On the Stability of Signal Detection

**Abstract**

The probabilities of false alarm and false dismissal for sure signals embedded in additive noise is studied. The noise is contaminated Gaussian, and the framework is such that continuous-time problems can be studied.
ON THE STABILITY OF SIGNAL DETECTION

by

A. F. Gualtierotti*

Abstract

One studies probabilities of false alarm and false dismissal for sure signals imbedded in additive noise as functions of contamination when the noise is contaminated Gaussian.

*Department of Statistics, University of North Carolina, Chapel Hill, NC 27514.

IDHEAP, Sciences humaines 1, U. Lausanne, CH-1015 Lausanne.

Research partially supported by ONR Contracts: N00014-75-C-0491 and N00014-81-K-0373.
1. INTRODUCTION

One of the basic problems in statistical communication theory is that of detecting, at the end of a communication channel, a signal imbedded in noise. The Gaussian model is a model for which the noise as well as the received signal are Gaussian processes with equivalent laws, so that the resulting detection problem is nonsingular. Unfortunately "most solutions to problems involving signals in Gaussian noise are based on a knowledge of the autocorrelation function or spectrum (if the noise is stationary). In fact, of course, as mentioned before, they are rarely, if ever, known precisely" (19). This observation motivates the paper just quoted and, in it, its author investigates the stability of the decision functions based on a model with Gaussian noise and covariance function $R$ in two ways. First he considers the case of Gaussian noise with covariance function $R+A$ instead of just $R$. However "the restriction to Gaussian measures were better removed even though the noise is nearly Gaussian, if the stability criteria are to be believable in practice" (19). Stability is then understood as a form of continuity of the decision functions when the topology is that of weak convergence, a concept which is now widely used in the field of robust statistical procedures (11). Of course "this formulation does not allow much hope for actually calculating useful bounds on errors" (19). Since the most important elements of a detection problem are the probabilities of false alarm and false dismissal, one may try to study the stability problem by taking an intermediate position between the two described above. This is the object of the present work and the idea is to look at specific contaminations of Gaussian laws in order to find analytic
expressions for the probability of false alarm and that of false dismissal as functions of contamination. The aim is not so much to produce models for specific Gaussian disturbances as to get some feel for what these can do to the probabilities of false alarm and false dismissal.

It shall be assumed that signals and noises have finite energy, that is are stochastic processes with square integrable paths. These will induce probability measures on the Borel sets of $L^2(0,1)$, the space of equivalence classes of square integrable functions. $L^2(0,1)$ is a standard separable Hilbert space and it is notationally simpler to work with a general separable Hilbert space $H$, which we shall do. $B(H)$ is the sigma algebra generated by the open sets of $H$ and is the smallest sigma algebra which makes all continuous linear functionals $\langle \cdot, h \rangle$, $h$ in $H$, measurable. $\langle \cdot, \cdot \rangle$ is the inner product of $H$ and the corresponding norm is $|| \cdot ||$. All measures considered shall be defined on $B(H)$.

They are typically obtained as follows. If $X(\omega, t)$ is a stochastic process defined on the probability space $(\Omega, \mathcal{A}, \mu)$ such that $\int X^2(\omega, t) \, dt$ is finite for $\mu$-almost every $\omega$, it induces, on $B(H)$, a measure $P_X$ defined by

$$P_X(x \in H: \langle x, h_1 \rangle \in B_1, \ldots, \langle x, h_n \rangle \in B_n) =$$

$$= P(\omega \in \Omega: \langle X(\omega, \cdot), h_1 \rangle \in B_1, \ldots, \langle X(\omega, \cdot), h_n \rangle \in B_n),$$

where $h_1, \ldots, h_n$ belong to $H$ and $B_1, \ldots, B_n$ belong to $B([0,1])$, the Borel sets of $\mathbb{R}$. Here $\langle X(\omega, \cdot), h \rangle = \int_{0}^{T} X(\omega, t) \, h(t) \, dt$.

If $S$ denotes the signal and $N$ the noise, the detection problem $(N, S+N)$ is nonsingular if $P_{S+N}$ and $P_N$ are mutually absolutely continuous, that is $P_{S+N}(B) = 0$ if and only if $P_N(B) = 0$ for Borel $B$ in $B(H)$. The singularity of a detection problem on a space of infinite dimension is notoriously unstable (19). Consequently,
for reasons analogous to those which require detection problems to be non-
singular (3, 14, 18), attention shall be restricted to laws which are not only
in the neighbourhood of Gaussian laws but also equivalent to these.

One tool which has proved useful in investigations dealing with robust
statistical procedures is that of contamination (21). Let P be the law of the
chosen model; its contaminations are of the form $P^C = (1-\varepsilon)P + \varepsilon Q$, where $Q$
is some probability measure and $0 < \varepsilon < 1$. If $P = P_X$ and $Q = Q_Y$ and $U$ is a
random variable, independent of $X$ and $Y$, which takes the value 1 with pro-
bability $1 - \varepsilon$ and the value 0 with probability $\varepsilon$, then $P^C$ is the law of $Z =
UX + (1-U)Y$, that is the paths of $Z$ are obtained by sampling among those of
$X$ and $Y$ according to the law of $U$. In this paper we shall use two types of
contaminations. The first type ($T1$-contaminations) is constructed with mea-

sures $Q$ which are Gaussian and the second type ($T2$-contaminations) with
measures $Q$ which are absolutely continuous with respect to $P$ and have a
quadratic form as Radon-Nikodym derivative. The use of such simple conta-
minations has a number of advantages. First of all, their covariance has the
form $R+A$ so that we have a nongaussian generalization of the stability study
within the Gaussian model, as in (19). Moreover, the situation is at the same
time sufficiently simple so that one can obtain analytic information on the pro-
babilities of interest, and sufficiently complex so that one can exhibit features
which do not appear in the Gaussian perturbation model.

The first part of the paper deals with the basic properties of the contamina-
tions just introduced. We show in particular that they share with Gaussian
laws many properties, the main exception being that they do not necessarily
converge weakly to laws of the same type. In the second part of the paper we use these contaminations to study the stability of the detection problem in case of sure signal and additive Gaussian noise. Besides giving a number of quantitative and qualitative results involving actual and nominal probabilities of false alarm, we are able to show that weak convergence alone fails in general to control the behaviour of the parameters which determine the probabilities of false alarm: what really matters is the way in which this convergence takes place. Finally, we show that the type of contamination which is considered plays an important role. Thus contaminations by Gaussian laws result in worse pathologies than contaminations of the second type considered: the former have in a sense "too many degrees of freedom" for weak convergence to control.

We consider here only the detection of sure signals. For some results in the case of stochastic signals, see (7). Proofs are outlined in the appendix.

2. CONTAMINATIONS

P will always denote a Gaussian probability on $B(H)$. It is determined by its mean $m_p$ and its covariance operator $R_p$, and we shall write $P \sim N(m_p, R_p)$. $m_p$ is an element in $H$ and is uniquely determined by the relation

$$E_P(\langle \cdot, h \rangle) = \langle m_p, h \rangle$$

for all $h$ in $H$.

In case $P = P_X$, $m_p$ is the mean function of $X$. The operator $R_p$ is uniquely determined by the relation

$$E_P(\langle \cdot, h \rangle - \langle m_p, h \rangle)(\langle \cdot, k \rangle - \langle m_p, k \rangle) = \langle R_p h, k \rangle$$

for all $h$ and $k$ in $H$.

In case $P = P_X$ and $R_X(s, t)$ is the covariance of $X$, $R_p$ is defined by

$$R_p f(t) = \int_0^T R_X(t, u) f(u) \, du.$$
Let \((a \theta b)(x) = \langle b, x \rangle a\). Then \(R_p\) has the representation

\[
R_p = \sum_{k=1}^{\infty} \lambda_k e_k \theta e_k
\]

where \(\lambda_k > 0\), \(\sum \lambda_k = \infty\), and \(\{e_k\}\) is a complete orthonormal set in the closure of the range of \(R_p\). Since this latter set is the support of \(P\), one can always suppose it is \(H\). \(e_k\) is an eigenvector of \(R_p\) corresponding to the eigenvalue \(\lambda_k\). It has been shown that Gaussian measures \(P\) on the Borel sets of \(L_2(0,1)\) are of the form \(P_X\), \(X\) a Gaussian process with paths in \(L_2(0,1)\) (15). The Fourier transform of \(P\), \(FTP\), is given by the relation

\[
FTP(h) = \mathbb{E}_{P}(\exp(i \langle \cdot, h \rangle)) = \exp(\langle i m_p, h \rangle - \frac{1}{2} < R_p h, h >).
\]

If \(P = P_X\), \(FTP(\cdot)\) is the characteristic function of the random variable \(\langle X(\cdot), h \rangle\), \(h\) real.

For \(T_1\)-contaminations, we choose \(Q \sim N(m_Q, R_Q)\), while for \(T_2\)-contaminations we choose \(dQ = C^{-1} ||A^2(x-a)||^2 dP\). \(A\) is an operator which is linear, continuous, positive and self-adjoint, \(a\) is a fixed element in \(H\) and \(C = \text{trace} \{A(R_p + (m_p - a) \theta (m_p - a))\}^{1/2}\). \(C\) is chosen to be \((c+C)^{-1}\), where \(c\) is nonnegative. We have already indicated that contaminations can be thought of as the laws of processes \(Z = UX + (1-U)Y\), where \(X\) has law \(P\) and \(Y\) has law \(Q\). We are going to add the requirement that \(X\) and \(Y\) have the same set of paths. One way to achieve this is to require that \(P\) and \(Q\) be mutually absolutely continuous, so that they have the same support. In this way, we discard cases such as those for which \(X\) would have continuous paths and \(Y\) discontinuous ones, for which no doubt would exist as to which of \(P\) and \(P^c\) obtains. For examples of \(T_2\)-contaminations, see (8,9). In particular \(T_2\)-contaminations of stationary processes yield harmonizable ones. We thus have

**Proposition 1**
If $P_C$ is a T1-contamination, $P$ and $P_C$ are mutually absolutely continuous if and only if $P$ and $Q$ are. $P$ and its T2-contaminations are always mutually absolutely continuous. /

The first assertion follows from the dichotomy theorem for Gaussian measures (10) and the second is a consequence of a zero-one law for Gaussian measures (2).

Assumption

$P$ and $P_C$ are henceforth assumed to be mutually absolutely continuous. /

We shall first need the mean $m_c$ and the covariance $R_c$ of $P_C$. A direct calculation yields

**Proposition 2**

For T1-contaminations $m_c = (1-\varepsilon)m_P + \varepsilon m_Q$ and

$$R_c = (1-\varepsilon)R_P + \varepsilon R_Q + \varepsilon(1-\varepsilon)(m_p - m_Q)(m_p - m_Q),$$

and for T2-contaminations $m_c = m_p + 2(c+C)^{-1}R_p A(m_p - a)$ and

$$R_c = R_p + 2(c+C)^{-1}R_p A R_p^{-1}R_p A (m_p - a) \Theta (m_p - a) A R_p. /$$

**Remark 1**

From Proposition 2 it follows that both types of contaminations have a mean and a covariance of the same form, that is

$$m_c = m_p + R_p^\frac{1}{2}b$$

and

$$R_c = R_p^\frac{1}{2}(I + T + t \Theta) R_p^\frac{1}{2}.$$

($P$ and $Q$ have been chosen equivalent in the T1 case so that in particular $m_Q$ belongs to the range of $R_P^\frac{1}{2}$). These are essentially the conditions which ensure that a Gaussian law with mean $m_c$ and covariance $R_c$ is equivalent to $P$ (16), so that, even if $P_C$ is wrongly assumed to be Gaussian, there is no way to discriminate surely between $P$ and $P_C$. 

For every fixed $h$ in $H$, one may define a real random variable $X(h)$ by setting $X(h)(x) = \langle x, h \rangle$, $x$ in $H$. For $T_1$-contaminations, the law of $X(h)$ with respect to $P^C$, defined by

$$P^C_X(B) = P^C_X(x \in H : X(h)(x) \in B), B \in B(\mathbb{R}),$$

is the obvious mixture of the laws of $X(h)$ with respect to $P$ and $Q$. For $T_2$-contaminations one has

Proposition 3

Let $a(h) = 2(c+C)^{-1} \langle R_P A(m_p-a), h \rangle$, 

$\beta(h) = 2(c+C)^{-1} \langle R_P A R_P h, h \rangle$, 

$\phi(t) = \exp(it\langle m_p, h \rangle - \frac{1}{2}t^2 \langle R_P h, h \rangle)$. 

The characteristic function $\phi$ of the random variable $X(h)$ with respect to $P_C$, a $T_2$-contamination, is given by

$$\phi(t) = (1 + i\alpha(h)t - \frac{1}{2}\beta(h)t^2)\phi(t).$$

Let $\alpha(h) = 1 - (c+C)^{-1} \langle R_P A R_P h, h \rangle / \langle R_P h, h \rangle$, 

$\beta(h) = 2(c+C)^{-1} \langle R_P A(m_p-a), h \rangle / \langle R_P h, h \rangle$, 

$\gamma(h) = (c+C)^{-1} \langle R_P A R_P h, h \rangle / \langle R_P h, h \rangle^{-2}$, 

$G(y) = \text{Gaussian density with mean } \langle m_p, h \rangle \text{ and variance } \langle R_P h, h \rangle$. 

The density $g$ corresponding to $\phi$ is given by

$$g(y) = (a(h) + b(h)(y - \langle m_p, h \rangle) + c(h)(y - \langle m_p, h \rangle)^2)G(y).$$

Remark 2

Suppose that $a = m_p$ and let $z = (y - \langle m_p, h \rangle) / \langle R_P h, h \rangle^{\frac{1}{2}}$, 

$$L^2 = (c+C)^{-1} \langle R_P A R_P h, h \rangle / \langle R_P h, h \rangle$$

and $K^2 = (2\pi \langle R_P h, h \rangle)^{\frac{1}{2}}$. 

The density $g$ of Proposition 3 can then be written
\[ g(y) = y(z) = K^2(1-L^2+L^2z^2)e^{\exp(-z^2)}. \]

If \( L^2 \leq 1/3 \), for which a sufficient condition is \((c+C)^{-1}||R^zAR^z|| < 1/3\), \( g \) has a unique maximum for \( z = 0 \). Otherwise \( g \) presents two symmetric peaks. Laws with such densities have been shown to model some types of noise (17). Since \( R^zAR^z \) is a compact operator, \( <R^zAR^zh,h>/<Rh,h> = <R^zAR^zk,k> \), with \( k = R^zh/||R^zh|| \), will be small in all but a finite number of directions. This shows the kind of departure from normality one can expect. A sink which would appear in the density of all functionals would eventually touch the origin, making the law singular with respect to the Gaussian one. To obtain two peaks with TI-contaminations, one would need distinct means. The peaks would then appear in every direction, but the sink would be controlled in size. Another interesting feature of the density \( g \) is that it is normal for \( h \) in the kernel of \( R^zAR^z \). The measure \( P^C \) can thus exhibit a Gaussian behaviour in many directions without actually being Gaussian. This "Gaussianness" is controlled by the range of \( A \). This feature is absent in the case of TI-contaminations.

One space of interest in linear inference problems is the closure in the real \( L_2 \)-space \( L_2(P^C) \) of the set \( \{X(h), h \in H\} \) of random variables. This subspace is called the linear space of \( P^C \) and is denoted \( L(P^C) \). This terminology arises as follows: if \( P^C = P^*_\lambda \), linear operations on \( X \) are of the form \( <X(\cdot, \cdot), h> = X(h) \). \( L(P) \) contains only normal random variables. The next result states that \( L(P^C) \) also contains random variables which all have a law of the same type.

Let \( P^C \) be a contamination and \( S_c \) be the operator \( R_c+mc \cdot m_c \). If \( f \) belongs to \( L(P^C) \), the following relation defines an element of \( H \) denoted \( h(f) \):

\[ h(f) = \int_x xf(x)P^C(dx). \]
For two such elements, \( h(f) \) and \( h(g) \), one can define the inner product
\[ \langle h(f), h(g) \rangle_{H(P^c)} = \int (fg). \]

Let \( H(P^c) = \{ h(f), f \in L(P^c) \} \) and define \( U_{p^c} : L(P^c) \to H(P^c) \) by \( U_{p^c}(f) = h(f) \).

\( H(P^c) \) is then a Hilbert space, contained in \( H \) as a subset, and \( U_{p^c} \) is a unitary operator. As a set, \( H(P^c) \) is the range of \( S^2_c \) and is isomorphic to the reproducing kernel Hilbert space of the process \( (X(h), h \in H) \) (Lemma 1, Appendix).

Let now \( B \) be the operator from \( L(P^c) \) to \( H \) defined by \( Bf = \int_{p^c} S_{p^c}^{-1} U_{p^c} f \). \( B \) is well defined and continuous (Lemma 4, Appendix). Since \( \langle R_{p^c} h, h \rangle < \langle S^2_c h, h \rangle \), \( R_{p^c}^{\frac{1}{2}} = S^2_c W_c \), where \( W_c \) is a bounded operator on \( H(5) \). Recalling that \( m_c = m_p + R_{p^c}^{\frac{1}{2}} b \) and that \( R_c = \int_{p^c} R_{p^c}^{\frac{1}{2}} (I + T + \varepsilon; \Theta t) R_{p^c}^{\frac{1}{2}} = \int_{p^c} R_{p^c}^{\frac{1}{2}} (I + T) R_{p^c}^{\frac{1}{2}} \) (Proposition 2 and Remark 1), one has

**Proposition 4**

Let \( f \) belong to \( L(P^c) \) and \( h, h_n \) belong to \( H \). Suppose that \( U_{p^c} f = S^2_c h \) and \( h_n \) converge to \( h \) in \( H \).

\[ h = \lim_{n} S^2_c h_n \] (We have assumed, which is no restriction, that \( P \) has full support. This implies that \( H \) is the closure of the range of \( S^2_c \)). Then:

1) there exists \( m \) in \( H \) such that \( \lim_{n} \langle m_p, h_n \rangle = \langle m, h \rangle \) (If \( m_p \) belongs to the range of \( R_{p^c}^{\frac{1}{2}} \), \( \langle m, h \rangle = \langle R_{p^c}^{\frac{1}{2}} m_p, Bf \rangle \));

2) \( E_{p^c} (f) = \langle m, h \rangle + \langle b, Bf \rangle \) (If \( m_p \) belongs to the range of \( R_{p^c}^{\frac{1}{2}} \), \( E_{p^c} (f) = \langle b + R_{p^c}^{\frac{1}{2}} m_p, Bf \rangle \));

3) \( V_{p^c} (f) = \| (I + T)^{\frac{1}{2}} Bf \|^2 = \| W_c h \|^2 \);

4) For \( T1 \)-contaminations, the characteristic function \( \phi \) of \( f \) is given by
\[ \phi(t) = (1-\epsilon)\phi_1(t) + \epsilon\phi_2(t), \] where \( \phi_1 \) is the characteristic function of a
\[ N(m, h, \|Bf\|_2^2) \] -random variable and \( \phi_2 \) is the characteristic function of a
\[ N(m, h, -\epsilon R^2_P(m - m_Q), Bf, \|W^*_c\|_2^2) \] -random variable; for T2-contaminations, \( \phi(t) = (1 + \text{i}t\langle Bf, Bf \rangle - \|T^2_Bf\|_2^2)\phi_1(t). \)

This result yields a useful computational tool, stated as a corollary.

**Corollary**

Let \( f_1, \ldots, f_n \) belong to \( L(P) \) and \( h_1, \ldots, h_n \) solve \( \sum_p f_i = S^3 h_i \). Write \( m_1, m_2, m_3 \) for the \( n \)-vectors with respective entries \( \langle m, h \rangle, -\langle B^* B, f \rangle \) and \( M_1, M_2, M_3 \) for the \( (n \times n) \)-matrices with respective entries \( \langle B^* B, f \rangle \) and \( \langle B^* T B, f \rangle \), where \( T \) solves \( L(P) \) for the \( (n \times n) \)-matrices with respective entries
\[ \phi^*_i, \phi^*_j \]
\[ \phi^*_i \]

Then, if \( M_1 \) has an inverse, \( (f_1, \ldots, f_n) \) has a density \( g \)
given by, in case of T1-contaminations, \( g(x) = (1-\epsilon)g_1(x) + \epsilon g_2(x) \), where \( g_1 \)
is the density of a \( N(m_1, M_1) \)-random vector and \( g_2 \) is the density of a \( N(m_2 - m_1, M_2) \)
random vector, and, in case of T2-contaminations,
\[ g(x) = (1-\text{trace}(M_1^{-1} M_2) - \langle x - m_1, M_1^{-1} m_3 + \epsilon M_3 M_1^{-1} (x - m_1), (x - m_1) \rangle)g_1(x). \]

There are two reasons which explain the effectiveness of Gaussian laws for solving problems of detection and estimation. The first is that their nonlinear
space \( L_2(P) \) is the direct sum of the symmetric tensor powers of their linear
space \( L(P) \) [12], which roughly means that the elements in \( L_2(P) \) are limits
of sums of polynomials of increasing order evaluated at an orthonormal basis for
\( L(P) \). The second reason is that the set of Gaussian laws is closed for the topo-
logy of weak convergence. It can be checked that the first result remains true
for T1- and T2-contaminations. The second however no longer obtains: the set of contaminations is not closed for weak convergence, as the following results show, and it is at the boundary that pathologies occur.

**Proposition 5**

Let \((P^n_c, n \in N)\) be a family of T1-contaminations which converge weakly to the probability \(M\). If \(\lim \inf \varepsilon_n = 0\) or \(\lim \sup \varepsilon_n = 1\), \(M\) is Gaussian. Otherwise \(M = (1-u)M_1 + uM_2\), where \(M_1\) and \(M_2\) are Gaussian and \(0 < u < 1\). \(M\) is not necessarily a T1-contamination and is Gaussian if and only if \(M_1 = M_2\) (in particular \(M_1\) and \(M_2\) need not be mutually absolutely continuous).

**Proposition 6**

Let \((P^n_c, n \in N)\) be a family of T2-contaminations which converge weakly to the probability \(M\). \(M\) has then a characteristic function \(\phi\) given by

\[
\phi(h) = (1+i\langle d, h\rangle - \langle S^{\frac{1}{2}}V S^{\frac{1}{2}}h, h\rangle) \exp(i\langle n, h\rangle - \frac{1}{2}\langle Sh, h\rangle),
\]

where \(S\) is a linear operator which is compact, nonnegative and selfadjoint and \(V\) is linear and bounded (to have a T2-contamination, one needs an \(S\) which has finite trace and a \(V\) of the form \(S^{\frac{1}{2}}WS^{\frac{1}{2}}\)). \(S\) has finite trace in one of the following two cases:

1. \(0 < \lim \inf \{C_n/(c_n + C_n)\} \lim \sup \{C_n/(c_n + C_n)\} < 1\);
2. if \(a_n(h)\) is zero when \(||R^{\frac{1}{2}}n h||\) is and \((c_n + C_n)^{-1} (c_n + C_n)^{-1} (\langle R A, R h, h\rangle/\langle R h, h\rangle)\) otherwise and if \(b_n(h)\) is zero when \(||R^{\frac{1}{2}}n h||\) is and \(b_n(h)/\langle R h, h\rangle\) otherwise, then \(\sup(a_n(h), n, h) < 1\) and \(\sup(b_n(h), n, h) < 1\).

3. **STABILITY OF SINGULARITY**

Consider two Gaussian measures \(P_1\) and \(P_2\) representing respectively the law of the noise and the law of the received signal in some detection problem. Detec-
tion is nonsingular if and only if \( m_1 - m_2 \) belongs to the range of the square root of \( \frac{1}{2} (R_1 + R_2) \) and if \( R_2 = R_2^2 (I+T)R_2^2 \) with \( T \) Hilbert-Schmidt and \( \sigma(T) > 1 \) (\( \sigma(T) \) denotes the spectrum of \( T \)) (16). Suppose now the actual laws one deals with are \( P_1^C \) and \( P_2^C \), \( T_1 \)- or \( T_2 \)-contaminations, possibly of a different type. The following result proves that the criterion for singularity which applies in the Gaussian case applies to \( T_1 \)- and \( T_2 \)-contaminations as well, so that singularity is stable over the class of \( T_1 \)- and \( T_2 \)-contaminations.

**Proposition 7**

Let \( P_1^C \) and \( P_2^C \) be \( T_1 \)- or \( T_2 \)-contaminations with respective means \( m_{c,1} \) and \( m_{c,2} \) and covariances \( R_{c,1} \) and \( R_{c,2} \). The detection problem \( (P_1^C, P_2^C) \) is either singular or nonsingular and it is singular if and only if \( m_{c,1} - m_{c,2} \) belongs to the range of the square root of \( \frac{1}{2} (R_{c,1} + R_{c,2}) \) and if \( R_{c,2} = R_{c,2}^2 (I+S)R_{c,2}^2 \), \( S \) Hilbert-Schmidt and \( \sigma(S) > 1 \). Furthermore, if the detection problem is nonsingular, 
\[
\frac{dP_2^C}{dP_1^C} = \frac{(dP_2^C/dP_2^C)}{(dP_1^C/dP_1^C)} \frac{(dP_1^C/dP_1^C)}{(dP_1^C/dP_1^C)} .
\]

4. **DETECTION OF SURE SIGNALS**

One observes a sure function which can have two possible sources: it is either a noise sample \( N \) or a distorted signal sample \( s + N \), where \( s \) is known. The nominal model is that model for which the law \( P \) of the noise \( N \) is Gaussian with mean zero and covariance \( R_P \) of the form \( \sum e_n \Theta e_n \), as defined in the introduction. The actual model is the model for which the law of \( N \) is \( P^C \), with mean zero. The nominal model represents the model actually used and the actual model represents the system which is being investigated, the "true" model. The corresponding detection problem is nonsingular whenever \( s \) belongs to the range of the square root of \( R_P \) (6), which is isomorphic to the
reproducing kernel Hilbert space of $R_p$ (Lemma 1, appendix). Since $R_p$ and $R_p^c$ have square roots with identical ranges (Lemma 4, appendix), the detection problem is nonsingular for the actual as well as for the nominal model. It shall henceforth be assumed that detection problems are nonsingular.

The maximum likelihood decision function for the nominal model is given by

$$D(X) = \frac{1}{n} \left( s_n / \lambda_n \right) X(e_n^\dagger) \quad \text{with} \quad s_n = \langle s, e_n \rangle \quad \text{and} \quad X(e_n^\dagger) = \langle y, e_n \rangle \quad (19).$$

A. The case of $T=1$-contaminations

By choice, $P^c = (1-\epsilon)\mathcal{P} + \epsilon Q$ with $Q \sim N(m_Q, R^\frac{1}{2}Q)$ and $P$ and $Q$ mutually absolutely continuous. Then $R^\frac{1}{2}Q = R^\frac{1}{2}P (I+T_Q)^\frac{1}{2}R^\frac{1}{2}P Q$ Hilbert-Schmidt and $Q(T_Q) > 1$ (16).

Since $m_p = m_c = 0$, $m_Q = 0$ (Proposition 2). Let $e_n, e_{n+1}, \ldots, e_{n+p}$ span the range of the projection $\pi_n, P' C_1 = ||\pi_n, P' R^{-\frac{1}{2}}s||^2 + \epsilon ||T_Q^\frac{1}{2}p_r R^{-\frac{1}{2}}s||^2$ and $C_2 = ||\pi_n, P' R^{-\frac{1}{2}}s||^2$.

Then $E_p^c \left\{ \sum_{i=1}^{n+p} (s_i / \lambda_i) X(e_i) \right\}^2$ equals $C_1$ if $X = N$ and equals $C_1 + C_2$ if $X = s+N$.

so that $D(X)$ belongs to $L_2(P^c)$. When $X = N$, let $Y_1$ be a normal random variable with mean zero and variance $||R_p^{-\frac{1}{2}}s||^2$, and, when $X = s+N$, let $Y_1$ have the same law and variance, but mean equal to the variance. Similarly, let $Y_2$ be normal with, when $X = N$, mean zero and variance $||R_Q^{-\frac{1}{2}}s||^2$, and, when $X = s+N$, mean $||R_p^{-\frac{1}{2}}s||^2$ and variance $||R_Q^{-\frac{1}{2}}s||^2$. Let $f_1$ and $f_2$ be the respective densities of $Y_1$ and $Y_2$. Then $D(X)$ has density $f(x) = (1-\epsilon)f_1(x) + \epsilon f_2(x)$. Finally let $p^a = p^c(D(X) > a | N)$ and $p^n = p(D(X) > a | N)$. Then $p^a - p^n = \epsilon (Q(D(X) > a | X = N) - p^n)$.

$\phi$ denotes the distribution function of a $N(0,1)$-random variable. The polar decomposition (22) yields $R^{-\frac{1}{2}}Q = R^{-\frac{1}{2}}P (I+T_Q)^\frac{1}{2}U$, where it is no restriction to suppose $U$ unitary, so that $R^{-\frac{1}{2}}Q = U^\dagger (I+T_Q)^\frac{1}{2}R^{-\frac{1}{2}}P$. Let $\varepsilon = a / ||R_p^{-\frac{1}{2}}s||^2$. 

\( \tilde{s} = R^{-1/2} s / \| R^{-1/2} s \|, \tau = \| (1+T_Q)^{-1} \tilde{s} \| \). Then \( \rho^n - \rho^a = \epsilon \{ \phi(t) - \phi(\tau) \} \).

**Proposition 8**

Fix \( s \) and \( \epsilon > \phi^{-1}(3/4) \). Then

1) \( \Delta = \sup \{ |p^a - p^n|, p^c \text{ a T1-contamination} \} = \phi(\tau)-\frac{3}{4} \);

2) there exists \( (p^c, k \in \mathbb{N}) \) such that \( W-lim_{k} p^c_k = P \) (\( W- \) indicating weak convergence) and \( \lim_{k} |p^a_k - p^n_k| = \Delta \).

The reason for 2) is that \( P \) and \( p^c \) can be close in the sense of weak convergence as well as close to being orthogonal. The result also shows that the discrepancy between \( p^a \) and \( p^n \) can increase with the level of the test.

**Proposition 9**

There exist sequences \( (p^c, s_k, k \in \mathbb{N}) \) such that \( W-lim_{k} p^c_k = P \), \( \lim s_k = 0 \) and \( \lim_{k} |p^a_k - p^n_k| = 1-\phi(\tau) \).

The reason for this phenomenon is that the class of admissible signal is too large for \( |p^a - p^n| \) to be controlled by weak convergence.

Another way to compare \( p^a \) and \( p^n \) is to look at the ratio \( p^a / p^n \). One has

\[ p^a / p^n = 1 - \epsilon + \epsilon(1 - \phi(\tau))/(1 - \phi(\tau)) \].

As a function of \( t \), \( p^a / p^n \) decreases from \( 1 - \epsilon + \epsilon/(1 - \phi) \) to \( 1 - \epsilon \). One thus has \( p^a < p^n \) if \( t > 1 \) and \( p^a > p^n \) if \( t < 1 \). When the level of the test increases, the behaviour of \( p^a / p^n \) depends on the relative behaviour of \( \epsilon \), \( \phi \), and \( t \). Suppose thus that \( t \) goes to infinity faster than \( \epsilon \) goes to infinity: \( p^a / p^n \) is unbounded, so that \( p^n \) goes to zero faster than \( p^a \). If now \( \epsilon \) goes to 1 and \( t \) goes to infinity, \( p^a / p^n \) goes to zero so that \( p^a \) goes to zero faster than \( p^n \). We have seen in Propositions 8 and 9 that this behaviour can occur as \( p^c_k \) converges weakly to \( P \).

As the examples considered show, weak convergence, when \( \epsilon \) does not go
to zero, depends on the behaviour of $R_1^T R_2^T$ whereas the behaviour of $|p^n - p^a|$ and $p^a, p^n$ depends on that of $T$, and these behaviours are not necessarily related.

Let now $P_c$ be the law of $s+N$ and $o^a$ be the probability of false alarm for the maximum likelihood detector for the actual detection model, that is $p^a = P_c (dP_c/dP_s)$. Again one would like to know to what extent $p^n = P (dP_s/dP_a)$ is a "good" approximation of $p^a$. Let $k = \frac{1-\epsilon + \epsilon (dQ/dP) - \frac{1}{2} \epsilon \exp (-\frac{1}{2} \langle G x, x \rangle)}{1+T_Q}$, so that $k$ represents the effect of contamination. The following assumptions allow one to obtain a manageable form for $k$ and will be made for the remaining part of this section.

**Assumptions**

1) $s$ belongs to the range of $R_p$ (which is a subset of the range of $R_p$).
2) $R_p$ and $R_Q$ have the same range.
3) $R_Q^{-1} R_p^{-1}$ has a symmetric closure $G$ whose domain has $P$-probability one.

In particular, when $R_p^{-1} R_Q^{-1}$ is bounded on the range of $R_p$, $G$ is bounded and self-adjoint (16). Let $k(x) = (1-\epsilon) \det \frac{1}{2} (I+T_Q) + \epsilon \exp (-\frac{1}{2} \langle G x, x \rangle)$. Then $k(x) = K(x-s)/K(x) (16)$. Thus, whenever $s$ belongs to the kernel of $G$, the actual and the nominal probabilities are the same. One however has

**Proposition 10**

The range of $k$ can be $\{x \in \mathbb{R} : 0 < x < \infty \}$.

Proposition 10 shows that the values taken by the actual and nominal likelihoods can be quite different. Since the global behaviour of $k$, as a function of $\epsilon, G$ and $T$, is difficult to analyze, one may try to evaluate $k$ at critical points. There are two of these of particular interest, 0 and $s$, at which one presumably would deci-
de that respectively no signal or a signal has been sent. Since \( k(s) = 1/k(0) \), it suffices to consider \( k(0) \). It is shown in the next proposition that \( k(0) \) can be as large as
\[
\exp\left\{ \frac{1}{2} \left| R^{-\frac{1}{2}} s \right|^2 \right\}
\]
which is the value taken by the nominal likelihood when the exact signal has been received. The actual likelihood has then the value one, whereas the nominal likelihood can be very small.

**Proposition 11**

\[
\sup \{k(0)|e.Q\} \geq \exp\left\{ \frac{1}{2} \left| R^{-\frac{1}{2}} s \right|^2 \right\}.
\]

Here again, the bound can be approximated when \( P^c \) is close to \( P \) in the weak sense. Formulae for \( p^a \) are quite complicated and only in particular cases can one obtain global bounds for \( p^a \) in terms of \( p^n \). Here is one case which shows what kinds of restrictions are then required.

**Proposition 12**

Let \( b_k = 1+(e_k/(1-e_k))\det^2\{1+T_{Q_k}\} \). Suppose that \( G_k \) is nonnegative and that
\[
b_k \leq M < \infty.
\]
Then, for \( a>0 \) and \( W\)-lim \( P^c_k = P \),
\[
P\{dP_s/dP>aM\} \leq \lim \inf k^{-a} P(\{dP^c_s/dP^c_s>a\}) \leq \lim \sup \sup k^{-a} P(\{dP^c_s/dP^c_s>a\}) \leq P\{dP_s/dP>aM\}.
\]

Thus a uniform bound on the admissible perturbations is required and the bounds obtained are good only if the perturbation is uniformly small.

**B. The case of T2-contaminations**

The features of T1-contaminations exhibited above indicate that stability results can only be obtained by restricting the class of signals and/or knowing some a priori uniform bounds on the perturbations one must consider (we have in particular assumed in Propositions 10, 11 and 12 - see the "Assumptions" preceding these statements - that the signal belongs to the range of \( R_p \), rather than to the range of \( R_p^{1/2} \), which is imposed by a nonsingular problem). In this sense, T1-contamina-
tions have "too many degrees of freedom". One can thus attempt to obtain stability results by restricting the class of contaminations considered. This is what one achieves by introducing T2-contaminations. Whereas $\epsilon$ and Q act as independent variables for T1-contaminations, for T2-contaminations, $\epsilon$ depends on Q in the sense that $\epsilon = C/(c+C)$, where C is a function of the parameters which determine Q (see in particular the proof of Proposition 13). Intuitively, T2-contaminations are obtained from T1-contaminations by keeping the first terms of the Taylor expansion of the likelihood of Q with respect to P. This indicates that classes of contaminations intermediate between T1- and T2-contaminations can be obtained by keeping more terms in the Taylor expansion considered above.

It is again assumed that $m_c = 0$, which is achieved by setting $R_p Aa = 0$. As in the case of T1-contaminations, one may check that $D(X)$ belongs to $L_2(P^c)$. Let $Y_1$ be the same random variable as in A, and $f_1$ be its density. Let $\epsilon = ||A^{1/2}s||^2/(c+C)||R_p^{1/2}||^2$ and $f_2(x) = x^2f_1(x)/||R_p^{-1/2}||^2$. With respect to $P^c$, $D(X)$ has then a density $f$ given by the relation $f(x) = (1-\epsilon)f_1(x) + \epsilon f_2(x)$. Let $I_A$ be the indicator of the closure of the range of A and $s, p^a, p^n$ and $s$ have the same meaning as in A. One then has,

$$K = ||A^{1/2}R_p^{1/2}s||^2/(c+\text{trace}(AR_p)||A^{1/2}a||^2),$$

$p^a - p^n = K I_A (s)^2 (x^2-1) \exp(-\frac{1}{2}x^2) dx/\sqrt{2\pi}$. The next proposition shows that the pathologies illustrated in Proposition 8 and 9 cannot occur in the case of T2-contaminations. The reason is clear: $\epsilon$ is not independent of Q. There is a restriction however: s is required to belong to the range of $R_p$.

**Proposition 13**

Let $s$ belong to the range of $R_p$ and $\text{W-lim} P_k^c = P$, then $\lim p^a_k = p^n$. / 

A simple limit calculation (4) gives the relative behaviour of $p^a$ and $p^n$: $p^n$ tends
to zero faster than $p^a$.

**Proposition 14**

For fixed $\varepsilon$ and large $n$, $p^2/n \sim \varepsilon^2$.

Finally, we consider again the actual likelihood and the actual probability of false alarm, still denoted $p^a$. In contrast to the case of T1-contaminations, the function $k$ of Proposition 10 is bounded for T2-contaminations. Indeed, if $K(x) = c+\frac{1}{|x-a|^2}$

$$dPc = \left(\frac{K}{(c+C)}\right)dP$$

so that, if $k(x) = K(x-s)/K(x)$, $dP^C_s/dP^C = k(dP_s/dP)$. Then

**Proposition 15**

Let $a_A$ belong to the kernel of $A$, $\lambda_m = \frac{1}{m}(1+(1+c/|A_s,s|)\frac{1}{2})$ and $\lambda_M = \frac{1}{M}(1-(1+c/|A_s,s|)\frac{1}{2})$

Then $k(a+\lambda_m s + a_A) < k(x) < k(a+\lambda_M s + a_A)$.

From Proposition 15 one can get a result which is more satisfactory than Proposition 12, but which is still insufficient in the sense that it depends on a priori knowledge on the perturbations, which is unlikely to be available.

**Proposition 16**

If $s$ is in the range of $R_p$ and $\lim \sup C_n/c_n < \infty$, then, when $W-lim P^C_n = P$

$$\lim P^C_n(dP^C_{n,s}/dP^C_{n,a}) = P(dP_s/dP^a).$$

5. **SUMMARY**

Let $p_1 = P(dP_s/dP^a)$, the probability of false alarm for the Gaussian model,

$p_2 = P^C(dP_s/dP^a)$, the probability of false alarm for the "exact" model when the detector is the likelihood for the Gaussian model, and $p_3 = P^C(dP^C_s/dP^C_{n,a})$, the probability of false alarm for the "exact" model. We have considered the case of a Gaussian detection model which is only an approximation to an "exact" contaminated model and compared the quantities $p_1$, $p_2$, and $p_3$ when the law $P^C$ of the "exact" model is close in the topology of weak convergence to $P$, the law of the
nominal model, which in the present case is Gaussian. We have seen that the quantities $|p_1-p_i|$ and $p_1/p_i$, $i=2,3$, depend on the type of contamination as well as on the distance between $P$ and $P^C$. Proposition 9 shows for example that one can have, for a T1-contamination, $|p_1-p_2| > e > 0$, though $s$ and the distance between $P$ and $P^C$ are negligible. Proposition 16 shows however that under fairly mild restrictions, $|p_1-p_3|$ is small as soon as $P$ and $P^C$ are close and $P^C$ is a T2-contamination. We have also shown that looking at the likelihood can be rather misleading when the model is not 'exact.'

6. OUTLINE OF PROOFS

The proof of Proposition 4 is based on four lemmas which we first state.

Lemma 1

Let $S_C$, $H(P^C)$ and $U$ be defined as in Proposition 4. For $f$ in $L(P^C)$ define

$$F_f: H \to \mathbb{R}$$

by the relation $F_f(g) = \int_H f(x)X(h)(x)P^C(dx)$. Let $H(S_C) = \{F_f, f \in L(P^C)\}$,

$$<F_f,F_g>_{H(S_C)} = \int_{P^C} (fg)$$

and $U_S: L(P^C) \to H(S_C)$ be defined by the relation $U_S f = F_f$.

Then

1) $H(S_C)$ is a Hilbert space (the reproducing kernel Hilbert space of $X$) and $U_{S_C}$ is a unitary operator,

2) $||f||_{H(P^C)} \leq ||S_C^{\frac{1}{2}}||_{H(P^C)} ||h||_{H(P^C)}$.

3) the two sets $H(P^C)$ and range $(S_C^{\frac{1}{2}})$ are equal,

4) $<S_C^{\frac{1}{2}}h,S_C^{\frac{1}{2}}k>_{H(P^C)} = <h,k>$, so that $S_C^{\frac{1}{2}}: H \to H(P^C)$ is unitary,

5) if $f_n = S_C^{\frac{1}{2}}h$ and $f_n = X(h_n)$, then $f = \lim f_n$ if and only if $h = \lim S_C^{\frac{1}{2}}h_n$. in $L(P^C)$ and $H$ respectively.
Lemma 1 follows from the definitions and first principles. These also yield

**Lemma 2**

If $R_p$ is injective, $\text{trace}(AR_p) = \frac{1}{2} \left| \left| R_p \frac{1}{2} A R_p \frac{1}{2} \right| \right|$ if and only if $A = h @ h$./

**Lemma 3**

If $P^C$ is a $T2$-contamination and $T+t@t$ is the operator of Proposition 2,

$s(T+t@t) > -1.$/

**pf.** It suffices to prove that $-1$ is not an eigenvalue of $T+t@t$ and this follows from the inequality $1 + (1 + \langle T, h h \rangle )^{\frac{1}{2}} - 1 \leq | |b| | < 1.$/

**Lemma 4**

Let $\tilde{T} = T + t @ t$. Then

1) $R^\frac{1}{2}_C$ and $R^\frac{1}{2}_P$ have the same range and $R^\frac{1}{2}_C = R^\frac{1}{2}_P (I+T)^{-\frac{1}{2}} V_C$, where $V_C$ is unitary.

2) the map $B: L(P^C) \to H$ defined by the relation $Bf = R^\frac{1}{2}_P S^{-1} U f$ is linear and continuous./

**pf.** 1) follows from Lemma 3, (5) and the polar decomposition (22). $B$ is well defined on $H^*$ and, if $f$ is in $H^*$, $| |Bf| | \leq K | |f| | _{L(P^C)}$./

**Proposition 4**

**pf.:** Let $f_n = X(h_n)$. Then $f = \lim f_n$ by Lemma 1, so that, by Lemma 4, $Bf = \lim R^\frac{1}{2}_P h_n$. Furthermore, $E_{P^C}(f) = \lim \langle m_p, h_n \rangle + \langle b, Bf \rangle$, so that $\lim \langle m_p, h_n \rangle$ is a continuous linear functional on $L(P^C)$. The first part of 1) follows by the Riesz theorem and Lemma 1. If $m_p$ belongs to the range of $R^\frac{1}{2}_P$, $\langle m_p, h_n \rangle$ tends to $\langle R^\frac{1}{2}_P m_p, Bf \rangle$. The result on the variance follows from Lemma 4 and that on the characteristic function from Proposition 3./

**Proposition 5**

**pf.:** For $\epsilon, x, \text{ and } y$ in the open unit interval, let $f_\epsilon(x, y) = (1-\epsilon) x + \epsilon y$. If $0 < \alpha < 1$, there is a $\beta$ in the open unit interval such that $f_\epsilon(x, y) > 1 - \beta$ implies $x > 1 - \alpha$ and
y \geq 1-\alpha. Consequently, if \( \{P^c_\lambda, \lambda \in \Lambda\} \) is relatively compact, so are \( \{P^c_\lambda, \lambda \in \Lambda\} \) and \( \{Q^c_\lambda, \lambda \in \Lambda\} \), provided \( \varepsilon_{\lambda} = \varepsilon \) and \( \varepsilon \) is in the open unit interval.

Suppose now \( \{P^c_\lambda, \lambda \in \Lambda\} \) converges weakly to \( P^c \) so that \( 0 < \lim \inf \varepsilon_{\lambda} < \lim \sup \varepsilon_{\lambda} < 1 \). If \( \{P_n, n \in N\} \) is a subsequence of \( \{P^c_\lambda, \lambda \in \Lambda\} \), let \( \{\varepsilon_n\} \) converge to \( \varepsilon \) in the open unit interval. Set \( P^c_n, P^c = (1-\varepsilon)P_n + \varepsilon Q_n \). \( \{P^c_n\} \) converges weakly to \( P^c \).

The remarks at the beginning imply that \( \{P_n\} \) is relatively compact. The proposition follows. To see that the limit is not necessarily a contamination, consider \( \{P^c_n\} \) with \( P^c_n = (1-\varepsilon)P + \varepsilon Q_n \). Set \( P \sim N(0, R_P), Q_n \sim N(0, R^{1/2}_P \{1-T_n\} R_P^{1/2}) \), \( T_n = ((1/n)-1)u\theta u \), \( u \) a unit vector.

Proposition 6

pf. : Because of (9), \( P^c \) has characteristic function given by the relation

\[
\phi(h) = \{1+i(a, h) - \langle Bh, h\rangle\} \exp\{i\langle m, h\rangle - \langle Sh, h\rangle\},
\]
where \( B \) and \( S \) are bounded linear operators. Since the convolutions \( \{P^c_n, P^c\} \) converge weakly to \( \{P^c_n, P^c\} \), the latter has characteristic function \( |\phi(h)|^2 = \langle a, h\rangle^2 + (1-\langle Bh, h\rangle)^2 \exp\{-\langle Sh, h\rangle\} \).

\( |\phi(h)|^2 \) is continuous in the \( S\)-topology and thus, if \( \{x_n\} \) converges weakly to zero, \( \lim |\phi(x_n)|^2 = 1. \) Consequently, \( \lim \sup \langle Sx_n, x_n\rangle < \infty \). If this limit is positive, choosing if necessary a subsequence, one may assume that \( \lim \langle Sx_n, x_n\rangle = \sigma^2 > 0 \).

One then has that \( \lim (1-t^2\langle Bx_n, x_n\rangle) = \exp(\sigma^2 t^2) \), so that \( \lim \langle Bx_n, x_n\rangle = 0 \) and consequently that \( \lim \langle Sx_n, x_n\rangle = 0 \). \( S^{1/2} \) is thus compact and thus so is \( S \). From the inverse Fourier transform of \( \phi \), one obtains that \( \langle Bh, h\rangle \geq \langle Sh, h\rangle \), so that

\[
B = S^{1/2}TS^{1/2}, \; T \text{ bounded}.
\]

1) yields, as in the proof of Proposition 5, that \( \{P_n\} \) is relatively compact, so that \( \{R_n\} \) is compact. 2) yields that \( P^c_n \) has a Fourier transform \( \phi_n \) such that

\[
|\phi_n(h)|^2 \leq \exp\{-1-\delta\} \langle R_n h, h\rangle\}, \text{ which also implies that } \{R_n\} \text{ is compact. That}
S has finite trace follows then from (20).

**Proposition 7**

**pf.** Recall that P and PC have been assumed equivalent. Using the result (1) "λ and u orthogonal and λ and v equivalent imply u and v orthogonal", one can see that the problem (P₁, P₂) is either singular or nonsingular and that it is nonsingular if and only if the problem (P₁, P₂) is nonsingular.

Suppose then that the problem (P₁, P₂) is nonsingular. The operators R₁, R₂, \(\frac{1}{2}(R_1+R_2)\) have all a square root with the same range (16) and, by Lemma 4 of 6., Rc, i and Ri, i=1, 2, have a square root with the same range. Thus \(m_{c,1}^{-1}m_{c,2} = m_1^{-1}m_2^{-1}b_1^{-1}R_1^{-\frac{1}{2}}b_2^{-1}R_2^{-\frac{1}{2}}\) belongs to the range of the square root of \(i(R_{c,1}+R_{c,2})\). The polar decomposition and the equality \(R_2 = R_1^{-\frac{1}{2}}(I+T)R_1^{-\frac{1}{2}}\) yield \(R_2^{-\frac{1}{2}} = R_1^{-\frac{1}{2}}(I+T)^{\frac{1}{2}}V\), where V is unitary (R₁ may be assumed injective by restricting attention to the support of P₁). Using again Lemma 4 of 6., one can write \(R_{c,2} = R_{c,1}^{\frac{1}{2}}W R_{c,1}^{-\frac{1}{2}}\), where W is a product of operators which are either unitary or of the form \((I+U)^{\frac{1}{2}}\), with U Hilbert-Schmidt and \(\sigma(U) > -1\). Unitary operators always appear in pairs involving the operator and its adjoint. Since \((I+U)^{-\frac{1}{2}}\) can always be written in the form \(I+U\) with \(\tilde{U}\) Hilbert-Schmidt, W has the form I+S. S Hilbert-Schmidt. To see that \(\sigma(S) > -1\), it is sufficient to remember that \(R_{c,1}^{\frac{1}{2}}\) and \(R_{c,2}^{\frac{1}{2}}\) have square roots with identical ranges, so that \(I+S\) is invertible. The reverse implication is proved similarly.

**Proposition 8**

**pf.** Let \(H(\tilde{s})\) be the subspace spanned by \(\tilde{s}\) and \(H(\tilde{s})^+\) be its orthogonal subspace. Let \(\pi(\tilde{s})\) denote the projection onto \(H(\tilde{s})\) and \(\pi(\tilde{s})^+\) the projection onto \(H(\tilde{s})^+\). Set \(T_Q = \theta \pi(\tilde{s})\). Then \(t = (1+\theta)^{\frac{1}{2}}\). Since \(\theta\) can be any real number strictly
larger than -1, t can take any positive value. As a function of t, \( \epsilon \{ \theta (\beta) - \frac{1}{2} \} \) is a convex function with a minimum at \( t = 1 \) and extreme values at \( t = 0 \), where its value is \( \epsilon (1 - \theta (\beta)) \). Furthermore, for \( \theta (\beta) > 3/4, \), \( \theta (\beta) - \frac{1}{2} > 1 - \theta (\beta) \).

\( \{P_k^c\} \) converges weakly to \( P \) if and only if either \( \{\epsilon_k\} \) converges to zero or \( \{Q_k\} \) converges weakly to \( P \). If the former obtains, \( \lim_k p_k^a - p_k^b = 0 \). Let \( \{a_k\} \) be a sequence of positive numbers whose sum is finite. Define \( \theta_k, l = 0 \) for \( l = 1, \ldots, k - 1 \), \( \theta_k = (1/\theta_k) - 1 \), \( \eta_k \) being defined below, and, for \( l > k \), \( \theta_{k,l} = 1 \). Let \( s_k = <s, e_k> \) and choose \( \{\theta_k\} \) such that \( \lim_k \theta_k s^2_k = \infty \). Then \( \frac{1}{k} (s^2_k/(1 + \theta_{k,l})) \geq \theta_k s^2_k \), so that \( t_k \leq (\theta_k s^2_k)^{-\frac{1}{2}} \) and \( \lim_k t_k = 0 \). Now, \( <R^c_p h, h> = <R^c P h, h> + \epsilon_k \eta_k \frac{1}{k} \lambda_k <e_k, h>^2 \).

Since the sequence \( \{\theta_k, l\} \) is uniformly bounded, \( \lim_k <R^c_p h, h> = <R^c P h, h> \).

Furthermore \( \text{trace}(R^c) \) is bounded and \( \lim \sum_k <R^c_k e_k, e_k> \) converges uniformly to \( \lim_k \sum_l p_k^c e_l \) zero as \( m \) increases to infinity. This insures that \( W \)-lim \( p_k^c = P \) \( (13) \). /

**Proposition 9**

**pf.** : Set \( s_k = \lambda_k^{-\frac{1}{2}} e_k \). Then \( ||R^{-\frac{1}{2}} p^s_k|| = 1 \), \( z = a \) and \( s_k = R^c_p s_k / ||R^c p^s_k|| = e_k \).

Let \( T^c_k = \theta_k \Pi(s_k) \). Then, as in the proof of Proposition 8, \( t_k = (1 + e_k)^{-\frac{1}{2}} \).

\( <R^c_p h, h> = <R^c P h, h> + \epsilon_k \theta_k \lambda_k \) and \( \text{trace}(R^c) = \text{trace}(R^c P) + \epsilon_k \theta_k \lambda_k \).

\( \sum_k <R^c_k e_k, e_k> \leq \sum_k <R^c_p e_k, e_k> + \epsilon_k \theta_k \lambda_k \). It thus suffices to choose \( \{\theta_k\} \) such that \( \lim_k \theta_k = \infty \) and \( \lim_k \theta_k \lambda_k = 0 \). /

**Proposition 10**

**pf.** : Fix \( g \) so that it is not orthogonal to \( s \) and \( r > -||R^c g||^2 \). Set \( \bar{g} = R^c g / ||R^c g|| \).
and \( \gamma = -\frac{1}{2} \left( \frac{|R_p^2 g|^2}{1+\gamma |R_p^2 g|^2} \right) \). Let \( \pi(g) \) be the projection onto \( H(g) \) and \( \pi^+(g) \) be that onto \( H^+(g) \). Then, if \( T = \gamma \pi^+(g) \), \((1+T)^{-1} = (1+\gamma |R_p^2 g|^2) \pi^+(g)-\pi^+(g) = 1+R_p^2 (\gamma \pi^+(g))R_p^2 \). So \( G = \gamma \pi^+(g) \) and \( K(x) = (1-\gamma) \exp \left( -\frac{1}{2} \gamma <g, x^2> \right) \).

Set \( <g, x> = y \), \( <g, s> = a \), \( -\frac{1}{2}|y|^2 = P(y) \). \( k \) can then be expressed in the form \( k(y) = \frac{1}{A+B \exp \left( P(y-a) \right)}/(A+B \exp \left( P(y) \right)) \).

If \( L(y) = \frac{1}{A+B \exp \left( P(y) \right)}/(A+B \exp \left( P(y) \right)) \), then \( \frac{d}{dy} k(y) = B k(y) \frac{L(y-a)-L(y)}{\lambda} \). Now, for \( \gamma <0 \), \( L \) is strictly increasing.

So the graphs of \( L(y-a) \) and of \( L(y) \) are parallel and do not meet (because of the assumption that \( a \) is different from zero, that is, \( g \) and \( s \) are not orthogonal).

The derivative of \( k \) has thus constant sign and \( k \) is a monotone function which is unbounded and positive.

**Proposition 11**

\[ k(0) = \frac{(\lambda+u)^{-1}(\lambda \exp \left( -\frac{1}{2} |R_p^2 s|^2 \right) + u \exp \left( -\frac{1}{2} (\frac{1}{2} |R_p^2 s|^2) \right) \exp \left( \frac{1}{2} |R_p^2 s|^2 \right)}{\lambda} \]

So, by convexity,

\[ k(0) \geq \exp \left( -\frac{1}{2} (\lambda+u)^{-1} \left( |R_p^2 s|^2 + u \left( (1+n)^{-1} |R_p^2 s|^2 \right) \right) \right) \exp \left( \frac{1}{2} |R_p^2 s|^2 \right) \]

If \( H_n \) is the subspace spanned by \( e_1, \ldots, e_n \), \( \pi \) the projection onto \( H_n \) and \( \pi^+ \) that onto \( H^+(n) \), set \( T = \kappa \pi \) and \( \kappa = 1-(1/(1+n)) \). Then \( \det \left( (1+n)^{-1} \right) = (1+n)^{-1} \left( \pi^+ \right)^2 + (1+n)^{-1} \left( \pi \right)^2 \). Taking the limit as \( n \) grows
yields $k(0) \geq \exp(\frac{1}{2} |R_P^{-\frac{1}{2}} s|^2)$. \\

Proposition 12
pf.: When $G_n$ is nonnegative, $s_n^{-1} \leq k_n(x) \leq s_n$ and thus

$$P_n^C(dP_s/dP_a > M) < P_n^C(dP_n, s/dP_n > a) \leq P_n^C(dP_s/dP_a > a).$$

Since $dP_s/dP$ is a continuous function, the result follows from the properties of weak convergence.

Proposition 13
pf.: Let $h^* = R_P^\frac{1}{2} h, ||R_P^\frac{1}{2} h||$. One then has that $E_Q \{\exp(iX(h))\} =$

$$\{1 - R_P^\frac{1}{2} A_P^\frac{1}{2} h^*, h^* < R_P h, h^*}\{\text{trace}(AR_P) + ||A\frac{1}{2} a||^2\} \exp(-\frac{1}{2}<R_P h, h\rangle).$$

So, if $W$-lim $Q_n = P$ and $k$ is a vector in the range of $R_P^\frac{1}{2}$, with norm equal to one,

$$\lim e_n = lim ||A_n^\frac{1}{2} R_P^\frac{1}{2} k||^2/\{\text{trace}(A_n R_P) + ||A_n^{\frac{1}{2}} a_n||^2\} = 0.\$$

Proposition 14
pf.: To obtain the extreme values of $k$, one studies the function $k(s^* + t x)$ as a function of $t$. It turns out that $s^*$ must be of the form $s^* = a + \lambda s + a_A$, where $a_A$ is in the kernel of $A$, for $s^*$ to be extremal. $\lambda$ is obtained using this representation.

Proposition 15
pf.: If $\lambda_m$ corresponds to $s_m^*$ and $\lambda_M$ to $s_M^*$ in the proof of Proposition 14,

$$P_n^C(D(X) > log_\lambda - log_\lambda(a + \lambda_m s + a_A)) \leq P_n^C(dP_s/dP_a > a) \leq P_n^C(D(X) > log_\lambda - log_\lambda(a + \lambda_M s + a_A)).$$

Then, if $d = 4c/<A s, s>$, $k(a + \lambda_m s + a_A) = (1 + (1 + d)^\frac{1}{2})/(1 + d)$, which increases strictly from zero to one as $d$ goes from zero to infinity. If now $W$-lim $P_n^C = P$ and $\lim sup C_n/c_n \leq = \lim <A_n s, s>/4c_n = 0$ and $\lim k(a + \lambda_n, m s + a_A) = 1.$

7. REFERENCES


(8) A.F. Gualtierotti, "On Shannon information, and channel capacity for a class of contaminated Gaussian noises," Information and Control 46, pp. 46-70, 1980.


(14) E. Parzen. "Probability density functionals and reproducing kernel Hilbert

(15) B.S. Rajput and S. Cambanis. "Gaussian processes and Gaussian measures,"

(16) C.R. Rao and V.S. Varadarajan, "Discrimination of Gaussian processes,"

NATO Advanced Study Institute: Communication systems and random process


(20) B. Simon, Trace Ideals and their Applications. Cambridge, England: Cambridge

(21) J.W. Tukey, "A survey of sampling from contaminated distributions,"

(22) J. Weidmann, Linear Operators in Hilbert Spaces. New York, NY: Springer
Verlag, 1980.