\mu}\left[\langle \xi_h \xi_{h'}\right], h,h' \in H.\end{aligned}$$

Moreover, by the theorem of normal correlation and the commutativity of the diagram (4.1), the conditional covariance operator is defined by

$$\begin{aligned}\langle C(X|X_2)h,h'\rangle &= E_{\mu}\left[\langle X - E\left[X|\mathcal{B}_2\right],h\rangle\langle X - E\left[X|\mathcal{B}_2\right],h'\rangle|\mathcal{B}_2\right] \\ &= E_{\mu}\left[\langle X - E\left[X|\mathcal{B}_2\right],h\rangle\langle X - E\left[X|\mathcal{B}_2\right],h'\rangle\right] \\ &= E_{\mu}\left[\left(\xi_h - E\left[\xi_h|\mathcal{B}_2\right]\right)\left(\xi_{h'} - E\left[\xi_{h'}|\mathcal{B}_2\right]\right)\right], \quad h,h' \in H.\end{aligned}$$

In terms of the Gaussian Hilbert spaces $H_2^{\mu} \subset H^{\mu}$, using the commutativity of the diagram (4.2) and the identification of the con ditional expectation with orthogonal projection, we conclude that

$$\langle Ch,h'\rangle = \langle \xi_h, \xi_{h'}\rangle_{H^{\mu}}, \qquad h,h' \in H \tag{4.4}$$

and

$$\langle C(X\,|\,X_2)h,h'\rangle = \left\langle \left(I - P_{H_2^{\mu}}\right)\xi_h, \left(I - P_{H_2^{\mu}}\right)\xi_{h'}\right\rangle_{H^{\mu}}, \; h,h' \in H. \tag{4.5}$$

Since the orthogonal projection $P_{H_2^{\mu}}$ is a metric projection of H^{μ} onto H_2^{μ}, we can express the dual optimization problem to the metric projection as follows: for any $h \in H$, using the decomposition $h=h_1+h_2$ with $h_1 \in H_1, h_2 \in H_2$,
we decompose $\xi_h = \xi_{h_1+h_2} = \xi_{h_1} + \xi_{h_2}$. Then, noting that $(I - P_{H_2^{\mu}})\xi_{h_2} = 0$, we obtain

$$\begin{aligned}\left\|\xi_h\right\|_{H^{\mu}}^2 &= \left\|\xi_{h_1} + \xi_{h_2}\right\|_{H^{\mu}}^2 \\ &= \left\|\left(I - P_{H_2^{\mu}}\right)\left(\xi_{h_1} + \xi_{h_2}\right)\right\|_{H^{\mu}}^2 + \left\|P_{H_2^{\mu}}(\xi_{h_1} + \xi_{h_2})\right\|_{H^{\mu}}^2 \\ &= \left\|\left(I - P_{H_2^{\mu}}\right)\xi_{h_1}\right\|_{H^{\mu}}^2 + \left\|P_{H_2^{\mu}}\xi_{h_1} + \xi_{h_2}\right\|_{H^{\mu}}^2.\end{aligned}$$

Since in the second term on the right-hand side $P_{H_2^\mu}\xi_{h_1} \in H_2^V$, there is a sequence $h_2^n, n = 1,\ldots$ such that the corresponding sequence $\xi_{h_2^n}$ converges to $-P_{H_2^\mu}\xi_{h_1}$ in $L^2(R,\mu,\mathcal{B})$ and therefore H^μ, we conclude that

$$\left\| \left(I - P_{H_2^\mu}\right)\xi_{h_1} \right\|_{H^\mu}^2 = \inf_{h_2 \in H_2} \left\| \xi_{h_1} + \xi_{h_2} \right\|_{H^\mu}^2 .$$

From the identifications (4.4) and (4.5), we conclude that

$$\langle C(X \mid X_2)h_1, h_1 \rangle = \inf_{h_2 \in H_2} \langle C(X)(h_1 + h_2), h_1 + h_2 \rangle, .$$

Therefore, Anderson and Trapp implies the assertion

$$C(X \mid X_2) = H_1(C).$$

The assertion in the non-centered case follows by simple translation.

Proof of Theorem 3.1 Since $P_{H_2}A^nP_{H_2} = P_{H_2^n}A^nP_{H_2^n}$, the range of $P_{H_2}A^nP_{H_2}$ is finite dimensional, and therefore closed, so that it follows from Corach, Maestripieri and Stojanoff that A^n is compatible with H_2 for all n. Now we utilize the approximation results of Butler and Morley for the shorted operator. By their Lemma 1, for $c \in H$ and for fixed n, it follows that there exists a sequence $y_m^n \in H_2, m = 1,..$ and a real number M such that

$$\begin{aligned} A_{11}^n c + A_{12}^n y_m^n &\to H_1(A^n)c, && m \to \infty \\ A_{21}^n c + A_{22}^n y_m^n &\to 0, && m \to \infty \\ \langle A_{22}^n y_m^n, y_m^n \rangle &\leq M, && \forall m. \end{aligned}$$

Since $A_{11}^n = A_{11}, A_{12}^n = A_{12}P_{H_2^n}, A_{21}^n = P_{H_2^n}A_{21,}$ and $A_{22}^n = P_{H_2^n}A_{22}P_{H_2^n}$ this can be written as

$$\begin{aligned} A_{11}c + A_{12}P_{H_2^n}y_m^n &\to H_1(A^n)c, && m \to \infty \\ P_{H_2^n}A_{21}c + P_{H_2^n}A_{22}P_{H_2^n}y_m^n &\to 0, && m \to \infty \\ \langle A_{22}^n P_{H_2^n}y_m^n, P_{H_2^n}y_m^n \rangle &\leq M, && \forall m. \end{aligned}$$

Since these equations only depend on $P_{H_2^n}y_m^n$ we can further assume that $P_{(H_2^n)^\perp}y_m^n = 0, m = 1,\ldots$ where $P_{(H_2^n)^\perp}$ is the orthogonal projection onto $(H_2^n)^\perp \subset H_2$. That is, we can assume that $P_{H_2^n}y_m^n = y_m^n, m = 1,\ldots$ and therefore

$$\begin{aligned} A_{11}c + A_{12}y_m^n &\to H_1(A^n)c, && m \to \infty \\ P_{H_2^n}A_{21}c + P_{H_2^n}A_{22}y_m^n &\to 0, && m \to \infty \\ \langle A_{22}^n y_m^n, y_m^n \rangle &\leq M, && \forall m. \end{aligned} \tag{4.6}$$

It follows from $H_1(A^n) \leq A^n$ that $\left\|\sqrt{H1(A^n)}\right\| \leq \left\|\sqrt{A^n}\right\|$ for the unique square root. Consequently, it follows that $\|H_1(A^n)\| \leq \|A^n\|$ for all n and since $\|A^n\| \leq \|A\|$ for all n it follows that $\|H_1(A^n)\| \leq \|A\|$ for all n. Consequently, the sequence $H_1(A^n)c$ is bounded. Therefore there exists a weakly convergent subsequence. Let n' denote the index of any weakly convergent subsequence, so that

$$\mathcal{H}_1(A^{n'})c \xrightarrow{\omega} d', \quad n' \to \infty \tag{4.7}$$

for some d' depending on the subsequence. Now the strong convergence of the lefthand side to the righthand side in (4.6) is maintained for the subsequence n' and, since for the subsequence the first term on the righthand side converges weakly to d', it follows that we can define a monotonically increasing function $m(n')$ and use it to define a new sequence $\hat{y}^{n'} : y_{m(n')}^{n'}$ such that

$$\begin{aligned} A_{11}c + A_{12}\hat{y}^{n'} &\underset{\omega}{\to} d' && n' \to \infty \\ P_{H_2^{n'}}A_{21}c + P_{H_2^{n'}}A_{22}\hat{y}^{n'} &\to 0, && n' \to \infty \\ \langle A_{22}\hat{y}^{n'}, y^{n'} \rangle &\leq M, && \forall n'. \end{aligned} \tag{4.8}$$

Since $P_{H_2^n}$ is strongly convergent to P_{H_2} it follows that $P_{H_2^{n'}}$ is strongly convergent to P_{H_2}, so that $P_{H_2^{n'}}A_{21C}$ converges to $A_{21}c$ and $P_{H_2^{n'}}A_{22}\hat{y}^{n'}$ converges to $-A_{21}c$. Moreover, by Reid's inequality, Corollary 2, we have

$$\left\|A_{22}\hat{y}^{n'}\right\|_{H_2}^2 \leq \|A_{22}\| \left\langle A_{22}\hat{y}^{n'}, \hat{y}^{n'} \right\rangle \leq \|A_{22}\| M, \tag{4.9}$$

for all n', so that the sequence $A_{22}\hat{y}^{n'}$ is bounded. Since weak convergence of a bounded sequence on a separable Hilbert space is equivalent to the convergence with respect to each element of any orthonormal basis, it follows that $A_{22}\hat{y}^{n'}$ is weakly convergent to $-A_{21}c$. From (4.8), we obtain

$$\begin{aligned} A_{12}\hat{y}^{n'} &\xrightarrow{\omega} d' - A_{11}c, \qquad n' \to \infty \\ A_{22}\hat{y}^{n'} &\xrightarrow{\omega} -A_{21}c, \qquad n' \to \infty . \end{aligned} \tag{4.10}$$

From Kakutani's generalization of the Banach-Saks Theorem it follows that we can select a subsequence $\grave{n}$ of n' such that the Cesaro means of $A_{22}\hat{y}^{n'}$ and $A_{12}\hat{y}^{n'}$ converge strongly in (4.10). That is, if we consider the Cesaro means

$$z^{\grave{n}} = \frac{1}{\grave{n}}\sum_{i=1}^{\grave{n}} \hat{y}^{n'}$$

we have

$$\begin{aligned} A_{12}z^{n'} &\to d' - A_{11}c, \qquad n' \to \infty \\ A_{22}z^{n'} &\to -A_{21}c, \qquad n' \to \infty . \end{aligned}$$

Since $A_{22} \geq 0$ it follows that the function $y \to \langle A_{22}y, y\rangle$ is convex, so that $\langle A_{22}z^{\grave{n}}, z^{\grave{n}}\rangle \leq M$ for all $\grave{n}$, so that

$$\begin{aligned} A_{11}c + A_{12}z^{\grave{n}} &\to d', \qquad \grave{n} \to \infty \\ A_{21}c + A_{22}z^{\grave{n}} &\to 0, \qquad \grave{n} \to \infty \\ \langle A_{22}z^{\grave{n}}, z^{\grave{n}}\rangle &\leq M, \qquad \forall \grave{n}. \end{aligned}$$

It therefore follows from Theorem 1 of Butler and Morley that

$$d' = H_1(A)c$$

Consequently, by (4.7), we obtain that

$$H_1(A^{n'})c \xrightarrow{\omega} H_1(A)c, \qquad n' \to \infty. \tag{4.11}$$

Since this limit is independent of the chosen weakly converging subsequence, it follows that the full sequence weakly converges to the same limit, that is we have

$$H_1(A^{n})c \xrightarrow{\omega} H_1(A)c, \qquad n' \to \infty, \tag{4.12}$$

and since c was arbitrary we conclude that

$$H_1(A) = \omega - \lim_{n\to\infty} H_1(A^n).$$

Proof of Theorem 3.3 Let us first establish the assertion when C is compatible with H_2. Consider the operator $\hat{C}$: H →H defined by

$$\hat{C} := H_1(C) + P_2CP_2.$$

Since C is compatible with H_2, there exists an oblique projection $Q \in P(C,H_2)$, and Proposition 4.2 of Corach, Maestripieri and Stojanoff asserts that for $E := 1-Q$, we have

$$H_1(C) = CE = E^*CE. \tag{4.13}$$

Since $Q^* C = CQ$ it follows that $E^* C = CE$, and since Q is a projection, it follows that $QE = EQ = 0$ and that E is a projection. Moreover, since $R(Q) = H_2$ it follows that $ker(E) = H_2$, so that we obtain $P_2Q = Q$ and $EP_1 = E$ and therefore $Q^* P_2 = Q^*$ and $P_1E^* = E^*$. Consequently, we obtain

$$\begin{aligned} (P_1+Q)^*\hat{C}(P_1+Q) &= (P_1+Q)^*(E^*CE + P_2CP_2)(P_1+Q) \\ &= (P_1+Q)^*(E^*CE + P_2CQ) \\ &= E^*CE + P_2CQ \\ &= CE + CQ \\ &= C \end{aligned}$$

that is

$$C = (P_1+Q)^* \hat{C}(P_1+Q). \tag{4.14}$$

Since Q is a projection onto H_2, it follows that P_1+Q is lower triangular in its partitioned representation and therefore the fundamental pivot produces an explicit, and most importantly continuous, inverse. Indeed, if we use the partition representation

$$Q=\begin{pmatrix} 0 & 0 \\ \hat{Q} & 1 \end{pmatrix},$$

we see that

$$(P_1+Q)=\begin{pmatrix} 0 & 0 \\ \hat{Q} & 1 \end{pmatrix}$$

from which we conclude that

$$(P_1+Q)^{-1}=\begin{pmatrix} 0 & 0 \\ -\hat{Q} & 1 \end{pmatrix}$$

Without partitioning, using $P_1Q= 0$ *and* $QP_2 =P_2$, we obtain

$$\begin{aligned}(2-P_1+Q)(P_1+Q) &= 2P_1+2Q-(P_1^2+P_1Q+QP_1+Q^2) \\ &= 2P_1+2Q-P_1-P_1Q-QP_1-Q \\ &= P_1+Q-QP_1 \\ &= P_1+QP_2 \\ &= P_1+P_2 \\ &= 1\end{aligned}$$

and so confirm that

$$(P_1+Q)^{-1} = 2-P_1-Q \tag{4.15}$$

Following the proof of Lemma 4.3 of Hairer, Stuart, Voss, and Wiber, let $N(m,C)$ denote the Gaussian measure with mean m and covariance operator C and consider the transformation

$$(P_1+Q)^{-*}: H \to H,$$

where we use the notation A^{-*} *for* $(A^{-1})^* = (A^*)^{-1}$. From (4.14) we obtain

$$(P_1+Q)^{-*}C(P_1+Q)^{-1} = \hat{C} \tag{4.16}$$

so that the transformation law for Gaussian measures, see Lemma 1.2.7 of Maniglia and Rhandi, implies that

$$((P_1+Q)^{-*})_* N(m,c) = N\left((P_1+Q)^{-*}m, \hat{C}\right).$$

Since

$$(P_1+Q)^{-1}=\begin{pmatrix} 0 & 0 \\ -\hat{Q} & 1 \end{pmatrix}$$

we obtain

$$(P_1+Q)^{-*}=\begin{pmatrix} 0 & -\hat{Q}^* \\ 0 & 1 \end{pmatrix}$$

and therefore

$$(P_1+Q)^{-*}m=\begin{pmatrix} m_1-\hat{Q}^*m_2 \\ m_2 \end{pmatrix}.$$

Since the partition representation of $\hat{C}$

$$\hat{C}=\begin{pmatrix} (H_1(C))_{11} & 0 \\ 0 & C_{22} \end{pmatrix}$$

the components of the corresponding Gaussian random variable are uncorrelated and therefore independent. That is, we have

$$N\left((P_1+Q)^{-*}m, \hat{C}\right) = N\left(m_1-\hat{Q}^*m_2, (H_1(C))_{11}\right)N(m_2, C_{22}).$$

This independence facilitates the computation of the conditional measure as follows. Let $X = (X_1,X_2)$ denote the random variable associated with the Gaussian measure $N(m,C)$ and consider the transformed random variable $Y = (P_1+Q)^{-*} X$ with the product law $N\left(m_1 - \hat{Q}^* m_2, (H_1(C))_{11}\right) N(m_2, C_{22})$.

then,

$$\begin{aligned} Y_1 &= X_1 - \hat{Q}^* X_2 \\ Y_2 &= X_2 \end{aligned}$$

can be used to compute the conditional expectation as

$$\begin{aligned} E[X_1 | X_2] &= E[X_1 - \hat{Q}^* X_2 | X_2] + E[\hat{Q}^* X_2 | X_2] \\ &= E[Y_1 | Y_2] + E[\hat{Q}^* X_2 | X_2] \\ &= E[Y] + \hat{Q}^* X_2, \end{aligned}$$

obtaining

$$E[X_1 | X_2] = E[Y] + \hat{Q}^* X_2, \tag{4.17}$$

so that we conclude that

$$E[X_1 | X_2] = m_1 + \hat{Q}^*(X_2 - m_2).$$

A similar calculation obtains the covariance

$$C(X | X_2) = H_1(C) \tag{4.18}$$

thus establishing the assertion in the compatible case.

For the general case, we do not assume that C is compatible with H_2. Consider an ordered orthonormal basis for H_2, let H_2^n denote the span of the first n basis elements, let $P^n := P_{H_1} + P_{H_2^n}$ denote the orthogonal projection onto $H_1 \oplus H_2^n$ and consider the sequence of Gaussian measures $\mu_n := P_*^n \mu$ with the mean $P^n m$ and covariance operators

$$C^n := P^n C P^n, \quad n = 1....$$

As asserted in Theorem 3.1, C^n is compatible with H_2 for all n, and the sequence $H_1(C^n)$ converges weakly to $H_1(C)$. Let $C(X_1 | H_2^n)$ and $C(X_1|X_2)$ denote the conditional covariance operators associated with the measure μ. Then we will show that $C(X_1 | H_2^n) = H_1(C^n)$, so that the assertion regarding the conditional covariance operators is established if we demonstrate that the sequence of conditional covariance operators $C(X_1 | H_2^n)$ converges weakly to $C(X_1|X_2)$.

To both ends, consider the Lebesgue-Bochner space $L^2(H,\mu,\mathcal{B})$ space of (equivalence classes) of H-valued Borel measurable functions on H whose squared norm

$$\|f\|^2_{L^2(H,\mu,\mathcal{B})} := \int_H \|f(x)\|^2_H \, d\mu(x)$$

is integrable. Since Fernique's Theorem, implies that the random variable X is square Bochner integrable, it follows that the Gaussian random variables $P^n X$ are also square Bochner integrable with respect to μ. Let us denote $\mathcal{B}_2 := \{(H_1,T) : T \in \mathcal{B}(H_2)\}$ and $\mathcal{B}_2^n := \{(H_1, T^n, (H_2^n)^\perp) : T^n \in \mathcal{B}(H_2^n)\}$, and let $\mu_n := P_*^n \mu$ denote the image under the projection. μ_n is a Gaussian measure on H with mean $P^n m$ and covariance C^n.

Now consider a function $f: H \to H$ which is Bochner square integrable with respect to μ and satisfies $f \circ P^n = f$. Then, using the change of variables formula for Bochner integrals, see Theorem 2 of Bashirov, *et al.* along with the fact that $(P^n)^{-1}\mathcal{B}_2 = \mathcal{B}_2^n$, and using the fact that for an arbitrary $\mathcal{B}_2^n$. measurable function g we have $g = g \circ P^n$, it follows that for $A \in \mathcal{B}_2$, we have

$$\begin{aligned} \int_A f d\mu_n &= \int_{(P^n)^{-1}A} f \circ P^n d\mu \\ &= \int_{(P^n)^{-1}A} f d\mu \\ &= \int_{(P^n)^{-1}A} E_\mu[f | (P^n)^{-1}\mathcal{B}_2] d\mu \\ &= \int_{(P^n)^{-1}A} E_\mu[f | (\mathcal{B}_2^n)] d\mu \\ &= \int_{(P^n)^{-1}A} E_\mu[f | (\mathcal{B}_2^n)] \circ P^n d\mu \\ &= \int_A E_\mu[f | (\mathcal{B}_2^n)] d\mu_n \end{aligned}$$

we obtain

$$E_{\mu_n}[f|\mathcal{B}_2]=E_{\mu}[f|\mathcal{B}_2^n], \tag{4.19}$$

and conclude that the sequence $E_{\mu_n}[f|\mathcal{B}_2], n=1...$ is a martingale corresponding to the increasing family of σ-algebras B_2^n. Moreover, it is easy to see that (4.19) holds for real valued functions f: $H \to R$ which are square integrable with respect to μ and satisfy $f \circ P^n = f$. With the choice f:= X_1, we clearly have $X_1 \circ P^n = X_1$, so that if we denote $X_2^n := P^n X_2$, we conclude that the sequence

$$E_{\mu_n}[X_1|X_2]=E_{\mu}[X_1|X_2^n], \qquad n=1,.... \tag{4.20}$$

is a martingale. Since conditional expectation is a contraction, it follows that the L^2 norm of all the conditional expectations are uniformly bounded by the L^2 norm of X. Then by the Martingale Convergence Theorem, Corollary V.2.2 of Diestel and Uhl, $E_{\mu_n}[X_1|X_2]$ converges to $E_{\mu}[X_1|X_2]$ in L^2 $(H,\mu,\mathcal{B})$. For the conditional covariance operators, observe that (4.20) implies that

$$X-E_{\mu_n}[X|X_2]=X_1-E_{\mu}\left[X_1|X_2^n\right] \tag{4.21}$$

for all n, so that for $h_1,h_2 \in H$, we have

$$\begin{aligned}\left\langle C_{\mu_n}(X|X_2)h_1,h_2\right\rangle &:= E_{\mu_n}[\left\langle X-E_{\mu_n}[X|X_2],h_1\right\rangle\left\langle X-E_{\mu_n}[X_1|X_2],h_2\right\rangle|X_2] \\ &= E_{\mu_n}[\left\langle X_1-E_{\mu}[X_1|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle|X_2]\end{aligned}$$

and since the integrand $f:=\left\langle X_1-E_{\mu}[X_1|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle$ satisfies $f \circ P^n = f$, it follows from (4.19) that

$$\begin{aligned}&E_{\mu_n}[\left\langle X_1-E_{\mu}[X_1|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle|X_2] \\ &E_{\mu_n}[\left\langle X_1-E_{\mu}[X|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle|X_2^n]\end{aligned}$$

so that using the theorem of normal correlation, we obtain

$$\left\langle C_{\mu_n}(X|X_2)h_1,h_2\right\rangle := E_{\mu}[\left\langle X_1-E_{\mu}[X_1|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle] \tag{4.22}$$

Since the theorem of normal correlation also shows that

$$\begin{aligned}\left\langle C_{\mu_n}(X|X_2)h_1,h_2\right\rangle &:= E_{\mu}[\left\langle X-E_{\mu}[X|X_2],h_1\right\rangle\left\langle X-E_{\mu}[X|X_2],h_2\right\rangle|X_2] \\ &= E_{\mu}[\left\langle X-E_{\mu}[X|X_2],h_1\right\rangle\left\langle X-E_{\mu}[X|X_2],h_2\right\rangle] \\ &= E_{\mu}[\left\langle X_1-E_{\mu}[X_1|X_2],h_1\right\rangle\left\langle X-E_{\mu}[X_1|X_2],h_2\right\rangle],\end{aligned}$$

the difference in the covariances can be decomposed as

$$\begin{aligned}&\left\langle C_{\mu_n}(X|X_2)h_1,h_2\right\rangle-\left\langle C_{\mu_n}(X_1|X_2)h_1,h_2\right\rangle \\ &= E_{\mu}[\left\langle X-E_{\mu}[X_1|X_2^n],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2^n],h_2\right\rangle] \\ &-E_{\mu}[\left\langle X_1-E_{\mu}[X_1|X_2],h_1\right\rangle\left\langle X_1-E_{\mu}[X_1|X_2],h_2\right\rangle] \\ &= E_{\mu}[\langle E_{\mu}[X_1|X_2]-E_{\mu}[X_1||X_2^n],h_1\rangle\langle X_1,h_2\rangle]+E_{\mu}[\langle X_1,h_1\rangle]\left\langle E_{\mu}[X_1|X_2]-E_{\mu}[X_1||X_2^n],h_2\right\rangle] \\ &+E_{\mu}[\langle E_{\mu}[X_1|X_2^n],h_1\rangle\langle E_{\mu}[X_1||X_2^n],h_2\rangle]-E_{\mu}[\langle E_{\mu}[X_1|X_2],h_1\rangle\langle E_{\mu}[X_1|X_2],h_2\rangle]\end{aligned}$$

where the last term can be decomposed as

$$\begin{aligned}&E_{\mu}[\langle E_{\mu}[X_1|X_2^n],h_1\rangle\langle E_{\mu}[X_1||X_2^n],h_2\rangle]-E_{\mu}[\langle E_{\mu}[X_1|X_2],h_1\rangle\langle E_{\mu}[X_1|X_2],h_2\rangle] \\ &= E_{\mu}[\langle E_{\mu}[X_1|X_2^n]-E_{\mu}[X_1|X_2],h_1\rangle\langle E_{\mu}[X_1||X_2^n],h_2\rangle] \\ &+E_{\mu}[\langle E_{\mu}[X_1|X_2],h_1\rangle\langle E_{\mu}[X_1|X_2^n]-E_{\mu}[X_1|X_2],h_2\rangle].\end{aligned}$$

Then since conditional expectation is a contraction on $L_2(H,\mu,\mathcal{B})$ it follows that $\|E_{\mu}[X_1|X_2]\|_{L_2(H\mu,\mathcal{B})} \leq \|X_1\|_{L_2(H\mu,\mathcal{B})}$ and $\|E_{\mu}[X_1|X_2^n]\|_{L_2(H\mu,\mathcal{B})} \leq \|X_1\|_{L_2(H\mu,\mathcal{B})}$ for all n. Moreover, since $E_{\mu}[X_1|X_2^n]$ converges $E_{\mu}[X_1|X_2]$ in $L_2(H\mu,\mathcal{B})$ it follows that $\langle E_{\mu}[X_1|X_2^n],h\rangle$ converges to $\langle E_{\mu}[X_1|X_2],h\rangle$ in $L_2(H,\mu,\mathcal{B})$ for all $h \in H$. Therefore, the Cauchy-Schwartz inequality applied four times in the above decomposition implies that

$$\lim_{n\to\infty}\left\langle C_{\mu_n}(X|X_2)h_1,h_2\right\rangle=\left\langle C_{\mu}(X_1|X_2)h_1,h_2\right\rangle, \qquad h_1,h_2 \in H$$

so that we obtain

$$C_{\mu}(X|X_2)=\omega-\lim_{n\to\infty}C_{\mu_n}(X|X_2).$$

Since C^n is compatible with H_2 for all n, and the compatible case demonstrated in (4.18) that

$$C_{\mu_n}(X|X_2)=H_1(C^n) \tag{4.23}$$

for all n, and Theorem 3.1 asserts that

$$H_1(C) = \omega - \lim_{n\to\infty}(C^n),$$

we conclude that $C_\mu(X \mid X_2) = H_1(C)$, establishing the assertion regarding the covariance operators.

For the means, observe that since μ is a probability measure, it follows that X and therefore X_1 lie in the Lebesgue-Bochner space $L^1(H,\mu,\mathcal{B})$, and since by Diestel and Uhl the conditional expectation operators are also contractions on $L^1(H,\mu,\mathcal{B})$ it also follows that $E_{\mu_n}[X_1|X_2]$ converges to $E_\mu[X_1|X_2]$ in $L_1(H,\mu,\mathcal{B})$. Therefore, Diestel and Uhl [13, Thm. V.2.8] implies $E_{\mu_n}[X_1|X_2]$ that converges to $E_\mu[X_1|X_2]$ a.e.-μ. Let the conditional means $E_\mu[X_1|X_2]$ be denoted by $E_\mu[X_1|X_2] = m_t, t \in H_2$ Then, since

$$p^n m = \begin{pmatrix} m_1 \\ P_{H_2^n} m_2 \end{pmatrix},$$

is the mean of the measure μ_n, the assertion in the compatible case demonstrated that the conditional means $E_{\mu_n}[X_1|X_2] = m_t^n, t \in H_2$ are

$$m_t^n = \begin{pmatrix} m_1 + \hat{Q}_n^*(t - P_{H_2^n} m_2) \\ t \end{pmatrix},$$

Since the conditional means $E_{\mu_n}[X_1|X_2]$ converge to the conditional means $E_\mu[X_1|X_2]$ a.e.-μ amounts to $m_t^n \to m_t$ for μ-almost every t, the first assertion regarding the means is also proved. Now suppose that Q_n eventually becomes the special element $Q_n = Q_{C^n}, H_2$ defined near (3.2). Then, by definition, $R(\hat{Q}_n) \subset \overline{R(C_{22}^n)}$ so that $ker(\hat{Q}_n^*) \supset R(C_{22}^n)^\perp$, but since $C_{22}^n = \prod_2 C^n \prod_2^* = \prod_2 P^n C P^n \prod_2^* = \prod_2 P_{H_2^n} C P_{H_2^n} \prod_2^*$, it follows that $R(C_{22}^n) \subset H_2^n$ and therefore $R(C_{22}^n)^\perp \supset (H_2^n)^\perp$ so that $ker(\hat{Q}_n^*) \supset R(H_{22}^n)^\perp$. Therefore $\hat{Q}_n P_{H_2^n} = \hat{Q}_n$, so that the final assertion follows from the previous.

Proof of Corollary 3.4 By Mourier's Theorem, there exists a Gaussian measure μ on H with mean 0 and covariance operator $C := A$. Looking at the end of the proof of Theorem 3.3, since conditional expectation is a contraction on $L_2(H,\mu,\mathcal{B})$ it follows that $\|E_\mu[X_1|X_2]\|_{L_2(H,\mu,\mathcal{B})} \le \|X_1\|_{L_2(H,\mu,\mathcal{B})}$ and $\|E_\mu[X_1|X_2^n]\|_{L_2(H,\mu,\mathcal{B})} \le \|X_1\|_{L_2(H,\mu,\mathcal{B})}$ for all n. Therefore, for $h \in H$, it follows from the Cauchy-Schwartz inequality that $\|\langle E_\mu[X_1|X_2^n], h\rangle\|_{L_2(H,\mu,\mathcal{B})} \le \|X_1\|_{L_2(H,\mu,\mathcal{B})}$ and $\|\langle E_\mu[X_1|X_2^n], h\rangle\|_{L_2(H,\mu,\mathcal{B})} \le \|X_1\|_{L_2(H,\mu,\mathcal{B})}$ for all n, uniformly for $h \in H$ with $\|h\|_H \le 1$. Therefo the Cauchy-Schwartz inequality applied four times in the decomposition at the end of the proof of Theorem 3.3 implies that

$$\lim_{n\to\infty}\langle C_{\mu_n}(X \mid X_2)h_1, h_2\rangle = \langle C_\mu(X_1 \mid X_2)h_1, h_2\rangle, \qquad h_1, h_2 \in H$$

Uniformly for $h_1, h_2 \in H$ with $\|h_1\|_H \le 1$ and $\|h_1\|_H \le 1$. Therefore, it follows that the sequence of covariance operators converges

$$C_{\mu_n}(X \mid X_2) \to C_\mu(X \mid X_2)$$

in the uniform operator topology.

According to Maniglia and Rhandi or Da Prato and Zabczyk, for a Gaussian measure μ with mean 0 and covariance operator C, we have

$$tr(C) = E_\mu \|X\|^2.$$

From (4.22), by shifting to the center, we obtain that

$$tr\left(C_{\mu_n}(X \mid X_2)\right) = E_\mu\left[\|X_1 - E_\mu[X_1 \mid X_2^n]\|^2\right]$$

And

$$tr\left(C_{\mu_n}(X \mid X_2)\right) = E_\mu\left[\|X_1 - E_\mu[X_1 \mid X_2]\|^2\right],$$

and therefore the difference

$$\begin{aligned} & tr\left(C_{\mu_n}(X \mid X_2)\right) - tr\left(C_\mu(X \mid X_2)\right) \\ &= E_\mu\left[\|X_1 - E_\mu[X_1 \mid X_2^n]\|^2\right] - E_\mu\left[\|X_1 - E_\mu[X_1 \mid X_2]\|^2\right] \\ &= E_\mu\left[\langle E_\mu[X_1 \mid X_2^n] - E_\mu[X_1 \mid X_2], E_\mu[X_1 \mid X_2^n] + E_\mu[X_1 \mid X_2] - 2X_1\rangle\right]. \end{aligned}$$

Therefore, the Cauchy-Schwartz inequality, the L^2 convergence of $E_\mu[X_1|X_2^n]$ to $E_\mu[X_1|X_2]$, and the uniform L^2 boundedness of $E_\mu[X_1|X_2^n]$, $E_\mu[X_1|X_2]$ and X_1, implies that

$$\lim_{n\to\infty} tr(C_{\mu_n}(X \mid X_2)) = tr(C_\mu(X \mid X_2)).$$

Since $C_{\mu_n}(X_1|X_2) \to C_\mu(X_1|X_2)$ in the uniform operator topology, it follows from Kubrusly that $C_{\mu_n}(X_1|X_2) \to C_\mu(X_1|X_2)$ in the trace norm topology. Since (4.23) asserts that $C_\mu(X|X_2) \to H_1(C)$, and Theorem 3.3 asserts that $C_{\mu_n}(X_1|X_2) \to H_1(C^n)$ the identification *A*:=*C* completes the proof [40-46].

Acknowledgment

The authors gratefully acknowledge this work supported by the Air Force Office of Scientific Research under Award Number FA9550-12-1-0389 (Scientific Computation of Optimal Statistical Estimators).

References

1. Krein M (1947) The theory of self-adjoint extensions of semi-bounded hermitian transformations and its applications. I Matematicheskii Sbornik 62: 431-95.

2. Anderson WN, Trapp GE (1975) Shorted operators. II. SIAM J Appl Math 28: 60-71.

3. Hairer M, Stuart AM, Voss J, Wiberg P (2005) Analysis of SPDEs arising in path sampling. Part I: The Gaussian case. Commun Math Sci 3: 587-603.

4. Stuart AM (2010) Inverse problems: a Bayesian perspective. Acta Numer 19: 451-559.

5. Dalecky YL, Fomin SV (1991) Measures and Differential Equations in Infinite-dimensional Space. Springer Science & Business Media 76.

6. Mandelbaum A (1984) linear estimators and measurable linear transformations on a Hilbert space. J Prob Theor Relate Area 65: 385-97.

7. LaGatta T (2013) Continuous disintegrations of Gaussian processes. Theory Probab Appl 57: 151-62.

8. Tarieladze V, Vakhania N (2007) Disintegration of Gaussian measures and average-case optimal algorithms. J Complex 23: 851-66.

9. Lee D, Wasilkowski GW (1986) Approximation of linear functionals on a Banach space with a Gaussian measure. J Complex 2: 12-43.

10. Traub JF, Wasilkowski GW, Wozniakowski H (1998) Information-Based Complexity. Academic Press, New York.

11. Diaconis P (1988) Bayesian numerical analysis. In Statistical Decision Theory and Related Topics 1: 163-75.

12. Poincare H (1896) Calculation of probabilities. Georges Carr'es, Paris.

13. Sul'din AV (1959) Wiener measure and its applications to approximation methods. I Izv Vyss Ucebn Introduced. Mathematics 1959: 145-58.

14. Larkin FM (1972) Gaussian measure in Hilbert space and applications in numerical analysis. Rocky Mountain J Math 2: 379-422.

15. Sard A (1963) Linear approximation. American Mathematical Society, Providence.

16. Kimeldorf GS, Wahba G (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann Math Statist 41: 495–502.

17. Shaw JEH (1988) A quasirandom approach to integration in Bayesian statistics. Ann. Statist., 16: 895-914.

18. O'Hagan A (1992) Some Bayesian numerical analysis. In Bayesian Statistics, Oxford Univ Press, New York 345-63.

19. Briol FX, Cockayne J, Teymur O, Yoo WW, Schober M, et al. (2016) Contributed discussion on article by Chkrebtii, Campbell, Calderhead, and Girolami. Bayesian Anal 11: 1285-93.

20. Owhadi H, Scovel C (2017) Universal scalable robust solvers from computational information games and fast eigenspace adapted multiresolution analysis. arXiv: 1703.10761.

21. Skilling J (1991) Maximum Entropy and Bayesian Methods: Seattle, chapter Bayesian Solution of Ordinary Differential Equations. Springer Netherlands, Dordrecht.

22. Schober M, Duvenaud DK, Hennig P (2014) Probabilistic ODE solvers with Runge-Kutta means. Advances in Neural Information Processing Systems 27: 739-47.

23. Owhadi H (2015) Bayesian numerical homogenization. Multiscale Model Simul 13: 812-28.

24. Hennig P, Osborne MA Girolami M (2015) Probabilistic numerics and uncertainty in computations. Proc R Soc 471: 20150142.

25. Hennig P (2015) Probabilistic interpretation of linear solvers. SIAM J Optim 25: 234-60.

26. Briol FX, Oates CJ, Girolami M, Osborne MA, Sejdinovic D (2015) Probabilistic integration: A role for statisticians in numerical analysis? arXiv: 1512.00933.

27. Raissi M, Perdikaris P, Karniadakis GE (2017) Inferring solutions of differential equations using noisy multi-fidelity data. J Comp Phy 335: 736-46.

28. Owhadi H (2017) Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. SIAM Rev 59: 99-149.

29. Cockayne J, Oates CJ, Sullivan T, Girolami MA (2016) Probabilistic meshless methods for partial differential equations and Bayesian inverse problems. arXiv: 1605.07811.

30. Perdikaris P, Venturi D, Karniadakis GE (2016) Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. SIAM J Sci Comput 38: 521-38.

31. Sch¨afer F, Sullivan TJ Owhadi H (2017) Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity. arxiv.org/abs/1706.02205.

32. Owhadi H, Scovel C (2019) Operator Adapted Wavelets, Fast Solvers, and Numerical Homogenization: from a game theoretic approach to numerical approximation and algorithm design.

33. Da Prato G, Zabczyk J (2014) Stochastic Equations in Infinite Dimensions. Cambridge University press 152.

34. Vakhania N, Tarieladze V, Chobanyan S (1987) Probability distributions on Banach spaces, Springer Science & Business Media 14.

35. Dudley RM (1989) Real Analysis and Probability, Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge 74.

36. Bogachev VI (1998) Gaussian Measures. American Mathematical Soc 62: 1-90.

37. Diestel J, Uhl JJ (1977) Vector measures. Number 15 in Mathematical Surveys and Monographs. Am Math Soc 15: 1.

38. Corach G, Maestripieri A, Stojanoff D (2001) Oblique projections and Schur complements. Acta Sci Math 67: 337–56.

39. Janson S (1997) Gaussian Hilbert Spaces. Cambridge University Press 129.

40. Corach G, Maestripieri A, Stojanoff D (2006) Projections in operator ranges. Proceedings of the American Mathematical Society 134: 765–78.

41. Butler CA, Morley TD (1988) A note on the shorted operator. SIAM J Matrix Anal Appl 9: 147-55.

42. Reid WT (1951) Symmetrizable completely continuous linear transformations in Hilbert space. Duke Math J 18: 41-56.

43. Maniglia S, Rhandi A (2004) Gaussian measures on Hilbert spaces. Quaderni of the Department of Mathematics of the University of Salento 2004: 1-24.

44. A Bashirov (2003) Partially Observable Linear Systems under Dependent Noises. Birkh¨auser Verlag.

45. Kubrusly CS (1986) On convergence of nuclear and correlation operators in Hilbert space. Mat Apl Comput 5: 265-82.

46. Kakutani S (1938) Weak convergence in uniformly convex spaces. Tˆohoku Math J 45: 188-93.