ON COVARIANCE OPERATORS

by

Charles R. Baker
Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 712

— September, 1970
On Covariance Operators

Charles R. Baker
University of North Carolina at Chapel Hill

INTRODUCTION

Let $H$ be a real and separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and Borel $\sigma$-field $\Gamma$. A "covariance operator" is any operator mapping $H$ into $H$ that is linear, bounded, non-negative, self-adjoint, and trace-class. If $\mu$ is a probability measure on $(H, \Gamma)$ and $\int_H |\xi|^2 d\mu(\xi) < \infty$, then $\mu$ has a mean element $\mu$ and covariance operator $R$ defined by

$$\langle \xi, y \rangle = \int_H \langle \xi, \xi \rangle \, d\mu(\xi)$$
$$\langle R \xi, y \rangle = \int_H \langle \xi - \xi, y \rangle \langle \xi - \xi, \xi \rangle \, d\mu(\xi),$$

for all $\xi, y$ in $H$ [1].

A probability measure $\mu$ on $(H, \Gamma)$ is said to be Gaussian if the probability distribution on $B[R^1]$ (Borel sets of the real line) induced from $\mu$ by a bounded linear functional is Gaussian, for all bounded linear functionals on $H$. Thus, if $P_{\mu}$ is defined by

$$P_{\mu}(r: r \in A) = \mu(x: \langle \xi, y \rangle \in A), \text{ for } A \in B[R^1],$$

then $\mu$ is Gaussian if and only if $P_{\mu}$ is Gaussian for all $y$ in $H$. Mourier [1] has shown that each Gaussian measure $\mu$ satisfies

$$\int_H |\xi|^2 d\mu(\xi) < \infty,$$

and that for each covariance operator $R$ in $H$ and element $\mu$ of $H$ there exists a Gaussian measure having $R$ as covariance operator and $\mu$ as mean element.
In applications, one usually has an underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and a \(\mathcal{F}/\mathbb{P}\) measurable mapping of \(\Omega\) into \(\mathcal{F}\), say \(X\). \(X\) then induces from \(\mathbb{P}\) a measure \(\nu_X\) on \((\mathcal{F}, \mathbb{I})\) in the usual way. An example of a \(\mathcal{F}/\mathbb{I}\) measurable map is a stochastic process \((S_t)_{t \in T}\), where \(T\) is a compact interval, the sample functions of \((S_t)\) belong almost surely to \(L^2(T)\) (Lebesgue measure), and \((S_t)\) is \(\mathcal{F}[T] \times \mathcal{B}[\mathbb{R}^1]\) measurable. The fact that \(S\) is \(\mathcal{F}/\mathbb{I}\) measurable follows easily from the fact that \(\mathbb{I}\) is the smallest \(\sigma\)-field such that all bounded linear functionals on \(\mathcal{F}\) are \(\mathcal{F}/\mathcal{B}[\mathbb{R}^1]\) measurable, and that \(\langle S, y \rangle\) is \(\mathcal{F}/\mathcal{B}[\mathbb{R}^1]\) measurable for all \(y\) in \(\mathcal{F}\) \([2]\).

Covariance operators thus arise naturally in problems involving Gaussian measures; typical are those involving mean-square continuous and measurable Gaussian processes on a compact interval. Examples of problems whose solutions can be given in terms of covariance and cross-covariance operators are the following: Equivalence of probability measures (see, e.g., \([3]\)); mean-square estimation \([4]\); and the computation of mutual information \([5], [6]\). Equivalence of probability measures has applications in the detection of signals in noise \([7]\).

In the problems cited above, a major aspect of the solution usually involves determination of the range of the square root of one or two covariance operators. For example, to show that two Gaussian probability measures on \((\mathcal{F}, \mathbb{I})\) are equivalent, one must show that the ranges of the square roots of the two covariance operators are identical. This can be a difficult problem. When \(\mathcal{F}\) is \(L^2(T)\), the integral operator representation of a covariance operator will often have kernel that is the Green's function of a linear differential operator in \(L^2(T)\), and the range of the covariance operator can then be easily determined (see, e.g., \([8]\), pp. 978-979). However, the range of the square root operator can be vastly more difficult to determine. Thus, methods for specifying the range of self-adjoint, non-negative bounded linear operators
are of some interest, and in this paper we obtain a number of such results. They are first stated for general bounded linear operators, and in most cases the simplification introduced when the operators are self-adjoint and non-negative are obvious. Some of these results are then further developed for the case when the operators are square roots of covariance operators. Finally, these results are applied to determine the range space of the square roots of several ubiquitous covariance operators in $L_2[T]$. 
Suppose that $R_1$ and $R_2$ are bounded linear operators mapping $H$ into $H$, that $P_1$ is the projection operator mapping $H$ onto $\text{range}(R_1^*)$ (the closure of the range of the adjoint of $R_1$), and that $N(R_1)$ is the null space of $R_1$. Several useful results on covariance and cross-covariance operators are a direct consequence of the following result.

**Proposition 1.** Range $(R_1) \subset \text{range} (R_2)$ if and only if there exists a bounded linear operator $G: H \rightarrow H$ such that $R_1 = R_2 G$, $G = G P_1 = P_2 G P_1$.

**Proof:** Sufficiency is clear. To prove necessity, suppose range $(R_1) \subset \text{range} (R_2)$; then for all $x$ in $H$ there exists $y$ in $H$ such that $R_1 x = R_2 y$. Define a mapping $G$ on $H$ by $G x = P_1 x = P_2 y$.

To see that $G$ is single-valued, suppose that $G x = z$ and $G x = z'$. Then $z = P_2 z + z'$, $z = P_2 z + z'$, where $z'$ and $z'$ are orthogonal to $\text{range}(R_2^*)$. Hence $R_1 x = R_2 P_2 z = R_2 z$, by the definition of $G$. $R_2 P_2 (z - y) = 0$ $\Rightarrow z - y \in N(R_2)$. If both $z$ and $y$ are in $N(R_2)$, then $P_2 z = P_2 y = 0$, and by the definition of $G$, $z = y = 0$. If either $z$ or $y$ is not in $N(R_2)$, then both must belong to $\text{range}(R_2^*)$, and then $z = y$ follows from $N(R_2) = \text{range}(R_2^*)$.

$G$ is clearly linear and defined everywhere in $H$.

Suppose next that $x_n \rightarrow x$, $G x_n \rightarrow y$. By the definition of $G$, $R_1 x_n = R_1 x \Rightarrow R_2 G x_n = R_2 y$. Since $G x_n \rightarrow y$, one has $R_2 G x_n \rightarrow R_2 y$; thus $R_1 x = R_2 y$. Hence $G x = y$, and $G$ is closed.

$G$ is thus a closed linear operator, defined everywhere in $H$; by the closed-graph theorem, $G$ is bounded.

**Corollary 1:** Suppose $R_1$ and $R_2$ are bounded linear operators in $H$. Then
(a) range \((R_1) \subseteq \text{range} \ (R_2) \iff \text{there exists a scalar } k < \infty \text{ such that } \langle R_1R_1^*y, y \rangle \leq k \langle R_2R_2^*y, y \rangle, \text{ all } y \in H;\)

(b) range \((R_1) \subseteq \text{range} \ (R_2) \iff R_1R_1^* = R_2Q R_2^*, \ Q \ a \ bounded \ linear \ operator;\)

(c) range \((R_2) \subseteq \text{range} \ (\{R_1R_1^* + R_2R_2^*\}^{\frac{1}{2}});\)

(d) range \((\{R_1R_1^* + R_2R_2^*\}^{\frac{1}{2}}) \subseteq \text{range} \ (R_2) \iff \text{range} \ (R_1) \subseteq \text{range} \ (R_2);\)

(e) For \(y \in H, \ y \in \text{range} \ (R_2) \iff \text{there exists a scalar } k < \infty \text{ such that } \langle y, x \rangle^2 \leq k \langle R_2R_2^*x, x \rangle, \text{ all } x \in H;\)

(f) range \((R_1) = \text{range} \ (R_2) \iff \text{there exists a bounded linear operator } G \text{ with bounded inverse such that } (R_1R_1^*)^{\frac{1}{2}} = (R_2R_2^*)^{\frac{1}{2}} G;\)

(g) range \((R_1) = \text{range} \ (R_2) \iff \text{there exists a bounded linear operator } Q \text{ having bounded inverse and such that } R_1R_1^* = (R_2R_2^*)^{\frac{1}{2}} Q (R_2R_2^*)^{\frac{1}{2}}.\)

**Proof:**

(a) \(|R_1^*y| \leq k|R_2^*y|, \text{ all } y \in H \implies \text{for any } y \in H, \)

\(|\langle y, R_1^*y \rangle| \leq k |y| |R_2^*y|, \text{ all } y. \)

Define a linear functional \(f_\|_y \) on \(\text{range} \ (R_2^*) \) by \(f_\|_y(R_2^*y) = \langle y, R_1^*y \rangle; \) from above, \(f \) is bounded on \(\text{range} \ (R_2^*) \) and can thus be extended by continuity to a bounded linear functional on \(\text{range} \ (R_2^*). \)

Hence, by Riesz' theorem, there exists an element \(z \) in \(\text{range} \ (R_2^*) \) such that \(\langle y, R_1^*z \rangle = \langle z, R_2^*y \rangle, \text{ all } y \in H, \text{ or } R_1^*y = R_2^*z. \)

Hence range \((R_1) \subseteq \text{range} \ (R_2). \) The converse follows from Proposition 1.

(b) - (e) Clear; for (e), use the operator \(R_1 \) defined by \(R_1x = \langle z, y \rangle y. \)

(f) From (a), range \((R_1) = \text{range} \ (\{R_1R_1^*\}^{\frac{1}{2}}). \) Suppose range \((R_1) = \text{range} \ (R_2) \). Then, there will exist bounded linear operators \(G_1, G_2 \) such that \((R_1R_1^*)^{\frac{1}{2}} = (R_2R_2^*)^{\frac{1}{2}} G_1 \) and \((R_2R_2^*)^{\frac{1}{2}} = (R_1R_1^*)^{\frac{1}{2}} G_2; \)
PG_1^*P = G_1, PG_2^*P = G_2, P the projection operator mapping H onto
range \( R_1 \). Let \( G_1 = G_1 + P \), with \( P \) the projection operator with range
\( N(R_1^*) \). Then \( (R_1 R_1^*)^\frac{1}{2} = (R_2 R_2^*)^\frac{1}{2} G_1^* \), since \( N(R_1^*) = N([R_1 R_1^*])^\frac{1}{2} \). If \( G_1 \) is a
subset of \( \mathbb{R} \), then both \( PG_1^*P \) and \( P^\perp \) are \( 0 \), since it is impossible otherwise for
\( PG_1^*P \neq P^\perp \). Now, if \( P^\perp \neq Q \), then \( P^\perp = Q \) and \( PG_1^*P \neq 0 \) if \( X \neq Q \),
since \( (R_1 R_1^*)^\frac{1}{2} X \neq 0 \). Hence \( G_1 \) exists. To see that \( G_1 \) is bounded, one
notes that \( G_2^* G_1^* = G_1^* G_2^* = I \), where \( G_2^* \equiv G_2 + P \).

(g) Clear.

\[ \textbf{Corollary 2:} \] Suppose \( R_i, i = 1, \ldots, N \) is linear and bounded. Let
\( R_0 = \sum_{i=1}^{N} R_i R_i^* \). Then

(a) \( (R_i R_i^*)^\frac{1}{2} = R_i A_i^* \), where \( A_i \) is partially isometric, \( A_i \) isometric on
\( \text{range} \ (R_i^*) \) and \( A_i y = 0 \) for \( y \in N(R_i) \);

(b) \( (R_i R_i^*)^\frac{1}{2} = R_i^\frac{1}{2} G_i \), \( G_i \) bounded, with \( \sum_{i=1}^{N} G_i G_i^* \) isometric on
\( \text{range} \ (R_0^\frac{1}{2}) \) and identically zero on \( N(R_0^\frac{1}{2}) \);

(c) \( \text{Range} \ (R_0^\frac{1}{2}) = \sum_{i=1}^{N} \text{range} \ [ (R_i R_i^*)^\frac{1}{2} ] = \sum_{i=1}^{N} \text{range} \ (R_i) \),
where \( \sum_i A_i \) is the linear manifold generated by \( U_i A_i \).

\[ \textbf{Proof:} \] (a) From (a) of Corollary 1, \( (R_i R_i^*)^\frac{1}{2} = R_i A_i^* \) for \( A_i \) bounded, and
\( A_i \) must be partially isometric, isometric on \( \text{range} \ (R_i^*) \), because
\( ||(R_i R_i^*)^\frac{1}{2} X|| = ||A_i R_i^* X|| = ||R_i X|| \), and one can define \( A_i y = 0 \) for
\( y \in N(R_i) \), by Proposition 1.

(b) \( (R_1 R_1^*) \leq R_0 \) for \( i = 1, \ldots, N \), so that \( (R_1 R_1^*)^\frac{1}{2} = R_0^\frac{1}{2} G_i \), \( G_i \)
bounded, \( G_i = G_i P_0 \), \( P_0 \) the projection operator with range equal to
\( \text{range} \ (R_0) \). Thus

\[ R_0 = R_0^\frac{1}{2} \left( \sum_{i=1}^{N} G_i G_i^* \right) R_0^\frac{1}{2} \]
so that \( \sum_{i=1}^{N} G_i G_i^* \) is isometric on \( \text{range}(R_0^{\frac{1}{2}}) \), and zero on \( N(R_0^{\frac{1}{2}}) \).

(c) It is obvious that \( \text{range}(R_0^{\frac{1}{2}}) \supseteq \sum_{i=1}^{N} \text{range}(R_i R_i^*)^{\frac{1}{2}} \), since 

\[
\text{range}(R_0^{\frac{1}{2}}) \supseteq \text{range}(R_i R_i^*)^{\frac{1}{2}}, \quad i = 1, \ldots, N,
\]

and \( \text{range}(R_0^{\frac{1}{2}}) \) is a linear manifold. From (b), 
\[
(R_i R_i^*)^{\frac{1}{2}} = R_0^{\frac{1}{2}} G_i, \quad \sum_{i=1}^{N} G_i G_i^* \text{ isometric on range } (R_0^{\frac{1}{2}}), \quad \text{zero on } N(R_0^{\frac{1}{2}}).
\]
Suppose that \( R_0^{\frac{1}{2}} \gamma = \xi \); then

\[
\sum_{i=1}^{N} (R_i R_i^*)^{\frac{1}{2}} G_i \gamma = R_0^{\frac{1}{2}} \sum_{i=1}^{N} G_i \gamma = \xi,
\]

so that \( \text{range}(R_0^{\frac{1}{2}}) \subseteq \sum_{i=1}^{N} \text{range}(R_i R_i^*)^{\frac{1}{2}} \). By (a) of Proposition 1, 
\[
\text{range}(R_i R_i^*)^{\frac{1}{2}} = \text{range}(R_i).
\]

**Corollary 3:** Suppose that \( R_1 \) and \( R_2 \) are linear, bounded, self-adjoint, and non-negative. Then

(a) \( \text{range } ([R_1 + R_2]^{\frac{1}{2}}) \supseteq \text{range } (R_2^{\frac{1}{2}}); \quad \text{range } (R_2^{\frac{1}{2}}) \supseteq \text{range } ([R_1 + R_2]^{\frac{1}{2}}) \) if and only if \( \text{range } (R_1^{\frac{1}{2}}) \subseteq \text{range } (R_2^{\frac{1}{2}}) \).

(b) Suppose \( H \) is \( L^2[T], \) \( T \) a compact interval, and \( R_1 \) and \( R_2 \) are integral operators with kernels defined by 
\[
R_1(t,s) = \int_{-\infty}^{\infty} e^{i\lambda(t-s)} \hat{a}_1(\lambda) d\lambda,
\]

with 
\[
\int_{-\infty}^{\infty} |R_1(t)| dt < \infty;
\]

then \( \text{range } (R_1^{\frac{1}{2}}) \subseteq \text{range } (R_2^{\frac{1}{2}}) \) if there exists a scalar \( k < \infty \) such that \( \hat{a}_1(\lambda) \leq k \hat{a}_2(\lambda) \) a.e. \( d\lambda \) on \( (-\infty, \infty) \).

**Proof:** Clear; for (b), use Parseval's theorem and convolution.
APPLICATIONS TO SPECIFIC SQUARE ROOTS

Let $H$ be $L_2[0,T]$ for $T$ finite; we proceed to specify the range of the square root for the covariance operators with the following kernels:

1. $R_1(t,s) = \min(t,s)$;
2. $R_2(t-s) = T - \max(t,s)$;
3. $R_3(t,s) = T - |t-s|$;
4. $R_4(t,s) = e^{-\alpha|t-s|}$, $\alpha > 0$;
5. $R_5(t,s) = \int_{-\infty}^{\infty} e^{-i\lambda (t-s)} \hat{H}_5(\lambda) \, d\lambda$,

where $\hat{H}_5$ is a rational spectral density function with denominator of degree exactly two more than the degree of the numerator.

The integral operators with these kernels are encountered quite often in communication theory problems. For example, $R_5$ could have kernel of the form $e^{-\alpha|t-s|} \cos \omega_0 (t-s)$ for $\alpha > 0$, $\omega_0 > 0$. The ranges of the covariance operators themselves are easy to find, since each kernel is the Green's function of a linear differential operator in $L_2[T]$ with constant coefficients; the ranges of $R_1$, $R_3$ and $R_4$ are given, for example, in [8]. However, for the square roots, only the range of $(R_{1/2})$ appears to be known (e.g. [3], [9]); it is included here for completeness.

(a) Range of $R_{1/2}$.

$$[R_{1/2}x](t) = \int_0^T \min(t,s)x_s \, ds = \int_0^t \int_u^T x_v \, dv \, du = [L^*x](t),$$

where

$$[Ly](t) = \int_0^t y_s \, ds, \quad [L^*y](t) = \int_t^T y_s \, ds.$$

Hence $R_1 = LL^*$, so that by Corollary 2 of Proposition 1, $R_{1/2} = LA$, $A$ isometric on $\text{range}(L^*)$. Since $\text{range}(L^*) = H$, $A$ is unitary. Hence $\text{range}(R_{1/2})$ consists of all elements of $L_2[0,T]$.
that are equal almost everywhere to an absolutely continuous function that vanishes at the origin and which has derivative belonging to $L^2[0,T]$.

(b) Range of $R_2^\frac{1}{2}$. Using integration by parts, $R_2 = L^*L$; using the same procedure as in (a), $R_2^\frac{1}{2} = L^*B$, $B$ unitary. Hence range $(R_2^\frac{1}{2})$ consists of all elements of $L^2[0,T]$ that are equal almost everywhere to an absolutely continuous function that vanishes at $T$ and has derivative belonging to $L^2[0,T]$.

(c) Range of $R_3^\frac{1}{2}$. $R_3 = R_1 + R_2$; hence by part (c) of Corollary 2, range $(R_3^\frac{1}{2})$ consists of all elements of $L^2[0,T]$ that are equal almost everywhere to an absolutely continuous function having derivative belonging to $L^2[0,T]$.

(d) Range of $R_4^\frac{1}{2}$. Rosanov has shown [10] that the zero-mean Gaussian measures having $R_3$ and $R_4$ as covariance operators are equivalent. This implies that $R_3 = R_4TR_4$, $T$ invertible ([3], Theorem 5.1) so that by (g) of Corollary 1, range $(R_3^\frac{1}{2}) = $ range $(R_4^\frac{1}{2})$.

(e) From (b) of Corollary 3, range $(R_5^\frac{1}{2}) = $ range $(R_4^\frac{1}{2})$. 
Applications to Discrimination of Gaussian Measures

Suppose that \( \mu_1 \) is a Gaussian measure with covariance operator \( R_1 \) and null mean element, and that \( \mu_1' \) is a Gaussian measure with covariance operator \( R_1 \) and mean element \( m_1 \). Then \( \mu_1 \) and \( \mu_1' \) are equivalent if and only if \( m_1 \in \text{range} \, (R_1^{1/2}) \), and are orthogonal otherwise [3]. The results of the preceding section thus give necessary and sufficient conditions on \( m_1 \) in order that \( \mu_1 \) and \( \mu_1' \) be equivalent, for the specific covariance operators treated there.

Further, if \( \mu_1 \) and \( \mu_j \) are equivalent, then \( \text{range} \, (R_1^{1/2}) = \text{range} \, (R_j^{1/2}) \) [3]. The preceding results show that \( \mu_1 \) (Wiener measure) is orthogonal to \( \mu_1 \) for \( i = 2, \ldots, 5 \); and that \( \mu_2 \) is orthogonal to \( \mu_i \) for \( i = 1, 3, 4, 5 \).

Finally, suppose that \( R_0 = \sum_{i=1}^{N} R_i \), where the \( R_i \) are unspecified covariance operators; let \( \mu_i, i = 1, \ldots, N \) be the Gaussian measure with null mean element and covariance operator \( R_i \), \( \mu_0 \) the Gaussian measure with null mean and covariance operator \( R_0 \). Then one sees from Corollary 3 that \( \mu_0 \bot \mu_i \) when \( \text{range} \, (R_1^{1/2}) \neq \text{range} \, (R_j^{1/2}) \) for \( j \in \{1, \ldots, N\} \). Moreover, if \( \text{range} \, (R_1^{1/2}) = \text{range} \, (R_0^{1/2}) \), then for \( \mu_0 \) to be equivalent to \( \mu_1 \) it is necessary that \( \text{range} \, (R_j^{1/2}) \neq \text{range} \, (R_i^{1/2}) \) for \( j \neq i \). Hence, if \( \mu_0 \) and \( \mu_i \) are equivalent, then \( \mu_0 \) and \( \mu_j \) are orthogonal for \( j \neq i \).
REFERENCES


(2) K. R. Parthasarathy, Probability Measures on Metric Spaces, Academic, New York, 1967; Chapter I, Section I.


