On Average Mutual Information and Capacity for a Channel without Feedback and Contaminated Gaussian Noise

A. F. Gualtierotti

Département de mathématiques-EPFL, 61, av. de Cour, CH-1007 Lausanne, Switzerland

We investigate the variation of Shannon information and channel capacity, when the statistics go through a family of contaminated-Gaussian laws, called QN-laws. There are results which allow one to decide when a joint QN-law has margins belonging to the same class, formulas for Shannon information in the latter case, and also for a channel adding independent noise, with statistics belonging to the QN-law class. Finally, there is a study of the structure of the channel capacity problem in the latter case.

1. INTRODUCTION

Normality assumptions are common in communication theory, and the solution to a given problem appears usually as a set of relations which involve the parameters of the Gaussian elements listed in the model. Thus the detection of a sure signal, corrupted by additive Gaussian noise, is non-singular when and only when this signal belongs to the range of the square root of an operator, which is obtained from the covariance of the noise (Grenander, 1950).

Many communication models involve infinite dimensional spaces, and, on these, measures are very sensitive to small changes in the values of the parameters which characterize them. Thus, if $R$ is a self-adjoint, positive operator on $L_2[0, T]$ with finite trace, the Gaussian law it determines is orthogonal to that corresponding to the operator $(1 + \epsilon)R$, however small $\epsilon$ may be (Rao and Varadarajan, 1963). This is in stark contrast to the finite dimensional case, where the normal densities corresponding to the matrices $\Sigma$ and $c\Sigma$ are always equivalent, for $c > 0$. Some problems in communication theory, among which are detection problems, and the computation of channel capacity, require that probability measures be equivalent. Since the assumption of normality is often an approximation of the real law which governs the system under consideration, and even sometimes an analytic convenience, it is necessary to investigate how a given set of relations, derived from a Gaussian model, behaves when these Gaussian laws undergo a perturbation. This necessity had been recognized a long time ago (Root, 1964).

In statistics, the insensitivity of a procedure or of a model against small deviations from the assumptions has been checked by letting the under-
ly ing distribution in the model belong to a family of laws which are contamina-
tions of the law in the model to be tested (Tukey, 1960; Huber, 1977). There are
several reasons which make contaminations useful and interesting. These are,
in the words of Huber (1977): "typical "good data" samples in the physical sci-
cences appear to be well modeled by an error law of the form \( F(x) = (1 - \epsilon)\Phi(x) + \epsilon\Phi(x/3) \), where \( \Phi \) is the standard normal cumulative, with \( \epsilon \) in the range between 0.01 and 0.1 (This does not necessarily imply that these samples contain between
1% and 10% gross errors, although this is often true—the above may just be a
convenient description of a slightly longer-tailed than normal distribution)."

When the model is infinite dimensional, any attempt to define a notion of
contamination meets with the following problem. Suppose \( P \) is a zero-mean,
Gaussian law on the Borel sets of \( L_2[0, T] \), representing the noise in some system.
Let \( Q \) be another zero-mean, Gaussian probability law on the same Borel sets.
A contamination of \( P \) can then be written: \( P_\epsilon = (1 - \epsilon)P + \epsilon Q, \epsilon \in [0, 1]. \) Now
\( P_\epsilon \) and \( P \) are equivalent if and only if \( P \) and \( Q \) are equivalent. Indeed, if \( P_\epsilon \) and \( P \)
are equivalent, \( P(A) = 0 \) implies \( P_\epsilon(A) = 0 \), so that \( Q(A) = 0. \) Thus \( Q \) is abso-
lutely continuous with respect to \( P \), which is sufficient to ensure that \( P \) and \( Q \) are
equivalent (Feldman, 1958; Hájek, 1958). Conversely, if \( P \) and \( Q \) are equivalent,
\( P \) and \( P_\epsilon \) are also equivalent. If now \( R \) and \( S \) are the respective covariance
operators of \( P \) and \( Q \), \( P \) and \( Q \) are equivalent if and only if \( S = R^{1/2}(I + TR)^{1/2} \),
where \( T \) belongs to the Hilbert–Schmidt class, and the eigenvalues of \( T \) strictly exceed \(-1 \) (Rao and Varadarajan, 1963). Thus, it can in principle be distin-
guished without error whether \( P \) or \( P_\epsilon \) obtains, as soon as \( S \) does not have the
representation in terms of \( R \) just given. One should then produce laws \( P_\epsilon \) which
are not only contaminations of \( P \), but also laws which are equivalent to \( P \). There
is an added technical restriction: the law \( P_\epsilon \) must be sufficiently simple so that
one can compute the likelihood \( dP_\epsilon/dP \), for one cannot usually assume, without
rather stringent hypotheses, which are, from a practical point of view, undesirable,
that the likelihood is, say, the exponential of a quadratic form (Rao and

We shall not consider contamination by a Gaussian law, because, though the
methods required are similar to those we shall present, another set of calculations
would be necessary. Furthermore, it is often required that the contaminating
law belong to a broadly defined class, for which information of the type required
to ascertain equivalence would not be available. Gualtierotti (1979a) introduced
a family of laws, called \( QN \)-laws, which are equivalent to Gaussian laws, are
contaminated-Gaussian, and in certain aspects qualitatively different from
Gaussian laws. \( QN \)-laws are defined by a relation \( dQ = q \, dP \), where \( P \) is Gaussian,
and \( q \) is a quadratic form. The density of the continuous linear functionals has
the form:

\[
(1 - c^2) \frac{1}{a(2\pi)^{1/2}} \exp \left( -\frac{1}{2} \frac{x^2}{a^2} \right) + c^2 \frac{1}{a(2\pi)^{1/2}} \frac{x^2}{a^2} \exp \left( -\frac{1}{2} \frac{x^2}{a^2} \right), \quad 0 \leq c^2 \leq 1.
\]
For appropriate values of $c^2$, the density will have only one peak, at the origin, and heavier tails than the Gaussian density with mean zero, and variance $a^2$. Another feature of $Q\!N$-laws is that their independence properties are quite different from those of the Gaussian laws. Indeed, the density of the form $x^2/(2\pi)^{1/2} \exp\left(-\frac{1}{2}x^2\right)$ is not decomposable (Lukacs, 1970). The Gaussian character of many communication models is justified by an appeal to the central limit theorem. But the independence properties of the approximating laws are usually quite different from those of the limit law, as can be seen most clearly in Donsker's theorem (Donsker, 1951), where the process approximating the Wiener process has polygonal paths which do not have independent increments. $Q\!N$-laws thus provide a means to check robustness with respect to lack of independence. A final characteristic of $Q\!N$-laws is that, with their use, it is possible to allow as much, or as little Gaussianness into the model as one may wish. Indeed, the constant $c^2$ in the formula for the density is proportional to the norm of $A^{1/2}Rh$, which is zero, whenever $Rh$ is in the kernel of $A^{1/2}$. It is thus possible to have a measure $Q$ which is not Gaussian in only one direction (choosing $A$ to be $a \otimes a$), or a measure $Q$ which is Gaussian in no direction.

$Q\!N$-laws can be generated from stochastic processes as Gaussian laws do (Gualtierotti, 1979b). For example, if $P$ is the Gaussian law induced by a stationary Gaussian process, $Q$ is the law of a harmonizable process. This paper has three closely related aims. The first is simply to find out what may happen when it cannot be assumed that the noise is Gaussian. There is indeed little information on the subject, and the methods which work in the Gaussian case are too specific to that case to provide much information on related non-Gaussian problems. In fact, one of the reasons for studying models in which the noise is a $Q\!N$-law was to understand better the Gaussian case. It can thus be shown that the calculation of channel capacity is an optimization problem where one has to determine the extremum of a semi-continuous function over a convex, relatively compact set of measures. Finally, as recommended in Hogg (1979), it is sound practice to assess the reliability of a model by its response to contamination. One can thus show that for a model in which the noise is a $Q\!N$-law, but is Gaussian except for one direction, the Gaussian laws have to be excluded from the calculation of channel capacity.

The paper has four parts. In the first, $Q\!N$-laws are defined, and the properties of these which will be used in the sequel are stated. The second part is necessary for the calculation of mutual information: relations between joint $Q\!N$-laws and their margins are given. Some examples which illustrate these relations follow. In the third part there is a formula for mutual information of a $Q\!N$-law with $Q\!N$-laws as margins. The fourth part is a study of mutual information for the independent channel without feedback, and a noise which is a $Q\!N$-law, and the related channel capacity problem.

There are many ways to model the problems considered in this paper. The one we have adopted is discussed at length in Baker (1978), where references
may be found, as well as the terms which appear in this paper without definition.

2. A Family of Laws

In this section, we shall define the family of laws mentioned in Section 1, and state the properties of its elements which we shall use. Further details may be found in Gualtierotti (1979a). We shall also describe the corresponding notation.

Since signals have finite energy, the basic space one works with is that of square integrable functions over a finite interval. But it is notationally simpler to write $H$ for that space, and thus one may as well start with a real and separable Hilbert space $H$, with inner product $\langle \cdot, \cdot \rangle$, and norm $n(\cdot)$. The field of events on $H$ will be the Borel sets $\mathcal{B}(H)$. On $\mathcal{B}(H)$, we shall first consider Gaussian measures, always denoted $P$, with characteristics $m$ and $R$. $X(h)$ denotes the random variable $\langle \cdot, h \rangle$, and expectation with respect to $P$ is written $E_P$. Then the mean $m$ is that element of $H$ for which $E_P X(h) = \langle m, h \rangle$, all $h$ in $H$. The covariance $R$ is the operator on $H$ defined by $E_P (X(h) - \langle m, h \rangle)(X(k) - \langle m, k \rangle) = \langle Rh, k \rangle$. $R$ is linear, self-adjoint, non-negative, and has finite trace.

We shall often write $P \sim N(m, R)$.

For each $P$, we shall consider neighbouring measures defined as follows. $a$ will denote a real number, $A$ fixed element of $H$, and $A$ a linear operator on $H$ which is bounded, self-adjoint, and nonnegative. $r(A)$ denotes the square root of $A$, whether $A$ is a number or an operator. The constant $c_A$ is given by the relation $c_A^{-1} = a^2 + \text{trace}(AR) + n^2(r(A)(m - a))$. Further, let $q(x)$ be $c_A(a^2 + n^2 r(A)(x - a))$. Then $E_P q = 1$, so that the relation $dQ = q \ dP$ defines a probability measure. For such a measure, we shall always write $Q \sim QN(q, P)$, or, when it will be necessary to specify the parameters, $(q, P), (m, R))$. It is always possible to assume that $a$ is either zero or one, and we shall do so.

If $Q \sim QN((a, A), (m, R))$, $b := 2c_A r(A)(m - a)$, and $b \otimes b(x) := (x, b)b$, one has

1. $P$ and $Q$ are mutually absolutely continuous and $dQ/dP = q$.

2. The expectation $m_Q$ of $Q$ is $m + r(R)b$, and its covariance operator $S$ is $r(R)(I + 2c_A r(R)Ar(R) - b \otimes b)r(R)$. Furthermore no eigenvalue of $2c_A r(R)Ar(R) - b \otimes b$ equals $-1$.

3. The characteristic function $FT(Q) = FT(P)$, where $FT(Q)$ is the characteristic function of $P$. 

$$(1 + i(h, r(R)b - c_A(h, RARh)) FT(P),$$
(4) The density of $X(h)$ is given by, letting $A_o := 2\sigma_o A$,

$$f(x) = \left(1 - \frac{(h, RA_o Rh)}{2(h, Rh)}\right) + \frac{(h, RA_o(m - a))}{2(h, Rh)} (x - (m, h)) + \frac{(h, RA_o Rh)}{2(h, Rh)^2} \cdot (x - (m, h))^2 \cdot (r(2\Pi(h, Rh))^{-1} \exp \left(-\frac{(x - (m, h))^2}{2(h, Rh)}\right).$$

Remark. The measure $Q$ is Gaussian if and only if the range of $R$ is contained in the kernel of $A$, that is, if the support of $P$ is contained in the kernel of $A$ (Itô, 1970). We shall always suppose that there are points in the support of $P$ which lie outside the kernel of $A$. The construction of the measure $Q$ was motivated by a theorem of Girsanov (1960) which has been very useful in many areas of communication theory (Lipster and Shiryayev, 1977).

3. Transformations and Margins of QN-Laws

If $X$ is the input message into a channel and $Y$ the output, it is useful to suppose $Y = TX + N$, where $T$ is a transformation of $X$ and $N$ the channel noise (Baker, 1978). The first result in this section gives conditions which guarantee that $Y$ has a QN-law, when the joint law of $TX$ and $2V$ is a QN'-law.

Here $H := H_0 \times H_0$, where $H_0$ is a real and separable Hilbert space, and $H$ has the usual Hilbert space structure given by the inner product $(x, y) := (x_1, y_1)_0 + (x_2, y_2)_0$, where $x := (x_1, x_2), y := (y_1, y_2)$, and $(\cdot, \cdot)_0$ is the inner product of $H_0$. $J_i$ will be the map $x \mapsto x_i, i = 1, 2$. Set $J := J_1 + J_2$ and $R_j := JR_j^*$. On $H$ we shall consider measures $P \div N(m := (m_1, m_2), R)$ and $Q \div QN(a, a := (a_1, a_2), R), (m, R)$.

Let $K$ be the closure of the range of $r(R) J^*$ and $L$ that of the range of $r(R_1)$. The polar decomposition (Reed and Simon, 1972, Theorem VI.10, p. 197) says there exists a unique partial isometry $U_{K,L} : H_0 \to H$ such that the initial set of $U_{K,L}$ is $L$, the final set $K$, and $r(R) J^* = U_{K,L} r(R_1)$. 

PROPOSITION 3.1. $Q_j := Q \cdot J^{-1}$ is a QN-law if and only if

(5) $r(A_o) r(R) U_{K,L} (r(R_1))^{-1}$ has an extension $B$ to $L$, which is linear, and bounded, and

(6) there is an element $d$ in $L$ such that $r(A_o)(m - a) - Bd$ is orthogonal to the range of $r(A_o) r(R) U_{K,L}$.

When (5) and (6) hold, and $B$ is extended to $H_0$ by setting $Bh := 0$, when $h$ is in $L^*$, $dQ_j = q_j dP_j$, where $P_j := P \cdot J^{-1}$, and $q_j$ is obtained from

(7) $(0, Jm - d, B^*B)$, when $a = 0$, $r(A_o)(m - a) = Bd$, and range $(r(R) (A)) \subset K$;

(8) $(1, Jm - d, (2 - \text{trace}(r(R_1) B^*Br(R_1)) - n^2(Bd))^{-1}B^*B)$, otherwise.
Proof. Let $FT(Q_j)$ be the characteristic function of $Q_j$ and $FT(P_j)$ that of $P_j$. The change of variables formula yields $FT(Q_j)(h) = FT(Q)(J^*h)$ and $FT(P_j)(h) = FT(P)(J^*h)$. Thus, from (3), one has

$$FT(Q_j)(h) = (1 + i(J^*h, RA_0(m - a)) - \frac{1}{2}(J^*h, RA_0RJ^*h) \cdot FT(P_j)(h).$$

Also, (5) is equivalent to (10) below:

(10) There exists $B: L \rightarrow H$, linear and bounded, such that

$$U_{K,L}^*_r(R) A_0 r(R) U_{K,L} = r(R_j) B^*Br(R_j).$$

Indeed, that (10) follows from (5) is obvious and, when (10) holds, the range of $U_{K,L}^*_r(R)r(A_0)$ is contained in that of $r(R_j)$ (Baker, 1973a, Corollary, (a), p. 179). But on $L$, $r(R_j)$ is an injection and its range is dense in $L$. Thus

$$\text{trace}(r(A_0)r(R) U_{K,L} r(R_j))^{-1}(r(R_j)x) \leq c_0(r(R_j)x),$$

and (1) holds.

Furthermore, when (5) holds, and if $d$ is a point in $L$, (6) is equivalent to (11) below:

$$B^*r(A_0)(m - a) = B^*Bd.$$  

Indeed, the closure of the range of $r(A_0)r(R) U_{K,L}$ is the same as that of $B$. Then (Bachman and Narici, 1966, (20.37), p. 363)

$$\ker(B^*) = (\text{closure}(\text{range}(B)))^\perp = (\text{closure}(\text{range}(r(A_0)r(R) U_{K,L})))^\perp.$$  

Thus, given hypotheses (5) and (6) and equivalent conditions (10) and (11), one can write (9) as

$$FT(Q_j)(h) = (1 + i(h, R_jB^*Bd)_0 - \frac{1}{2}(h, R_jB^*BR_jh)_0) \cdot \exp(i(Jm, h)_0 - \frac{1}{2}(Rsh, h)_0).$$

Finally, when (5) or, equivalently, (10) and (6) or, equivalently, (11) hold, then

(13) if $a = 0$, $r(A_0)(m - a) = Bd$, and range$(r(R)r(A_)) \subseteq K$, one has that

$$\text{trace}(r(R_j) B^*Br(R_j)) + n^a(Bd) = 2;$$

(14) otherwise, $\text{trace}(r(R_j) B^*Br(R_j)) + n^a(Bd) < 2.$

Indeed, (10) gives

$$\text{trace}(r(R_j) B^*Br(R_j)) = \text{trace}(U_{K,L}^*_r(R) A_0 r(R) U_{K,L})$$

and (11) gives $B^*Bd = B^*r(A_0)(m - a)$, and $n(Bd) \leq n(r(A_0)(m - a))$ with equality if and only if $Bd = r(A_0)(m - a)$. Furthermore, $\text{trace}(U_{K,L}^*_r(R) Br(R) U_{K,L}) \leq \text{trace}(AR)$ with equality if and only if range$(r(R)r(A)) \subseteq K$. Thus

$$\text{trace}(r(R_j) B^*Br(R_j)) \leq 2c_0(\text{trace}(U_{K,L}^*_r(R) Br(R) U_{K,L}) + n^a(r(A)(m - a))).$$

(15)
Under condition (13), the left-hand side of (15) is equal to $2c_0(\text{trace}(A) + n^2(\text{trace}(A)(m - a))) = 2$ and otherwise, the second factor on the right-hand side of (15) is strictly less than one.

A direct calculation, using (13) and (14), shows that (12) is the characteristic function of the laws corresponding to (7) and (8).

Remark 3.1. For computations, as we shall see in the examples, an alternate formulation of (5) and (10) is useful: if $D_H$ is the diagonal in $H$, and $K_D$ the closure of $RD_H$ in $H$, if $\Pi_D$ is the projection with range $K_D$, then (5) and (10) are equivalent to

$$r(A_0)\Pi_D = B\Pi_D,$$

for some $B$ which is linear and bounded. (16)

The latter relation shows also that the existence of an extension depends on the behaviour of $r(A)R$ on the diagonal $D_H$. For example, if $RD_H \subseteq \ker(A)$, $B = 0$ will do. The resulting measure is then Gaussian.

To evaluate the Shannon information of a $QN$-law, it is necessary to know what the margins of that law are. The next proposition states under which conditions the margins of a $QN$-law are themselves $QN$-laws. Proofs shall be skipped, since they are essentially the same as those involved in Proposition 1.

The notation shall be as follows. $H_1$ and $H_2$ are two real and separable Hilbert spaces and $H = H_1 \times H_2$. Again $Q \div QN(q, P)$. Define then $P_i = P \cdot J_i^{-1}$, $Q_i = Q \cdot J_i^{-1}$, $i = 1, 2$. $P_i$ is Gaussian, with mean $m_i$, and covariance $R_i$. The polar decomposition yields $r(R) J_i^* = V_i r(R_i)$, where $V_i$ is the unique partial isometry with initial set $K_i = \text{closure}(\text{range}(r(R_i)))$, and final set $L_i = \text{closure}(\text{range}(r(R) J_i^*))$. One then has

**Proposition 3.2.** $Q_i \div QN(q_i, P_i)$ if, and only if

(17) $r(A_0)r(R) V_i(r(R_i))^{-1}$ has an extension $B_i$ to $K_i$, which is linear and bounded, and

(18) there exists an element $d_i$ in $K_i$ such that $r(A_0)(m - a) - B_id_i$ is orthogonal to the range of $r(A_0)r(R) V_i$.

When (17) and (18) hold, and $B_i$ is extended to $H_i$ by setting $B_i h_i = 0$, when $h_i$ is in $K_i^\perp$, $q_i$ is obtained from

(19) $(0, m_i - d_i, B_i\cdot B_i^*)$, when $a = 0, r(A_0)(m - a) = B_id_i$, and range $(r(R)r(A)) \subseteq L_i$;

(20) $(1, m_i - d_i, (2 - \text{trace}(r(R_i) B_i^* B_i r(R_i)) - n^2(B_id_i)^{-1} B_i^* B_i), otherwise.

Remark 3.2. Here again $Q_i$ is Gaussian when range($RJ_i^*$) $\subseteq \ker(A)$. Furthermore, the statement analogous to that of Remark 3.1 is $r(A_0) \prod_{K(i)} = B_i J_i \prod_{K(i)}$ where $K(i)$ is the closure in $H$ of $RJ_i^* H_i$. 
Remark 3.3. It is sometimes convenient to have relation (17) in terms of operators on $H_1$. This can be done using some relations to be found in Gualtierotti (1979c). One can then see in particular that the existence of an extension depends on the behaviour of the cross covariance operator obtained from $R$.

Examples

Propositions 3.1 and 3.2 contain restrictions expressed in analytic terms. By considering below the simplest cases: $A = k \otimes k$ and $R$ diagonal, we shall see that these restrictions fix the relative positions of certain elements and subspaces, which are related to the parameters of the $QN$-laws involved. In a sense the quantity of information measures distances and angles (Gualtierotti, 1979c). But first we shall give an example showing how $QN$-laws can be obtained from stochastic processes.

1. Construction of a $QN$-law from a stochastic process. Let $P$ be the law induced on $L_2[-T, T]$ by a zero-mean Gaussian process with covariance $R(t) = \exp(-\text{abs}(t))$, where $\text{abs}(x)$ denotes the absolute value of $x$. Let $f_n$ be an element from an orthonormal sequence in $H(R)$, the reproducing kernel Hilbert space associated with $R$. Finally let $A(s, t)$ be the series $\sum a_n f_n(s) f_n(t)$, where the $a_n$'s are strictly positive, and their sum is less than one half. One can then construct a stochastic process $X$ with covariance $R \circ A$, and for which $\int_{-T}^{T} X(t) dt$ has characteristic function (Gualtierotti, 1979b): $2(1/2 - t^2(Ah, h)_{L_2(T)}) \cdot \exp(-(t^2/2)(Rh, h)_{L_2(T)})$, where $L_2(T)$ is the space of square integrable functions over the interval from $-T$ to $T$ (with respect to Lebesgue measure), $A$ and $R$ the operators on $L_2(T)$ with respective kernels $A$ and $R$. If $F$ is the spectral measure associated with $R$, and if $h(t) := (\exp(it \cdot), h)_{L_2(T)}$, then $(Rh, h)_{L_2(T)} = (h, h)_{L_2(F)}$, $(Ah, h)_{L_2(T)} = (U^* J_A U h, h)_{L_2(F)}$, $J_A$ is the operator on $H(R)$ with eigenvalues $a_n$ and eigenvectors $f_n$, $U$ the unitary map which identifies $L_2(F)$, the closure in $L_2(F)$ of the linear span of the functions $\exp(it \cdot)$, $t$ between $-T$ and $T$, with $H(R)$. Finally,

$$n_{L_2(T)}^2(Ah) \leq T(Ah, h)_{L_2(T)} \leq Tn(J_A)(Rh, h)_{L_2(T)} \leq \frac{Tn(J_A)}{\frac{1}{2} - \frac{1}{2} \exp(-2T)} n_{L_2(T)}^2(Rh),$$

which means (Douglas, 1966) that $A = RCR$, where $C$ is linear and bounded. One then sees, with the help of (3) that the process $X$ induces a $QN$-law with quadratic form proportional to $C$.

2. The case $A = k \otimes k$, $R$ arbitrary. To check the conditions of Proposition 3.1, one can use (6) and (16). $B$ has to be of the form $k \otimes b_0$, $b_0$ in $H_0$, and $r(2\pi \sigma) n^{-1}(k)k - (b_0, b_0)$ has to be in the kernel of $\Pi_D$. If $k$ is in the kernel of $\Pi_D$, $B = 0$ will do, otherwise, the sum of the components of $\Pi_D k$ has to be
different from zero. This takes care of (6). Equation (16) becomes:
\[(r(2c_{\theta})^{-1}(k) \times (m - \mathbf{a}, k) - (d, b_0) \cdot (U_{K, L}^T r(R) k, x)_0 = 0.\] If \(k\) is in the kernel of \(U_{K, L}^T r(R)\), no problem arises. Otherwise, if \(b_0 \neq 0\), \(d\) can always be appropriately chosen. Similar restrictions apply to Proposition 3.2.

3. The case \(R\) diagonal, \(A\) arbitrary. When \(P = P_1 \otimes P_2\), \(R\) is a diagonal matrix with elements \(R_1\) and \(R_2\), and \(r(R)\) is also diagonal, with \(r(R_1)\) and \(r(R_2)\) as diagonal elements. Thus \(r(R) J^* x = (r(R_1)x, r(R_2)x)\), and \(R_j = R_1 + R_2\). The operator \(r(R_1)(r(R + R_2))^{-1}\) has an extension \(W_1\) to \(L\), which is linear and bounded. Letting \(W_1\) be zero on \(L^\perp\), \(W_1\) is defined on \(H_0\). \(W_2\) can be defined similarly. If \(W = (W_1, W_2)\), \(W\) is a partial isometry, with initial set \(L\), final set \(K\), and \(r(R) J^* = Wr(R_j)\). Thus \(W = U_{K, L}\). If now \(R_1\) and \(R_2\) commute, so that their composition is a positive operator (Bachman and Narici, 1966, Theorem 22.8, p. 415), then \(R_j^{-1}\) has a bounded linear extension \(W_1\) to \(K\). \(W_2\) is defined similarly, and \(W'\) is \((W_1', W_2')\). Then \(W' r(R_j) = r(R) U_{K, L}\), and \(B\) can be chosen to be \(r(A_0) W'\). Hypothesis (5) is thus checked. Hypothesis (6) becomes \(m - a - W'd\) is in \(\ker(W^* r(R) r(A_0))\). This will be the case for example when \(A\) is an injection, and \(m - a\) is in the closure of the range of \(R\).

If \(R\) is an injection, \(K_1 = H_1\), and \(r(R) J_j = J^* r(R_1)\), so that \(B_1 = r(A_0) J^*_j\) solves (17). Relation (18) becomes \(m - a - (d_1, O_2)\) is in \(\ker(J r(R) r(A_0))\).

4. The case \(A = k \otimes k, R\) diagonal. Suppose \(a = m = 0\), \(R\) injective, \(k_i \neq 0\), \(q(x) = (R k, k)^{-1}(k, x)^2\), \(n(k) = 1\). \(B_i\) can then be chosen to be \(r(2(R k, k)^{-1} k \otimes k_i)\). The range of \(r(R) r(A)\) is not contained in \(L_i\), and \(\text{trace}(r(R_j) B_i^* B_j r(R_j)) = 2(R k, k)^{-1}(R k_i, k_i)\), so that \(Q_i = QN((1, 0, (R k, k)^{-1} k_1 \otimes k_1), (0_1, R_1))\).

**PROPOSITION 3.3.** Suppose, for \(i = 1, 2, Q_i = QN(q_i, P_i) = QN((a_i, a_i, A_i), m_i, P_i)). Set \(a = (a_1, a_2), \ m = (m_1, m_2), P = P_1 \otimes P_2, Q = Q_1 \otimes Q_2, \) and \(R := \text{the diagonal matrix with diagonal elements } R_1\) and \(R_2\). Then the range of \(Q_i\) is a \(QN\)-law if and only if at least one of the indices 1 and 2, say \(i\), \(\text{range}(R_i) \subseteq \ker(A_i)\). \(Q_i\) is then Gaussian, and \(A_i\) can be chosen to be the null operator. If \(j\) is the index in \((1, 2)\) which is different from \(i\), and \(\text{range}(R_j) \subseteq \ker(A_j)\), \(Q = P\). Otherwise \(Q = QN(q_j, J_j, P)\).

**Proof.** Suppose first that \(Q\) is a \(QN\)-law, that is, \(Q = QN(q', P') = QN((a', a', A'), (m', R'))\). One has that, for every Borel set \(B\), \(Q(B) = \int_B q_B (q, q') \cdot J_B (p, p') (a, a', A', (m', R'))\). The relation is indeed true for every rectangle. If \(P'\) is orthogonal to \(P\) and \(P'(B) = P(B') = 0\), then \(P'(B') = 0\) also, for \(Q\) and \(P'\) are mutually absolutely continuous (1). Thus \(P' = 0\), which is impossible. Consequently (Rao and Varadarajan, 1963, Theorem 4.1, p. 308, and Theorem 5.1, p. 312),

\[P' = P \quad \text{and} \quad R' = r(R)(I + T)r(R). \]
From (21), the equality
\[ \text{abs}^2(FT(Q_1)(h_1) \cdot FT(Q_2)(h_2)) = \text{abs}^2(FT(Q)(h)) \]
and (3) one has
\[
(1 - \frac{1}{2}(h, R'A_0 R'h))^2 - (h, R'A_0(m' - a'))^2 \cdot \exp(-\frac{1}{2}(R'TR'h, h)) \]
\[ = ((1 - \frac{1}{2}(h_1, R_1 A_0 R_1 h_1))^2 - (h_1, R_1 A_0(m_1 - a_1))^2) \]
\[ \cdot ((1 - \frac{1}{2}(h_2, R_2 A_0 R_2 h_2))^2 - (h_2, R_2 A_0(m_2 - a_2))^2). \]  
(22)

Setting \( h = tk \), one sees that \( T \) must be "zero." Thus (22) is an equality between two polynomials in \( t \): on the left-hand side the degree is four, and on the right-hand side it is eight. Thus the coefficients of the terms of degree exceeding four must be zero. In particular, one has \( (h_1, R_1 A_0 R_1 h_1) \cdot (h_2, R_2 A_0 h_2) = 0 \), which means that \( \ker(R_1 A_0 R_1) \times H_2 \cup (H_1 \times \ker(R_2 A_0 R_2)) = H \), or that one of the sets in the union is \( H \). Suppose then that \( \ker(R_1 A_0 R_1) = H_1 \). One then has
\[
FT(Q)(h) = (1 + i(h, J^*_2 R_2 A_0 J_2(m - a)) - \frac{1}{2}(h, J^*_2 R_2 A_0 R_2 h)) \cdot \exp(i(m, h) - \frac{1}{2}((J^* R_1 J_1 + J^*_2 R_2 J_2)h, h)).
\]
But \( J^*_1 R_1 J_1 + J^*_2 R_2 J_2 = R \), and \( R_2 = J^*_2 R_2 J_2 \) (Gualtierotti, 1979c), so that
\[
FT(Q)(h) = FT(P)(h) \cdot (1 + i(h, J^*_2 R_2 A_0 J_2(m - a)) - \frac{1}{2}(h, J^*_2 R_2 A_0 R_2 h)).
\]

\( FT(Q) \) thus has the required form.

Conversely, if, for example, \( FT(Q_1)(h_1) = FT(P_1)(h_1) \), then (3)
\[
FT(Q)(h) = (1 + i(h, J^*_2 R_2 A_0 J_2(m - a)) - \frac{1}{2}(h, J^*_2 R_2 A_0 R_2 h)) \cdot FT(P)(h).
\]
Defining \( A' \) as above, one sees that \( Q \) is a QN-law.

4. The Average Mutual Information of a QN-Law with QN-Laws as Margins

In this section, it is assumed that \( Q \div QN(q, P) \), and that \( Q_i := Q \div J_i^{-1} \div QN(q_i, P_i), i = 1, 2 \). Then \( P \) and \( Q_1, P_i, Q_2 \), and \( Q_i, i = 1, 2 \), are mutually absolutely continuous (1). Consequently \( P_1 \otimes P_2 \) and \( Q_1 \otimes Q_2 \) are mutually absolutely continuous, and \( d(Q_1 \otimes Q_2)/d(P_1 \otimes P_2) = q_1 \cdot q_2 \) (Hewitt, and Stromberg, 1965, (21.29) Theorem, p. 394). Furthermore, \( P \) and \( P_1 \otimes P_2 \) are mutually absolutely continuous if and only if \( Q \) and \( Q_1 \otimes Q_2 \) are. We shall write \( P_\odot \) for \( P_1 \otimes P_2 \), \( Q_\odot \) for \( Q_1 \otimes Q_2 \), \( P_\odot \) for \( dP/dP_\odot \), and \( Q_\odot \) for \( dQ/dQ_\odot \).

We shall assume that \( Q \) and \( Q_\odot \) are mutually absolutely continuous (necessary and sufficient conditions are stated in Gualtierotti (1979a) and are the same as
those for equivalence in the Gaussian case). We shall obtain a formula for
$I(Q) := E_\theta \log(q_\theta)$. Now \( q_\theta = (q_1 \cdot q_2)^{-1} \cdot p_\theta \) (Hewitt, and Stromberg, 1965, (19.44) (Chain rule), p. 328), \( p_\theta \) is known, and obtained as follows. If \( P \div N(m, R) \), \( P_\theta \div N(m, R_\theta) \), where \( R_\theta \) is the diagonal operator with diagonal elements equal, respectively, to \( R_1 \) and \( R_2 \), where \( R_i = \sum_j R_j \delta_{ij} \), \( i = 1, 2 \). Since, by hypothesis, \( P \) and \( P_\theta \) are mutually absolutely continuous, \( R = r(R_\theta) \times (I + T)r(R_\theta), n(T) < 1 \), and \( T \) is Hilbert–Schmidt (Baker, 1973a, Theorem 6, p. 287). Let \( t: H \to H \) be defined by \( \mathbf{x} = \mathbf{x} - m \). Then \( P \cdot t^{-1} \) and \( P_\theta \cdot t^{-1} \) are mutually absolutely continuous Gaussian measures, with mean zero, and respective covariances \( R \) and \( R_\theta \). Their Radon–Nikodym derivative \( p_\theta \) is given by \( p_\theta(\mathbf{x} + m) \). Let \( r_n \) and \( e_n \) be, respectively, the positive eigenvalues and the corresponding eigenvectors of \( R_\theta \). Set \( f_n = (r(r_n))^{-1} e_n \) and \( X_n(x) = (\mathbf{x}, e_n) \). Also, let \( t_n \) and \( t_n \) be, respectively, the eigenvalues and corresponding eigenvectors of \( T \). Set \( Y_p = \sum_n (t_p, e_n) X_n \). Then, with respect to \( P_\theta \), \( (Y_p, P) \) is a family of independent, \( N(0, 1) \)-random variables, and, with respect to \( P \), a family of independent, \( N(0, 1 + t_n) \)-random variables. Furthermore, if \( L_n P_n(x) = -\frac{1}{2} t_n(1 + t_n)^{-1} - 1 \cdot Y_p^2 - \log(1 + t_n)) \), \( p_\theta(x) = \lim L_n P_n(x) \) (Rao and Varadarajan, 1963, p. 318). Finally, the eigenvalues \( t_n \) come in pairs which distinguish themselves from each other only by sign (Baker, 1973a, Theorem 3(b), p. 281). We shall then write, when necessary, \( t_n^+, t_n^-, t_n^+, t_n^- \), with the convention that \( t_n^- = -t_n^+ \), \( t_n^+ \geq 0 \). The corresponding \( Y_p \)'s will be written \( Y_p^+ = \sum_n (t_p^+, e_n) X_n \), and we shall also use the notation \( Y_p^+ = (r(1 + t_n^+))^{-1} r(t_n^+) Y_p^+ \).

Finally, it is no restriction to suppose that \( m = 0 \), for the quantity \( I(Q) \) is invariant under translations (Baker, 1978, Corollary, p. 77).

**Lemma 4.1.** If \( \Pi_\theta \) is the projection onto the closure of the range of \( R_\theta \),

\[
E_\theta \log(p_\theta) = I(P) + c_\theta \sum_{i=1}^n t_i(1 + t_i)^{-1} n^2(r(A)r_\theta)(I + T)\Pi_\theta t_i.
\]

**Proof.** \( \log(p_\theta) \) is in \( L_2(P) \) (Baker, 1978, Proposition 2, pp. 77–79). Since \( n^2(r(A)(x - a)) \) is also in \( L_2(P) \), \( E_\theta \log(p_\theta) n^2(r(A)(x - a)) \) exists, and is finite, so that \( E_\theta \log(p_\theta) \) exists, and is finite. Then, \( E_\theta \log(p_\theta) = c_A a^2 + n^2(r(A)a)I(P) + 2c_A E_\theta \log(p_\theta)(\Lambda a, \cdot) + c_A E_\theta \log(p_\theta)(\Lambda a, \cdot) \). \( \log(p_\theta) \) is the limit in \( L_2(P) \) of the sequence \( L_n \) (Baker, 1978, Proposition 2, pp. 77–79), and \( (\Lambda a, \cdot) \) is in \( L_2(P) \). Thus, since \( L_n (\Lambda a, \cdot) \) involves only odd powers of zero-mean normal random variables, \( E_\theta \log(p_\theta)(\Lambda a, \cdot) = \lim E_\theta L_n (\Lambda a, \cdot) = 0 \). Consequently one only need an expression for \( E_\theta \log(p_\theta)(A, \cdot) = \lim E_\theta L_n (A, \cdot) \times n^2(r(A)) \). But \( L_n \) can be written as:

\[
L_n(x) = \frac{1}{2} \sum_{i=1}^n (Y_{i}^+)^2 - (Y_{i}^-)^2 - \frac{1}{2} \sum_{i=1}^n \log(1 - t_i^+)^2.
\]

(Baker, 1978, Proposition 2, pp. 77–79), and one has thus to compute integrals of the type \( E_\theta (Y_{i}^+)^2 n^2(r(A)) = \sum_{i=1}^n E_\theta (Y_{i}^+)^2 (r(A)\mathbf{h}_i, \cdot)^2 \), where
(h_j, j) is a complete, orthonormal set in H. Let \( U_j(x) := (r(A)h_j, x) \). Then,

\[
E_p \exp(i(uY_j^+ + vU_k)) = \exp(-\frac{1}{2}n^2(r(I + T)(ut_j^+(1 + t_j^+)^{-1} \Pi \otimes t_j^+ \\
+ vr(R_\otimes r(A)h_k))).
\]

Thus \( Y_j^+ \) and \( U_k \) have a joint normal distribution with mean zero, and covariance matrix with entries

\[
c_{11} = t_j^+(1 + t_j^+)^{-1}n^2((I + T) \Pi \otimes t_j^+), \quad c_{22} = n^2(r(R)r(A)h_k), \\
c_{12} = r(t_j^+(1 + t_j^+)^{-1})(I + T) \Pi \otimes t_j^+, \ r(A)h_k).
\]

Consequently, \( E_p(Y_j^+)^2U_k^2 = E_p(Y_j^+)^2E_pU_k^2 + 2E_pY_j^+U_k = t_j^+ \times \\
n^2(r(R_\otimes r(A)h_k) + 2t_j^+(1 + t_j^+)^{-1}(r(A)r(R_\otimes)(I + T) \Pi \otimes t_j^+, h_k)^2). \)

Similarly, \( E_p(Y_j^+)^2U_k^2 = -t_j^+n^2(r(R)r(A)h_k) - 2t_j^+(1 + t_j^+)^{-1}(r(A)r(R_\otimes) \Pi \otimes t_j^+, h_k)^2. \)

Thus, \( E_p(Y_j^+)^2n^2(r(A)) = \pm t_j^+ \text{trace}(AR) \pm 2t_j^+(1 + t_j^+)^{-1}n^2(r(A)r(R_\otimes) \times \\
(I + T) \Pi \otimes t_j^+, h_k)^2, \) and \( E_pL_n^2n^2(r(A)) = -\frac{1}{2} \text{trace}(AR) \sum_{i=1}^n \log(1 - (t_i^+)^2) + \\
\sum_{i=1}^n t_i^+(1 + t_i^+)^{-1}n^2(r(A)r(R_\otimes)(I + T) \Pi \otimes t_i^+, h_k)^2. \)

The result follows from Baker (1978, Proposition 2, p. 77).

Let the entropy of \( N \) with respect to \( M \) be denoted \( H_M(N) \). Then:

**Proposition 4.1.** If \( Q \div QN(q, P) \) has margins \( Q_i \div QN(q_i, P_i) \), then

\[
I(Q) = I(P) + H_p(Q) = H_{P_1}(Q_1) - H_{P_2}(Q_2) + c_o \sum_{i=1}^\infty t_i(1 + t_i) n^2(r(A)r(R_\otimes)t_i)
\]

**Proof.** Since the behaviour of \( T \) only matters on the closure of the range of \( R_\otimes \), one can suppose that the support of \( T \) is that range. We already know, from the proof of Lemma 4.1, that \( E_Q \log(p_\otimes) \) exists and is finite. We now check that \( E_Q \log(p_\otimes) \) exists and is finite. To that end let \( I_1 \) be the indicator of \( (q \leq 1) \), and \( I_2 \) that of \( (q > 1) \). Then \( E_Q \log(q) = -E_QI_1 \log(q) + E_QI_2 \log(q) \leq E_QI_1 \times \\
(1 - q) + E_QI_2(q - 1) \leq 2 + E_pq^2 < \infty, \) where \( g(x) := x \cdot \log(x) \). Finally, since \( q_i = q_i \cdot I_i(x), E_Q \log(q_i) = E_{q_i} \log(q_i), \) which exists and is finite. The result then follows from Pinsker (1960) (2.4.7), p. 20, and Lemma 4.1.

**Remark 4.1.** If \( R \) is diagonal, \( I(Q) = H_p(Q) = H_{P_1}(Q_1) - H_{P_2}(Q_2) \). An upper bound for \( H_p(Q) \) is given by \( \log(1 + c_o(4n^2(r(R)A\otimes) + \text{trace}(AR))). \)

**Remark 4.2.** If \( P_{XY} \), the law of \( (X, Y) \), is of the form \((1 - \epsilon)P + \epsilon Q, \) where \( P \) and \( Q \) are Gaussian, and \( Q \) is strongly equivalent to \( P \) (Hajek, 1962), then \( I(P_{XY}) \) has the following form:

\[
I(P_{XY}) = I(P) + H_p(P_{XY}) - H_{P_1}(P_X) - H_{P_2}(P_Y) + \frac{\epsilon}{2} \text{trace}(TT_\otimes)
\]
where $\hat{T}$ is the trace-class operator characterizing the equivalence of $P$ and $Q$, and $T_\otimes$ is the Hilbert–Schmidt operator characterizing the equivalence of $P$ and $P_1 \otimes P_2$. Thus, in both cases of contamination considered, the average information has the same form: it is made up of the information of $P$, a contribution of the contaminating factor, and a mixing term.

**Remark 4.3.** As the reviewer remarked, the paper of Ihara (1978) exhibits formulas similar to those obtained here. The main differences are: here we give exact formulas, and do not suppose that $X$ and $Y$ are independent; Ihara assumes that $X$ and $Y$ are independent, states inequalities, but admits a wider class of laws than the ones used in the present paper. In the last section of the paper, where we also assume that $X$ and $Y$ are independent, the interest in having exact formulas lies in the fact that one can study the structure of the problem, and identify conditions which allow channel capacity to be achieved.

**Example.** (See Example 3.4). $H_p(Q) = 2r(II)^{-1}(I(\frac{\beta}{\gamma}) \log(2) + I'(\frac{\beta}{\gamma}))$. Set $u^2 := (R_2k_2, k_2)_2$, $v^2 := (R_1k_1, k_1)_1$, and $w^2 := u^2 + v^2$, and let $N$ denote the standard normal distribution on the real line. Then,

$$H_{P_1}(Q_1) = E_N(w^2 + (1 - w^2)(\cdot)^2) \log(w^2 + (1 - w^2)(\cdot)^2),$$

$$H_{P_2}(Q_2) = E_N((1 - w^2) + w^2(\cdot)^2) \log((1 - w^2) + w^2(\cdot)^2).$$

Since $x \log(x)$ is convex, and provided that $0 < w^2 < 1$, $H_{P_1}(Q_1) + H_{P_2}(Q_2) < H_p(Q)$. Thus the values of $R_1k_1$ and $R_2k_2$ have a definite influence on the value of $I(Q)$.

### 5. The Average Mutual Information for the Independent Channel Without Feedback, When the Noise is Distributed According to a QN-Law

A description of the framework adopted in this section is to be found in Baker (1978, p. 75). The difference in notation is due to the presence, for QN-laws, of a number of parameters larger than that which is necessary in the Gaussian case.

$Q_1$ is an arbitrary probability measure, $P_2 \sim N(m_2, R_2)$, and $Q_2 \sim QN(q_2, P_2)$. $Q_2$ has covariance $S_2$. Set $Q := Q_1 \otimes Q_2$, $Q' := Q_1 \otimes P_2$, and, if $Q_1 \sim QN(q_1, P_1)$, $P := P_1 \otimes P_2$. $\alpha : H_1 \to H_2$ is a Borel measurable map, $f := a \cdot J_1 + J_2$, and $F := (J_1, f)$. Translation by $c$ is denoted $t_c$. The following probabilities will be considered: $Q_{a} := Q \cdot t_{a(c)}^{-1}$, $Q_{a,c} := Q_1 \cdot a^{-1}$, $Q_{a,f} := Q \cdot f^{-1}$, and $Q_{F} := Q \cdot F^{-1}$. When $Q_1$ is a QN-law, we shall consider similar probabilities, replacing the letter $Q$ by the letter $P$, and finally, we shall write $Q'_{a}$ and $Q'_{F}$, when $Q_2$ is replaced by $P_2$. If $Q_{a,c}$ has a covariance, it will be written $R_a$. 
Again, if $X$ is the input message, and $Y$ the output, the chosen model is $Y = a(X) + N = f(X, N)$, and $(N, Y) = F(X, N)$. Thus, if $Q_1$ is the law of $X$, and $Q_2$ that of $N$, $Q_F$ is the law of $(N, Y)$, when the law of $(X, N)$ is $Q_1 \otimes Q_2$, $Q_{2,f}$ is the law of $Y$, and $Q_2,a$ that of $a(X)$.

**Lemma 5.1.** (a) When $Q_{2,a}(\text{range}(r(S_2))) = 1$, $Q_2$ and $Q_{2,f}$, respectively $Q_F$ and $Q_{1} \otimes Q_{2,f}$, are equivalent ($\equiv$ mutually absolutely continuous).

(b) If $R_a$ exists, and $R_a = r(S_2)C_2^r(S_2)$, where $C_2$ has finite trace, and if $m_a$, the mean of $Q_{2,a}$, is in the range of $r(S_2)$, then $Q_{2,a}(\text{range}(r(S_2))) = 1$, so that (a) obtains.

(c) Suppose $Q_1 = QN(q_1, P_1)$, and $a$ is linear, and bounded. If $Q_F$ and $Q_1 \otimes Q_{2,f}$ are equivalent, then $Q_{2,a}(\text{range}(r(S_2))) = 1$.

(d) Suppose $Q_1 = QN(q_1, P_1)$, where $P_1$ has mean zero, and $a$ a linear and bounded. If $Q_F$ and $Q_1 \otimes Q_{2,f}$ are equivalent, then $R_a = r(S_2)C_2^r(S_2)$ where $C_2$ has finite trace.

**Proof.** $P_2$ and $Q_2$ are equivalent (1). Thus $Q$ and $Q^G$, and consequently $Q_{2,f}$ and $Q_{2,f}^G$, respectively $Q_F$ and $Q_F^G$, are equivalent. Now $r(R_2)$ and $r(S_2)$ have the same range ((2); Baker, 1973a, Corollary, (b), p. 280), so that $Q_F^G$ and $Q_1 \otimes Q_{2,f}^G$, as well as $P_2$ and $Q_{2,f}^G$, are equivalent (Baker, 1976, Corollary 1a), p. 6, and Theorem 1, p. 5). Part (a) is thus true. Let us prove (b). Since $r(R_2)$ and $r(S_2)$ have the same range, $m_a$ is in the range of $r(R_2)$. Furthermore, (2) and the polar decomposition yield: $r(S_2) = r(R_2)r(I_2 + T_2)U_2$, and consequently $R_a = r(R_2)C_2^r(R_2)$, where $C_2'$ has finite trace. Thus $Q_{2,a}(\text{range}(r(R_2))) = 1$ (Baker, 1976, Corollary 1b), p. 6). The proof of (c) goes as follows. $F$ is an injection, and $F^{-1}(x) = (x_1, x_2 - a(x_1))$. Thus, by the change of variables formula, and for Borel $B$, $Q_F(B) = I_F(B)(q_1 \cdot I_1)(q_2 \cdot (J_2 - a \cdot J_1))$. Consequently, $Q_F$ is absolutely continuous with respect to $P_F$. Suppose then that $P_F$ and $Q_F$ are not equivalent, and that $Q_F(B) = 0$, but $P_F(B) > 0$. One must then have $a_1 = a_2 = 0$, and also, for almost every $x$ in $B$, with respect to $P_F$, $n_2^2(A_1(x_1 - a_1)) \cdot n_2^2(A_2(x_2 - a(x_1)) \cdot a_1) = 0$. Thus, in $B$, all $P_F$-almost surely, either $x$ is in $(a_1 + \ker(A_1)) \times H_2$, or it is in $(J_2 - a \cdot J_1)^{-1}(a_2 + \ker(A_2))$. Consequently: $P_F(B) \leq P_F((a_1 + \ker(A_1)) \times H_2) + P_F((J_2 - a \cdot J_1)^{-1}(a_2 + \ker(A_2)))$. But $P_F((a_1 + \ker(A_1)) \times H_2) = P_1(a_1 + \ker(A_1)) = 0$ (Baker, 1973b, Theorem 1, p. 293), and similarly $P_F((J_2 - a \cdot J_1)^{-1}(a_2 + \ker(A_2))) = P_2(a_2 + \ker(A_2)) = 0$. Thus $P_F(B) > 0$ is impossible, and $P_F$ and $Q_F$ are equivalent. Again $P_1$ and $Q_1$ are equivalent (1), and so then are, respectively, $P_{2,f}$ and $Q_{2,f}$, $P_{1} \otimes P_{2,f}$ and $Q_{1} \otimes Q_{2,f}$, $P_F$ and $P_1 \otimes P_2$. The result follows from Baker (1976, Corollary 3a), p. 7). Part (d) is proved similarly.

Assume now that $n_2^2(a(\cdot))$ is integrable with respect to $Q_1$, so that $R_e$ exists. $I(Q_F)$ will not have a different value if we suppose that $m_a = m_2 = 0$ (Baker, 1978, Corollary, p. 77). Whenever we write a Radon–Nikodym derivative involving $a(x_1)$, we shall assume the latter lies in the range of $r(S_2)$. We shall also use
the following notation whenever it makes sense: \( p_{x_1} = dP_{x_1}/dP_2, \)
\( q_{x_1} = dQ_{x_1}/dQ_2, \)
\( q_{x_1} = dQ_{x_1}/dP_{x_1}. \)
\( I_{\text{Leb}} \) denotes integration with respect to Lebesgue measure.

**Lemma 5.2.** (a) For almost every \( x_1, \) with respect to \( Q_2, \)

- (i) \( Q_2, x_1 \) and \( Q_2 \) are equivalent,
- (ii) \( q_{x_1} = \frac{p_{x_1} \cdot q_{x_1}^{-1} \cdot \left( q_{x_1} \cdot t_{a(x_1)}^{-1} \right)}{p_{x_1} \cdot q_{x_1}^{-1} \cdot \left( q_{x_1} \cdot t_{a(x_1)}^{-1} \right)}, \)
- (iii) \( \log(q_{x_1}) = \log(p_{x_1}) + \log(q_{x_1} \cdot t_{a(x_1)}^{-1}) - \log(q_{x_1}) \), \( Q_2 \)-almost surely (and thus \( P_{x_1}, P_{x_1} \), \( \ldots \)).

(b) \( \log(q_2 \cdot t_{a(x_1)}^{-1}) \) is \( Q_{2, x_1} \)-integrable, and \( E_{Q_{2, x_1}} \log(q_2 \cdot t_{a(x_1)}^{-1}) = H_{P_2}(Q_2). \)

(c) \( \log(q_2) \) is \( Q_{2, x_1} \)-integrable, and \( E_{Q_{2, x_1}} \log(q_2) \) exists and is finite.

(d) \( \log(p_{x_1}) \) is \( Q_{2, x_1} \)-integrable, and, for \( a(x_1) \) in the range of \( r(R_2), \)

\[
E_{Q_{2, x_1}}(p_{x_1}) = \frac{1}{n} \sum_{n} r_{n}^{-1}(a(x_1)) e_{2, n}^{2} - 2c_{0}(a(x_1), A_2 a_2),
\]

where \((e_{2, n}, n)\) is the family of eigenvectors of \( R_2 \) corresponding to the positive eigenvalues \((\epsilon, n)\).

**Proof.** If \( Q_2 = QN((a_2, a_2, A_2), (m_2, R_2)) \), then \( Q_{2, x_1} = QN((a_2, a_2 + a(x_1), A_2), (m_2 + a(x_1), R_2)) \). Thus equivalence obtains (Gualtierotti, 1979a), and \( q_{x_1} = q_{x_1} \cdot p_{x_1} \cdot q_{x_1}^{-1}. \) It is easy to see, from the definitions, that \( c_{Q_{2, x_1}} = c_{Q_2}, \)

and that \( q_{x_1}(x_2) = q_{x_1} \cdot t_{a(x_1)}^{-1}(x_2). \) Finally, since \( Q_2 \) and \( P_2, \) as well as \( P_{x_1} \) and \( P_2, \) are equivalent, the right-hand side of (a, ii) will contain, as factors, 0 and \( \infty \) only for sets of measure 0 (WRT \( Q_2 \)). Part (a) is thus checked. Part (b) follows from the proof of Proposition 4.1. We now prove (c). One has \( E_{Q_{2, x_1}}(\log(q_{x_1})) \leq E_{Q_{2, x_1}}(q_{x_1} - 1) < \infty. \) Furthermore, if \( G \) is the distribution function of \( Q_{2, x_1} \), \( q_{x_1}^{-1}, E_{Q_{2, x_1}}(\log(q_{x_1})) = E_{G_I(1/n, 1/x)}(x) \log(1/x) = \lim_{n} E_{G_I(1/n, 1/x)}(x) I_{\text{Leb}}(1/n, 1/x) = \lim_{n} E_{G_I(1/n, 1/x)}(x) I_{\text{Leb}}(1/n, 1/x)(1/t)G(1/t) - G(1/n)). \) But \( G(1/t) - G(1/n) = Q(1/n < q_{x_1} \cdot t_{a(x_1)} < 1/t). \) So, if \( q_{2, x_1}(x_2, t_{a(x_1)}) > 1/t \) large enough, say \( t = t_0, q_{2, x_1} > 1/t. \) Consequently, \( \lim_{n} E_{G_I(1/n, 1/x)}(x) I_{\text{Leb}}(1/n, 1/x)(1/t)G(1/t) - G(1/n) \leq 0. \) It is thus sufficient to consider the case \( a_2 = 0. \) Define \( D_2 x_2 \) to be \( r(c_{Q_2})r(A_2)(x_2 + a(x_1) - a_2). \) Then \( (1/n < q_{2, x_1} \cdot t_{a(x_1)} < 1/t) = D_2^{-1}(1/n < n_2^2(x_2) < 1/t), \) so that, if \( B \) is the set \( (1/n < n_2^2(x_2) < 1/t), \)

\( G(1/n) - G(1/t) = E_{P_2, D_2^{-1}I_{n_2^2(x_2) < 1/t}} < 2(1/n + n_2^2(r(A_2)a(x_1))). \) If \( Q_2 \) is not Gaussian, as supposed, \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero. \( P_2 \) has a covariance proportional to \( r(A_2)R_2, \) which is different from zero.
pulations involving \( u_2 \) are also independent of \( x_1 \), except for the factor containing \( a(x_1) \). The bounds thus obtained for \( E_Q \) are independent of \( x_1 \), with the exception of the \( Q_1 \)-integrable factor \( n_2 \phi(a(x_1)) \). Consequently \( E_Q \) is independent of \( x_1 \), with the exception of the \( Q_1 \)-integrable factor \( n_2 \phi(a(x_1)) \). For similar reasons, one can replace the "\(-\)" sign in the last expression by a "\(=\)" sign. Using the change of variables formula and Fubini's theorem, one gets: \( \infty > E_Q \) \( E_Q \phi(a(x_1)) \) \( \phi(a(x_1)) \) \( E_Q \phi(a(x_1)) \)

Let us now prove (d). Since \( P_2 \) is the random variable \( (r(R_2))^{-1} e_2, n_2 \). Then, for fixed \( x_1 \) such that \( a(x_1) \) is in the range of \( r(R_2) \), \( \log(p_{x_1}) = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2) x_n - \frac{1}{2} \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) (Rao and Varadarajan, 1963, p. 317). Furthermore, \( \phi_2, n_2 = (e_2, n_2^2 + n_2 \phi_2(r(A_2)(x_2 - a(x_1)) \phi_2)) \cdot dP_2, n_2 \), and \( E_{P_2, n_2} \phi_2(a(x_1)) = \frac{1}{4} \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \). This last equality can be found in Baker (1979, Application). It is thus sufficient to check that \( \phi_2, n_2(r(A_2)(x_2 - a(x_1)) \phi_2) \) is \( P_2, n_2 \)-integrable. To that end, one must alter the form of \( \phi_2(a(x_1)) \) given above. Let \( Y_n \) be the random variable \( (\theta_n)(r(R_2))^{-1} e_2, n_2 \). Then \( X_n = Y_n + (\theta_n)^{-1}(a(x_1), e_2, n_2) Y_n - \frac{1}{2} \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( Y_n = \sum Y_n \phi_2 \) \( Y_n = \sum Y_n \phi_2 \phi_2 \). Let \( (e_2, \theta_n, \phi_2) \) complete \( (\phi_2, n_2) \) into an orthonormal basis, and define \( \phi_2(r(A_2)(x_2 - a(x_1))) \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \). Consider now the family of random variables \( Z_{n, p} \) defined by \( Z_{n, p} = (\phi_2, n_2)(r(A_2)(x_2 - a(x_1))) \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \). It is easy to check that \( \sup_p \phi_2, n_2(\phi_2, n_2) < \infty \). The family \( (\phi_2, n_2, \phi_2) \) is thus uniformly integrable, and consequently \( E_{P_2, n_2} \phi_2(r(A_2)(x_2 - a(x_1))) \phi_2 = \lim \phi_2, n_2 \phi_2(\phi_2, n_2) = \frac{1}{4} \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \) \( \phi_2 = \sum (r(\theta_n))^{-1}(a(x_1), e_2, n_2)^2 \). The result follows.

Lemma 5.3. Suppose \( Q_{n, f} := (Q_1^{(n)} \otimes Q_2) \cdot (a_n \cdot J_1 + J_2)^{-1} \) converges weakly to \( Q_1^{(n)} \), \( a_n^1 \) has mean zero, and a covariance \( R_n = r(S_2) C_n r(S_2) \), and furthermore that \( \sup_n \text{trace} (C_n) \) converges to \( Q_1^{(n)} \), and that \( Q_1^{(n)} \cdot a_n^1 \) has mean zero, and a covariance \( R_n = r(S_2) C_n r(S_2) \), and furthermore that \( \sup_n \text{trace} (C_n) \) converges to \( Q_1^{(n)} \), and that \( Q_1^{(n)} \cdot a_n^1 \) has mean zero, and a covariance \( R_n = r(S_2) C_n r(S_2) \), and furthermore that \( \sup_n \text{trace} (C_n) \) converges to \( Q_1^{(n)} \), and that \( Q_1^{(n)} \cdot a_n^1 \) has mean zero, and a covariance \( R_n = r(S_2) C_n r(S_2) \), and furthermore that \( \sup_n \text{trace} (C_n) < \infty \). Then

(a) \( Q \) is of the form \( Q_2 f \) and equivalent to \( P_2 \),

(b) the covariance of \( Q_{n, f} := (Q_1^{(n)} \otimes Q_2) \cdot (a_n \cdot J_1 + J_2)^{-1} \) has the form \( \hat{R}_n = r(S_2) C_n r(S_2) \), and \( \hat{R}_n \) has mean \( n_{a_n} \),

(c) if furthermore, \( E_{Q_2} \phi_2, n_2^2 \phi_2 \phi_2, n_2 \phi_2 = K \) converges to \( K < \infty \), some \( d > 0 \), then \( R_n \) converges weakly to \( \hat{R}_n \).

Proof. The hypothesis, relation (1), and Lemma 5.1 ensure that \( Q_{n, f}^{(n)} \) is equivalent to \( P_2 \). The proof of the present lemma is based on the uniform
integrability of the family \((dQ_{2,n}^*(dP_2, n))\), with respect to \(P_2\), which will be checked in several steps. In the first two steps the index \(n\) will be dropped to alleviate the notational burden. The following notation will be useful. If \(C\) is some operator, \(C'\) denotes an operator obtained from \(C\) in a way to be stated. \(p_n\) is the projection with range spanned by \(e_{a,1}, \ldots, e_{a,n}\); \(a\) the measurable map \(p_n \cdot a\), and \(f_p\) the map \(a_p \cdot f + f_2; P_2, t_a^{-1}(x)\) will be written \(P_2,x^{1/2}\), and \(q_n\) will stand for \(\sum_{i=1}^n r^{-1}_i(a(x), e_{a,i}) \alpha(x, e_{a,i}) q_2(x - a(x))\); finally, the function \(1/e + x \log(x)\), \(x > 0\), will be denoted \(g\).

The first step is to compute \(dQ_{2,n}^*(dP_2)\). One has:

\[
Q_{2,n}^*(B) = E_0 I_B (a_p(x_1) + x_2) = E_0 I_B (a_p(x_1)) E_0 I_B (x_2) = E_0 I_B (x_2 - a_p(x_1)) \exp(q_n(x)),
\]

since \(a_p(x_1)\) is in the range of \(r(R_2)\). But \(q_n(x_2 - a_p(x_1)) \cdot \exp(q_n(x))\) is positive, and measurable with respect to \(B(H_1) \otimes B(H_2)\), so that, by Fubini’s theorem,

\[
dQ_{2,n}^*(dP_2) = E_0 q_2(x_2 - a_p(x_1)) \exp(q_n(x)).
\]

The second step is to show that \((dQ_{2,n}^*(dP_2, p)\) is uniformly integrable. Since \(g\) is non-negative, convex, and 

\[
\lim_{x \to 0} \left(1/x\right) g(x) = 0,
\]

it is sufficient to check that

\[
\sup_{a(x_1)} E_0 g(dQ_{2,n}^*(dP_2)) < \infty.
\]

Now Jensen’s inequality gives

\[
g(dQ_{2,n}^*(dP_2)) \leq E_0 g(q_2(x_2 - a_p(x_1)) \exp(q_n(x))),
\]

so that

\[
E_0 g(dQ_{2,n}^*(dP_2)) \leq E_0 g(q_2(x_2 - a_p(x_1)) \exp(q_n(x))).
\]

But \(q_n dP_2 = dP_{2,x^{1/2}}\), and thus the right hand side of the last inequality can be written:

\[
1/e + E_0 q_2(x_2 - a_p(x_1)) (\log(q_2(x_2 - a_p(x_1))) + q_n(x))
\]

\[
= 1/e + E_0 q_2(x_2) (\log(q_2(x_2))) + \frac{1}{2} \sum_{i=1}^n r^{-1}_i(a(x_1), e_{a,i})^2
\]

\[
= 1/e + H_{P_2}(Q_{2,n}) + \frac{1}{2} \sum_{i=1}^n C (C') (\log(q_2(x_2)) + \frac{1}{2} \text{trace}(C')) < \infty,
\]

where \(C' = r(I_2 + T_2) U_2 C U_2^* r(I_2 + T_2)\).

One can now show uniform integrability. Since \(a(x_1)\) is \(Q_{2,n}\)-almost surely in the range of \(r(R_2)\) (Lemma 5.1), \(Q_{2,n}\)-almost surely, \(a_p(x_1)\) converges to \(a(x_1)\), so that, if \(F'\) is bounded, and continuous, \(E_{Q_{2,n}} f' = E_0 f' \cdot f_p = E_0 f_{Q_{2,n}} f' \cdot f_p\), and thus, applying twice the dominated convergence theorem,

\[
\lim_{n \to \infty} E_{Q_{2,n}} f' = E_{Q_{2,n}} f'.
\]

So \(Q_{2,f,p}\) converges weakly to \(Q_{2,f}\). But the functional \(E_{L} \log(dL/dM)\) is lower semicontinuous in \(L\), for the topology of weak convergence of probability measures (Bretagnolle, 1979, Exposé 3, (2.3)). Consequently,

\[
E_{Q_{2,f}} g(dQ_{2,n}^*(dP_2)) \leq \lim_{n \to \infty} E_{Q_{2,f,p}} g(dQ_{2,n}^*(dP_2)) \leq H_{P_2}(Q_{2,n}) + \frac{1}{2} \text{trace}(C') < \infty.
\]

Finally, replacing \(Q_{2,n}\) by the measures \(Q_{2,f}^{(n)}\), one has

\[
\sup_n E_{Q_{2,f,p}} g(dQ_{2,n}^*(dP_2)) < \infty.
\]
One proves (a) as follows. Taking if necessary a subsequence, since $Q_{a,d}^{(n)} := Q_{a,d}^{(n)} \cdot \alpha^{-n}$ is relatively compact (Parthasarathy, 1967, Theorem 2.2, p. 154), one can suppose that $Q_{a,d}^{(n)}$ converges to some probability $Q_a$. Now the function $n_2^a(\cdot)$ is positive, and (lower semi-)continuous, and the function $J'(M) := E_0 n_2^a(\cdot)$ is lower semicontinuous for the topology of weak convergence (Dellacherie and Meyer, 55 Théorème, p. 155), so that $\int J'(Q_a) dQ_a = \lim_{n \to \infty} J'(Q_{a,d}^{(n)})$.

$Q_a$ has thus a mean $m_a$, and a covariance $R_a$. Similarly, one obtains that $(R_a h_2, h_2) + (m_a, h_2)^2 = E_{Q_a} h_2^2 \leq \lim_{n \to \infty} E_{Q_{a,d}^{(n)}} h_2^2 \leq c_1 \times n_2^a(\sigma(S_h, h_2))$. Consequently, $m_a$ is in the range of $\sigma(S_h, h_2)$, and $R_a = \sigma(S_h, h_2) \sigma(S_a) R_a$. $C_a \sigma(S_a)$ where $C_a$ is linear, and bounded (Baker, 1973a, Lemma 1, p. 279). $C_a$ can be chosen so that the closure of the range of $C_a$ is contained in that of $\sigma(S_h, h_2)$, so that, if $(f_p, s_p, p)$ are the eigenelements of $S_h$, then $(C_a f_p, s_p) \leq \lim_{n \to \infty} (C_n f_p, s_p)$, and thus, by Fatou’s lemma, and the hypothesis, $C_a$ has finite trace.

The result will follow from Lemma 5.1 if $Q_a$ is shown to be of the form $Q_{a,d}$, for indeed $FT(Q_a) = FT(Q_{a,d}) FT(Q_a)$. Let $D_a := (h_2: Q_a(h_2) > 0)$. $D_a$ is countable, and its elements shall be denoted $h_{2,i}$, and $\sum q_{0,i}$, $q_{0,i}$. For each $h_{2,i}$ a point $h_{1,i}$ is chosen uniquely in $H_1$, and $h_0$ is chosen arbitrarily in $H_0$. $a_{q_0,i}$ is defined as follows: $a_{q_0,i}(x) := h_{q_0,i}$, if $x = h_{1,i}$, and $a_{q_0,i}(x) := h_{2,i}$, if $x \neq h_{1,i}$. $a_{q_0,i}$ is measurable. If $q_0 > 0$, one sets $Q_{a,d}^{(n)} := 1/q_0 \sum q_{0,i} d_{h_{2,i}}$, $Q_{a,d}^{(n)} := 1/q_0 \sum q_{0,i} d_{h_{2,i}}$, where $d_i$ is the probability measure concentrated at $h_i$. If $q_0 < 1$, one sets $Q_{a,d}^{(n)} := (1 - q_0)^{-1} (Q_0 - q_0 Q_{a,d}^{(n)})$. $P_1$ denotes a Gaussian measure on $H_1$ with full support. If $q_0 = 1$, $Q_0 = Q_{a,d}^{(n)}$, and $Q_{a,d}^{(n)}$ is a measure for which $Q_{a,d}^{(n)}(h_a) = 0$, all $h_a$, and thus one can find Borel sets $N_0 \subseteq [0, 1]$, $N_a \subseteq H_a$, and a map $F_2 : (N_0, \sigma(N_0)) \to (N_a, \sigma(N_a))$ (the latter being complement in $[0, 1]$) such that: $\text{Leb}(N_a) = Q_{a,d}^{(n)}(N_a) = 0$, $F_2$ is a measurable bijection with measurable inverse, and $Q_{a,d}^{(n)} = F_2^{-1}(\text{Leb}(N_a))$. $P_1$ is Gaussian with full support, one gets similarly two sets $N_0$, $N_a$, and a map $F_1$. If $N_0 := N_0 \cup N_0$, and $N_a := F_1^{-1}(N_a \cap N_0)$, then $Q_{a,d}^{(n)}(N_a) = 0$. If furthermore $a_{q_0,i}(x) := a_{q_0,i}(x) I_1(x) I_2(x)$, where $I_1$ is the indicator of $(h_{1,i}, i)$, and $I_2$ that of $(h_{1,i}, i) \cap (H_1 \cap N_1 \cap N_0)$, $a_{q_0,i}$ is measurable, and $Q_0 := (q_0 Q_{a,d}^{(n)} + (1 - q_0) P_1) - a_{q_0,i}$ makes sense. It is then possible to check that, for every bounded, continuous $F'$, $E_{Q_0} F' = E_{Q_0} F'$. The proof of (b) and (c) is based on the following remark: if $G_0$ denotes $dQ_{a,d}^{(n)} / dP_2$, and $G$, $dQ_0 / dP_2$, then $G_n$ converges to $G$ in the $\sigma(L_1(P_2), L_\infty(P_2))$-topology. If $B_2 := \{x_2: \text{abs}((x_2, h_2) \leq p)\}$, then $\text{abs}((E_{Q_{a,d}^{(n)}} I_{B_2} \sigma(\cdot, h_2)) \leq r(E_{Q_{a,d}^{(n)}}(\cdot, h_2)) \leq r(Q_{a,d}^{(n)}(B_2, \sigma(\cdot, h_2)))$ (writing $Q_{a,d}^{(n)} = G_n dP_2$, and using Hölder’s inequality). Thus: $\text{abs}((m_0 - m_{a,d}, h_2) \leq \text{abs}(E_{Q_{a,d}^{(n)}} I_{B_2} \sigma(\cdot, h_2)) + E_{Q_{a,d}^{(n)}} I_{B_2} \sigma(\cdot, h_2) \leq E_0 I_{B_2} \sigma(\cdot, h_2) + (m_0 - m_{a,d}) r(Q_{a,d}^{(n)}(B_2, \sigma(\cdot, h_2)))$. Convergence in the weak topology of $L_1(P_2)$ yields (the index is $n$): $\text{abs}((m_0 - m_{a,d}, h_2) \leq E_0 I_{B_2} \sigma(\cdot, h_2) + ct. r(Q_0(B_2)))$. The result follows by letting $p$ increase. This terminates the proof of (b). For (c), one has: $0 \leq ((R_n + S_n) h_2, h_2) - (R_{a,d} h_2, h_2) \leq \text{abs}(E_{Q_{a,d}^{(n)}} L^2 I_{B_2} -
$E_0 L^2 I_{B_e} + E_0 L^2 I_{B'_e} + E_0 \omega^2 L^2 I_{B'_e}$, where $L(x_2) := (x_2 - m_O, h_2)$. Furthermore,

$$E_0 \omega^2 L^2 I_{B'_e} = E_{P_2} \text{abs}^{2-1/u}(L) G_n^{1-1/u} \cdot \text{abs}^{1/u}(L) G_n^{1/u} I_{B'_e} \leq E_{P_2} \text{abs}^{2-1/u}(L) G_n^{1-1/u} E_0^{1/u} \text{abs}(L) I_{B'_e}.$$

But $(2 - 1/u)v = 1 + v$, so that

$$E_0 \omega^2 L^2 I_{B'_e} \leq n_2^{1-1/u}(h_2) \cdot E_{O_{Q_2}} n_2^{1-1/u}(\cdot - m_O) \cdot (n_2^2 r(R_n + S_2) h_2) + (m_O, h_2) \delta^{1/2u} \cdot (Q_2(n, (B_{P_2})))^{1/2u}.$$

Choosing $v \leq 1 + d$, the last term becomes a constant times the last factor. So, as above, $(R_0 h_2, h_2)_2 = \lim_n ((R_n + S_2) h_2, h_2)_2$, and the result follows from (a).

**Proposition 5.1.** If $Q_{2,a}$ has covariance $R_a = r(S_2) C a r(S_2)$, where $C a$ has finite trace, then

$$I(Q_r) = \frac{1}{2} \sum_n r_n^{-2}(R_n e_{2,n}, e_{2,n}) + H_{P_2}(Q_2) - E_{Q_2,f} \log(dQ_{2,f}/dP_2).$$

**Proof.** This formula follows from a result in Baker (1979), and the ensuing considerations. The form of $R_a$, and Lemma 5.1 ensure that $a(x_1)$ is $Q_1$-almost surely in the range of $r(S_2)$, and consequently that $Q_2$ and $Q_{2,a}$ are $Q_1$-almost surely absolutely continuous (Gualtierotti, 1979a). Furthermore, $(dQ_{2,x_1}/dQ_{2,f})(x_2) = q_2(x_2 - a(x_1)) p_{e_{2,n}}(x_2)(dP_2/dQ_{2,f})(x_2)$. The first factor on the right-hand side is $B(H_1) \otimes B(H_2)$-measurable, and the middle factor is $B(H_1) \otimes B(H_2)^G$-measurable (Baker, 1979, Application). It then follows, since $P_2$ and $Q_2$ are equivalent, that $(dQ_{2,x_1}/dQ_{2,f})(x_2)$ is measurable with respect to $B(H_1) \otimes B(H_2)^G$. Finally, one has to secure that $\log(dQ_{2,f}/dQ_2)$ is $Q_2,f$-integrable. But $\log(dQ_{2,f}/dQ_2) = -\log(q_2) + \log(dQ_{2,f}/dP_2)$. In view of Lemma 5.2, one must show that, with respect to $Q_{2,f}$, $\log(dQ_{2,f}/dP_2)$ is integrable. But that follows from the proof of Lemma 5.3. All the conditions in the result of Baker (1979) are thus met. Lemma 5.2 then yields

$$E_{Q_2,x_1} \log(q_2) = \frac{1}{2} \sum_n r_n^{-1}(a(x_1), e_{2,n})^2 - 2c_{a}(a(x_1), A_{2,a}) + H_{P_2}(Q_2) - E_{Q_2,x_1} \log(q_2).$$

This last expression has to be integrated with respect to $Q_1$. But the right-hand side is $B(H_2)$-measurable, and thus one gets the result integrating with respect to $Q_1$. 
Remark 5.1. When the contaminating law is Gaussian, the formula is the same, but one must interpret $Q_e$ adequately.

Remark 5.2. Proposition 5.1 and its analogue in the Gaussian case illustrate well the differences which exist between the Gaussian and non-Gaussian cases: the formula for the Gaussian case is obtained by setting $Q_2 = P_2$ (Baker, 1979). It is shown in Baker (1978, Lemma 6) that, when $R_o$ is fixed, the maximum of $I(Q_F)$ is achieved for $Q_2 = P_2$ Gaussian, and the method consists in showing that the expectation in the formula is minimum for the only Gaussian measure which has the required covariance. In the non-Gaussian case, one cannot even assert the existence of a maximum: Gaussian laws are excluded, as shown in the next statement, and the hypotheses are not sufficient to insure that $I(Q_F)$ is semi-continuous over a compact set (for the topology of weak convergence).

Theorem 5.1. If $Q_2$ is not Gaussian, $Q_{2,t}$ is not Gaussian.

Proof. $Q_{2,t}$ has mean $-R_2A_2a_2$ and covariance $R_o + S_2 = R_o + R_2 + R_2A_2R_o - r(R_2)(r(R_2)A_2a_2 \otimes r(R_2)A_2a_2)r(R_2)$, where $A_o = 2o_a A_2$, so that, if $Q_{2,t}$ is Gaussian, $(1 - i(h_2, R_2A_2a_2) - \frac{1}{2}h_2^2(r(A_2) R_2h_2)) FT(Q_{2,t})(h_2) = \exp(-i(h_2, R_2A_2a_2) - \frac{1}{2}(R_2A_2R_2h_2, h_2) - \frac{1}{2}(R_2h_2, h_2) - \frac{1}{2}(R_2A_2a_2, h_2)^2_2$. If $u := (h_2, R_2A_2a_2)_2$, $v_2 := (R_2h_2, h_2)_2$, $w_2 := n_2^2(r(R_2) R_2h_2)$, and $X := (\cdot, h_2)_2$, then $FT(Q_{2,t})(h_2) = \exp(-itu - \frac{1}{2}t^2w_2 - \frac{1}{2}(X, X)_2)$. (23)

Replacing $t$ and $-t$ in (23), multiplying the resulting equation with Eq. (23), “membre à membre,” one gets

$$((1 - \frac{1}{2}t_2^2w_2) e - \frac{1}{2}t_2^2u_2) \cdot \operatorname{abs}^2(\operatorname{cf}_x(t)) = \exp(-t^2(w_2^2 + v_2^2 - u_2^2)).$$ (24)

If $u = 0$, $w = 0$, for otherwise the left-hand side of (24) would be zero for finite $t$. Thus, if $H_0$ is the subspace orthogonal to $R_2A_2a_2$, one has: $H_0 \subseteq \ker(A_2 R_2)$, so that the subspace generated by $R_2A_2a_2$ must contain the closure of the range of $R_2r(A_2)$. Since $Q_2$ is not Gaussian, the latter cannot be the trivial subspace, and thus $R_2A_2a_2$ is different from zero. Consequently, $R_2r(A_2) h_2 = (h_2, c_2) R_2 A_2 a_2$, with $c_2 := n_2^2(r(A_2) a_2) \cdot r(A_2) a_2$. Let $a_2 := n_2^{-1}(A_2) r(A_2) a_2$. Then, if $p_j$ is the projection with range spanned by $a_2$, one can write $R_2r(A_2) h_2 = R_2r(A_2) p_j h_2$, and thus obtain $r(A_2) R_2 h_2 = (R_2r(A_2) a_2', h_2)_2 a_2'$. Consequently: $n_2(r(A_2) R_2 h_2) = (R_2r(A_2) a_2', h_2)_2$ and $(R_2r(A_2) a_2, h_2)_2 = n_2(r(A_2) a_2') (R_2r(A_2) a_2', h_2)_2$, so that $u = u_1 u_2$, where $u_1 := n_2(r(A_2) a_2)$, and $u_2 := (R_2r(A_2) a_2', h_2)_2$. One finally obtains: $(1 - tu_1 u_2 - \frac{1}{2}t^2 u_2^2) \cdot \operatorname{cf}_x(t) = \exp(-itu_1 u_2 - \frac{1}{2}t^2(u_2^2 + u_2)(1 - u_2^2))$. Let now $g(z)$ denote the function $\operatorname{cf}_x(t + iy) = g(z)$ is of the form $g_1(z) \cdot \operatorname{exp}(g_2(z))$, where $g_1$ and $g_2$ are polynomials of order two. $g_2(z) = \frac{1}{2}(2 - u_2^2) + (u_1 + u_2)y^2 - \frac{1}{2}t^2 u_2^2 - it u_2(u_1 + u_2 y)$, where, if
$Q_2$ is not Gaussian, $u_1^2 < 2$. Then $g_1(z) = 0$, if and only if
$\frac{tu_2(u_1 + u_2y)}{2} = 0$. Thus $t$ cannot be zero, and the only
possibility is $u_1 + u_2y = 0$. Then $t$ must be $\pm \text{abs}^{-1}(u_2)r(2(2 - u_1^2))$, and $z$ be

$\pm \text{abs}^{-1}(u_2)r(2(2 - u_1^2)) - iu_2u_2^{-1}$. $g$ is thus analytic in a neighbourhood of the
origin, and has two singularities on the same horizontal line, away from the
t-axis, and symmetrically located with respect to the $y$-axis. $c_{f_2}$ cannot then be a
characteristic function (Lukacs, 1970, Corollary to Theorem 7.1.1., p. 193).

**Remark 5.3.** A similar result obtains when the contaminating law is Gaussian.

**Theorem 5.2.** Let $N$ be a finite, positive integer. The following conditions are
needed: for the probability $Q_1$, and the measurable map $a$,

1. there is a linear subspace $L_2$ of $H_2$ containing the support of $Q_{2,a}$, and
   of dimension not exceeding $N(\dim H_1 \geq N, \dim \text{supp}(Q_2) \geq N)$,

2. $Q_{2,a} (\text{range}(r(S_2))) = 1$,

3. $E_{Q_1} \sum_n s_n^{-1}(a(x_1), f_{2,n}^2) \leq P_0 < \infty$, $P_0$ fixed; $(f_{2,n}, s_n)$ eigen elements
   of $S_2$,

4. there is $d > 0$ such that $E_{Q_{2,a}} n_{\frac{2+d}{2}}(x_2) \leq M < \infty$.

If $Q = ((Q_1, a): (25)-(28)$ obtain), then $\max_Q I(Q_F)$ exists.

**Proof.** The following remark will be useful: if a probability $P$ has mean zero,
and covariance $R$, the smallest linear subspace containing the support of $P$ is the
 closure of the range of $R$, denoted hereafter $\overline{R}$. Indeed, the support of $P$ is in $\overline{R}$
(Ito, 1970). Furthermore, if $L$ is the smallest subspace containing this support,
one can choose orthonormal $e_n$'s, $p_n$'s, and $g_\alpha$'s such that $(e_n, p_n)$ is a basis for $L$, $(e_n, p_n, g_\alpha, n, p)$ is a basis for $\overline{R}$, and $(e_n, p_n, g_\alpha, n, p, q)$ spans $H$. Let $p_L$ be the
projection onto $L$. Then $\text{trace}(R) = E_p \text{trace}(p_L^2(x_2)) = E_p \text{trace}(p_L^2(x_2)) = E_p \text{trace}(p_L^2(x_2))$, so
that the subspace spanned by the $p_n$'s and the $g_\alpha$'s is contained in the kernel of $R$, and
$L$ contains $\overline{R}$.

Now (26) and (27) ensure that $R_a$ exists and the first step consists in proving
that $Q$ contains a measure $Q_{2,a}^G := (Q_1^G \otimes Q_2^G)^{-1}(a_G f_1 + f_2)^{-1}$, where $Q_1^G$
is Gaussian, $a_G$ is linear, and bounded, and $Q_{2,a}^G$ has covariance $R_a$. Now (25) and
the remark at the beginning of the proof imply that $\overline{R}_a \subseteq L_2$, and (26) yields $\overline{R}_a =
\overline{R}_a \equiv \overline{S}_2$. $Q_1^G$ and $a_G$ are manufactured as in Baker (1978, p. 82). The useful
fact here is that the range of $a_G$ is contained in $\overline{S}_2$, and consequently, that
$(a_G(x_1), x_2)^2 \leq \text{trace}(r(S_2))$, so that (25) and (26) are true
for $Q_{2,a}^G$. Condition (28) then yields that $a_G(x_1)$ is $Q_1^G$-almost surely in the range of $r(S_2)$, so that (25) and (26) are true
for $Q_{2,a}^G$. Condition (27) is true trivially, and (28) holds because expectation of
the power of the norm is lower semicontinuous with respect to weak con vergence (Dellacherie and Meyer, 1975, 55 Théorème, p. 155). A consequence is
that (26) and Lemma 5.1 imply $R_a = r(S_2) C_{a} r(S_2)$, where $C_a$ has finite trace,
and can (and will) be chosen such that $\overline{C}_a \subseteq \overline{S}_2$. Then $\text{trace}(C_a) \leq P_o (27).
Let $Q' := (Q_2, r; (Q_1, a) \in Q)$. Since, because of (27),

$$E_{Q_2, r} \sum_{n \geq p} (x_2, f_{2,n})_2^2 = \sum_{n \geq p} (r(S_2)(I_2 + C_2) r(S_2) f_{2,n}, f_{2,n})_2 \leq (1 + P_0) \sum_{n \geq p} s_n,$$

$$\limsup_{Q' \rightarrow Q} E_{Q_2, r} \sum_{n \geq p} (x_2, e_n)_2^2 = 0,$$

for the orthonormal basis obtained when completing $(f_{2,n}, n)$. $Q'$ is thus relatively compact (Parthasarathy, 1967, Theorem 2.2, p. 154). The next step consists in proving that $Q'$ is closed. Let thus

$$Q^{(n)}_2 := (Q_1^{(n)} \otimes Q_2) \cdot (a_n \cdot J_1 + J_2)^{-1}$$

converge weakly to $Q$. Relations (27) and (28), and Lemma 5.3 ensure that $Q$ is equivalent to $P_2$, and can be written as $(Q_1 \otimes Q_2) \cdot (a \cdot J_1 + J_2)^{-1}$. Furthermore, $Q_{2,a}$ has covariance $R_a = r(S_2) C_2 r(S_2)$, where $C_2$ has finite trace, and $R_n$, the covariance of $Q^{(n)}_1 \cdot a^{-1} = r(S_2) C_n r(S_2)$, converges weakly to $R_a$. Thus, by Lemma 5.1, (26) holds for $Q$. Condition (28) holds because of the semicontinuity property already used, and (27) follows from Fatou’s lemma, since

$$\sum_n s_n^{-1}(R_a f_{2,n}, f_{2,n})_2 = \sum_n s_n^{-1} \lim_p (R_p f_{2,n}, f_{2,n})_2 \leq \lim_p \sum_n s_n^{-1}(R_p f_{2,n}, f_{2,n})_2 \leq P_0.$$

Thus, one need only check (25), and this is done using the dominated convergence theorem for operators (Simon, 1979, Theorem 2.16., p. 38). Indeed, $R_a$ and $R_n$ are both dominated by $S_2$, and $R_n$ converges weakly to $R_a$, so that convergence takes place in the trace-norm, and thus in the uniform norm. Consequently, the eigenvalues of $R_n$ converge to those of $R_a$ (Simon, 1979, Theorem 1.20., p. 18). But each $R_n$ has at most $N$ eigenvalues different from zero, so that $R_n$ has at most $N$ eigenvalues different from zero. The conclusion is thus that $Q'$ is compact. The proof will be complete when it is proved that $I(Q_r)$ is upper semicontinuous. Since the expectation term in Proposition 5.1 is lower semicontinuous (Bretagnolle, 1979, Exposé 3, (2.3)), it is sufficient to prove that the term $\sum_n r_n^{-1}(R_a e_{2,n}, e_{2,n})_2$ is continuous with respect to weak convergence for probability measures. It has already been stated that the right-hand side inequality of

$$\liminf_p \sum_n r_n^{-1}(R_p e_{2,n}, e_{2,n})_2 \leq \limsup_n r_n^{-1}(R_a e_{2,n}, e_{2,n})_2 \leq \limsup_n r_n^{-1}(R_p e_{2,n}, e_{2,n})_2$$

obtains, so that one need only prove the one on the left-hand side. If $C'_n$ is the operator $r(I_2 + T_2) U_n C_n U^*_2 r(I_2 + T_2)$, then $r(R_n)$ ker($R_n$) ker($C'_n$), so that $R_n \subseteq C'_n$, and thus $C'_n$ has finite dimensional range, as well as finite trace. Write

$$\sum_{k=1}^N c_k (g_k^{(p)}, e_{2,n})_2^2$$

for this operator. One has

$$(C'_p e_{2,n}, e_{2,n})_2 \leq P_0 n (I_2 + T_2) \sum_{k=1}^N (g_k^{(p)}, e_{2,n})_2^2,$$
and $\sum_n \sum_{k=1}^N (g_k^{(y)}, e_{2,n})_2^2 = \sum_{k=1}^N \sum_n (g_k^{(y)}, e_{2,n})_2^2 \leq N$. Consequently, again by Fatou's lemma and Lemma 5.3, the required inequality obtains, and the theorem is proved.

**Remark 5.4.** When the noise is Gaussian, one can restrict attention, in the optimization process, to $Q_1$ Gaussian, and a linear, and bounded (Baker, 1978, Lemma 6, p. 81). The same proof as that of Theorem 5.2 applies in the Gaussian case, though it does not provide the value of channel capacity. However, in the Gaussian case, (28) can be dispensed of, since (c) of Lemma 5.3 follows from the properties of Gaussian measure.

The proof given here shows that the conditions bearing on the family of messages are compactness requirements, and that the average mutual information has certain continuity properties. The conjunction of compactness and continuity is sufficient to guarantee that channel capacity can be achieved only when certain moments of order two are continuous with respect to weak convergence of measures. When the noise is Gaussian, this continuity comes free, but otherwise supplementary conditions have to be introduced. These are of a non-physical nature, since they involve moments of order higher than two.

Finally, contamination has the effect of shifting the set of laws over which the maximum of average mutual information is sought sufficiently far away so that there are no common laws with the set to be considered in the Gaussian case (Theorem 5.1). This is already true when contamination occurs only in one direction, that is when $A$ has the form $a \otimes a$. It would thus be of interest to have an accurate estimation of channel capacity as a function of $A$: one could then assess how much gain or loss of capacity would accompany a contamination of the Gaussian noise.

**Acknowledgments**

We thank Dr. C. R. Baker (Department of Statistics, University of North Carolina, Chapel Hill) for providing not only reprints, but all the supplementary explanations we needed. The reviewer is also to be thanked for his comments.

Received: January 12, 1979; revised: February 13, 1980

**References**


BRETAGNOLLE, J. (1979), Formule de Chernoff pour les lois empiriques de variables à valeurs dans les espaces généraux, Astérisque 68, 33–52.


