On the Stability of Bayes Estimators for Gaussian Processes

Ian W. McKeague

Florida State University
Tallahassee, FL 32306

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

We consider the Bayes estimator $\hat{\theta}$ for a Gaussian signal process observed in the presence of additive Gaussian noise under contamination of the signal or noise by QN-laws, introduced by Guatteri (1979). Upper bounds on the increase in the mean square error of $\hat{\theta}$ over the minimum possible mean square error under contaminated noise or contaminated signal are given. It is shown that the performance of $\hat{\theta}$ is relatively close to optimal for small amounts of contamination.
On the Stability of Bayes Estimators for Gaussian Processes

by

Ian W. McKeague

FSU Statistics Report M642
USARO Report No. D-56

September, 1982

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research supported by a Florida State University COFRS Grant and US Army Research Office Grant DAAG29-82-K-0168.

Key Words: Robustness, Bayes estimators, Gaussian processes, QN-laws.

On the Stability of Bayes Estimators for Gaussian Processes

by

Ian W. McKeague

Abstract

We consider the Bayes estimator $\delta_0$ for a Gaussian signal process observed in the presence of additive Gaussian noise under contamination of the signal or noise by QN-laws, introduced by Gualtierotti (1979). Upper bounds on the increase in the mean square error of $\delta_0$ over the minimum possible mean square error under contaminated noise or contaminated signal are given. It is shown that the performance of $\delta_0$ is relatively close to optimal for small amounts of contamination.
1. Introduction.

The Bayesian approach to the robust estimation of a signal in the presence of noise has been studied extensively in recent years. Some authors, including Blum and Rosenblatt (1967), Solomon (1972), Watson (1974) and Berger (1982) have discussed procedures which can be used when only vague information concerning the prior distribution is available. Others, including Box and Tiao (1968), Masreliez (1975) and Ershov and Liptser (1978) have constructed estimators which are robust with respect to contamination of the noise distribution.

The purpose of the present article is to study the performance of the usual Bayes estimator (denoted $\delta_0$) for Gaussian prior and additive Gaussian noise under certain deviations from normality in either the prior or the noise distribution. It is shown that the performance of $\delta_0$ is relatively close to optimal for small amounts of contamination. The main results of the paper give upper bounds on the increase in the mean square error of $\delta_0$ over the minimum possible mean square error under a specific contaminated prior or contaminated noise distribution. These results make it possible to assess the loss caused by the use of $\delta_0$ under non-Gaussian conditions. The contaminated Gaussian laws used in this paper are QN-laws (quasi-noise laws) which were introduced by Gualtierotti (1979). QN-laws form an appropriate class of contaminated Gaussian laws for some infinite dimensional models arising in communication theory (see Gualtierotti, 1980). Gualtierotti (1982) recently studied the stability
of signal detection under mixtures of Gaussian laws as well as QN-laws. Contamination by Gaussian mixtures was shown to lead to worse behavior than contamination by QN-laws. In the present paper attention is restricted to contamination by QN-laws.

Section 2 contains some preliminary material on measures on locally convex spaces and a derivation of the Bayes estimator for Gaussian prior and Gaussian noise on infinite dimensional spaces. Section 3 contains a discussion of QN-laws defined on locally convex spaces and a description of the posterior distribution when the prior or the noise is a QN-law. Upper bounds for the increase in the mean square error of $\delta_0$ over the minimum possible mean square error under a QN-law prior or QN-law noise are given in Section 4. Some examples, including an application to Kalman filtering, are discussed at the end of the paper.

2. Preliminaries.

Let $\mathcal{S}$ and $\mathcal{T}$ be measurable spaces, $\mu_{XY}$ a probability measure on $\mathcal{S} \times \mathcal{T}$, $\mu_X$ and $\mu_Y$ the projections of $\mu_{XY}$. The conditional distribution $\mu_{X|Y}$, if it exists, is defined to be a probability measure on $\mathcal{S}$ for a.e. $\mu_Y(y)$ such that $\mu_{X|Y}(A)$ is measurable as a function of $y$ for each fixed $A \in \mathcal{S}$ and

$$\mu_{XY}(A \times B) = \int_B \mu_{X|Y}(A) d\mu_Y(y) \quad \text{for all } A \in \mathcal{S} \text{ and } B \in \mathcal{T}.$$ 

It follows from the definition that $\mu_{X|Y} << \mu_X$ a.e. $d\mu_Y(y)$. The following lemma, which is proved using Fubini's theorem, states the abstract Bayes formula of Kallianpur and Striebel (1968).
Lemma 2.1. Suppose that the conditional distribution $\mu_Y | x$ exists and the map $(x,y) \mapsto \frac{d\mu_Y | x}{d\mu_Y} (y)$ is $S \times T$ measurable. Then the conditional distribution $\mu_X | y$ exists and

$$\frac{d\mu_X | Y}{d\mu_X} (x) = \frac{d\mu_Y | x}{d\mu_Y} (y) \text{ a.e. } d\mu_X \otimes \mu_Y (x,y).$$

The probability measure $\mu_{XY}$ is to be defined through a prior distribution $\mu_X$ on $S$ and a noise distribution $\mu_N$ on $T$. $S$ is the parameter space and $T$ is the observation space. Let $f:S \times T \rightarrow T$ be an $S \times T/T$ measurable map. Define $\mu_{XY}$ by $\mu_{XY} (A) = 

$$\mu_X \otimes \mu_N \{(x,y) : (x,f(x,y)) \in A\}.$$

It is easily seen that $\mu_Y | x$ exists and is equal to $\mu_N \circ f_X^{-1}$ where $f_X : T + T$ is defined by $f_X(y) = f(x,y)$. When $\mu_X | y$ exists it is called the posterior distribution.

Before going further we need to make a brief detour through the theory of probability measures on topological vector spaces. Let $E$ denote a locally convex topological vector space with topological dual $E'$. The cylindrical $\sigma$-algebra on $E$ is the $\sigma$-algebra generated by $E'$ and is denoted $\sigma(E')$. Let $\mu$ be a probability measure on $\sigma(E')$ such that $\int_E \langle f, x \rangle^2 \, d\mu(x) < \infty$, for all $f$ in $E'$. Then $\mu$ has a mean $m$ and a covariance operator $R$ and under mild conditions $m$ belongs to $E$ and $R$ maps $E'$ into $E$ (See Vakhania and Tarieladze, 1978). Schwartz (1964) showed that if $E$ is quasi-complete then each covariance operator $R:E' \rightarrow E$ has a unique Hilbert space $H$, which is a vector subspace of $E$, such that the natural injection $j$ of $H$ into $E$ is continuous and $R = jj^*$. The Hilbert space $H$ is called the reproducing kernel Hilbert space (RKHS) of $R$. If the RKHS is separable with a CONS
(e_\text{n}, \text{n} \geq 1) then the covariance operator admits a series representation
\[ R = \sum_n \text{j}_n \otimes \text{j}_n, \text{ where } (u \otimes u)(f) = <f, u \cdot u, \text{ for } u \in \mathcal{E}, \text{ } f \in \mathcal{E}'>, \text{ and the}
\text{series converges to } R \text{ in the strong operator topology:}
\[ \sum_n <f, \text{j}_n> \text{j}_n + Rf \text{ in } \mathcal{E} \text{ for all } f \text{ in } \mathcal{E}'. \text{ A probability measure}
\mu \text{ on } \sigma(\mathcal{E}') \text{ is Gaussian if each } f \text{ in } \mathcal{E}' \text{ is a Gaussian random}
\text{variable under } \mu. \text{ The methods used in this paper depend on the existence}
of a separable RKHS for the covariance operators of Gaussian measures.
\text{For this reason, we assume throughout that } \mathcal{E} \text{ is quasi-complete and each}
\text{Gaussian measure } \mu \text{ has a mean } m \in \mathcal{E}, \text{ a covariance operator } R: \mathcal{E}' \to \mathcal{E}
\text{and a separable RKHS. Such a Gaussian measure is specified by } \mu = N(m, R).

\text{Now assume that } \mu_{\text{N}} = N(0, R_{\text{N}}) \text{ on } \sigma(\mathcal{E}') \text{ with RKHS denoted } H_{\text{N}} \text{ and}
injection j_{\text{N}}: H_{\text{N}} \to \mathcal{E}, \mu_{\text{X}} = N(m_{\text{X}}, R_{\text{X}}) \text{ on } \sigma(H_{\text{N}}), (S, S) = (H_{\text{N}}, \sigma(H_{\text{N}})),
(T, T) = (E, \sigma(E')) \text{ and } f(x, y) = j_{n}(x) \cdot y. \text{ Let } L_{\text{N}} \text{ denote the closure of}
\mathcal{E}' \text{ in } L^{2}(\mathcal{E}, \mu_{\text{N}}), U_{\text{N}}: L_{\text{N}} \to H_{\text{N}} \text{ the unitary operator defined by}
U_{\text{N}} f = j_{n} f, \text{ for } f \in \mathcal{E}'. R_{\text{X}} \text{ is a trace-class operator on } H_{\text{N}} \text{ so it}
\text{has a series representation } R_{\text{X}} = \sum_n \tau_{n} e_{n} \otimes e_{n}, \text{ where } \{n, n > 1\} \text{ is a}
\text{CONS in } H_{\text{N}}, \tau_{n} \geq 0 \text{ and } \text{tr}(R_{\text{X}}) = \sum \tau_{n} < \infty. \text{ I denotes the identity}
\text{operator on } H_{\text{N}}.

\text{The following proposition, well known for finite dimensional spaces,}
gives the posterior distribution } \mu_{\text{X}|y} \text{ for Gaussian prior } \mu_{\text{X}} \text{ and}
\text{Gaussian noise } \mu_{\text{N}}.

\text{Proposition 2.2. Let } \mu_{\text{N}} = N(0, R_{\text{N}}), \mu_{\text{X}} = N(m_{\text{X}}, R_{\text{X}}). \text{ Then the posterior}
distribution } \mu_{\text{X}|y} \text{ exists as a probability measure on } \sigma(H_{\text{N}}) \text{ and is}
given by } \mu_{\text{X}|y} = N(m_{\text{X}|y}, R_{\text{X}|y}), \text{ where}
\[ m_{\text{X}|y} = \sum_n \frac{\tau_{n}}{1 + \tau_{n}} \left\{ \left[ U_{\text{N}}^{-1}(e_{n}) \right] (y) + \frac{e_{n} m_{\text{X}}}{\tau_{n}} \right\} e_{n}, \quad R_{\text{X}|y} = R_{\text{X}}(I + R_{\text{X}})^{-1}. \]
Proof. Denote \( [U_N^{-1}(e_n)](y) \) by \( a_n(y) \). The \( a_n \) are i.i.d. \( N(0,1) \) random variables under \( \mu_N \) so that \( m_x | y \in H_N \) a.e. \( du_N(y) \). But, \( \mu_N = \mathcal{N}_x^{-1} \sim \mu_N \) for each \( x \in H_N \) (cf. McKeague, 1982, Theorem 2.1) so that by Baker (1976) \( \mu_Y \sim \mu_N \). Thus \( m_x | y \in H_N \) a.e. \( du_Y(y) \) and the pair \( (m_x | y, R_x | y) \) defines a Gaussian measure on \( \sigma(H_N) \) a.e. \( du_Y(y) \).

Now check the conditions of Lemma 2.1. \( \mu_Y | x \) exists and is equal to \( \mu_N \cdot \mathcal{N}_x^{-1} \). The map \( (x,y) \mapsto du_Y | x / du_Y(y) \) is \( \sigma(H_N) \times \sigma(E') \) measurable since

\[
\frac{du_Y | x(y)}{du_Y} = \frac{\mathcal{N}_x^{-1}}{du_N} \frac{du_N}{du_Y}(y) = \frac{du_N}{du_Y}(y) \exp \left\{ [U_N^{-1}(x)] - \frac{1}{2} \| x \|_H^2 \right\}
\]

\[
= \frac{du_N}{du_Y}(y) \exp \left\{ \sum a_n(y) \langle e_n, x \rangle - \frac{1}{2} \langle e_n, x \rangle^2 \right\},
\]

where the Radon-Nikodym derivative \( du_N \cdot \mathcal{N}_x^{-1} / du_N \) is given in McKeague (1982, Theorem 2.1), for instance. Now applying Lemma 2.1, the characteristic functional \( \hat{u}_x | y(u) = \int_{H_N} e^{i \langle u, x \rangle} du_x | y(x) \), for \( u \in H_N \), as a function of \( u \), is proportional to \( \int_{H_N} \lim_{k \to \infty} Z_k(x) du_x(x) \), where

\[
Z_k(x) = \exp \sum_{n=1}^{k} \{ i \langle e_n, u \rangle \langle e_n, x \rangle + a_n(y) \langle e_n, x \rangle - \frac{1}{2} \langle e_n, x \rangle^2 \}.
\]

Provided that \( \{Z_k, k \geq 1\} \) is uniformly integrable, the result now follows from routine calculations since the \( \langle e_n, x \rangle, n \geq 1 \) are independent \( N(\langle e_n, m_x \rangle, \tau_n) \) random variables under \( \mu_x \). But
\[
\int_{H_N} |Z_k(x)|^2 \, d\mu_X(x) \leq \int_{H_N} \exp \left\{ 2 \sum_{n=1}^{k} \alpha_n (y) \langle \phi_n(x) \rangle d\mu_X(x) \right\} = \exp \left\{ 2 \sum_{n=1}^{k} (\alpha_n^2(y) + \alpha_n(y) \langle e_n, m \rangle ) \right\},
\]

which shows that \( \{Z_k, k \geq 1\} \) is a.e. \( d\nu_Y(y) \) uniformly integrable with respect to \( \nu_X \), as required.

\section*{3. \( QN \)-Laws.}

Let \( E_1 \) and \( E_2 \) be locally convex spaces. Suppose that \( \mu = N(m, R) \) on \( \sigma(E'_1) \) with RKHS denoted \( H \) and injection \( j : H \to E_1 \); also let \( A : E_1 \to E'_1 \) be a symmetric non-negative operator, \( a \in R \), \( a \in E_2 \) and \( J : E_1 \to E_2 \) be a continuous linear map. Provided

\[
c^{-1} \equiv \int_{E_1} (a^2 + \langle A(J(x)-a), J(x)-a \rangle) \, d\mu(x) \leq 0,
\]

define a probability measure \( \nu \) on \( \sigma(E'_1) \) by \( \nu = \mu \) if \( c^{-1} = 0 \), otherwise by the relation

\[
\frac{d\nu}{d\mu}(x) = c(a^2 + \langle A(J(x)-a), J(x)-a \rangle).
\]

The measure \( \nu \) is called a \( QN \)-law and was introduced on Hilbert space by Gualtierotti (1979). If \( J^*AJ \) has a separable RKHS then \( c^{-1} = 0 \) if and only if \( j^*J^*AJj \) is trace-class, and in this case

\[
c^{-1} = a^2 + \text{tr}(j^*J^*AJj) + \langle A(J(m)-a), J(m)-a \rangle.
\]

It is always possible to assume that \( a \) is either zero or one. We shall assume that \( a = 1 \) and write \( \nu = QN((J,a,A),\mu) \). When \( E_1 = E_2 \) and \( J \) is the identity map write \( \nu = QN((a,A),\mu) \). Gualtierotti (1980) calculated the mean and covariance operator of \( \nu \) for the case of a separable Hilbert space. It is possible to extend this result to separable Banach spaces as follows.
Lemma 3.1. Suppose that $E_1$ is a separable Banach space and $J^*A^J$ has a separable RKHS. Then the mean $m^Q$ and covariance operator $R^Q$ of $v$ are given by

$$m^Q = m + u,$$
$$R^Q = R + 2cRJ^*AJ - u\Theta u,$$

where $u = 2cRJ^*A(J(m)-a)$.

Proof. (Sketch) Assume that $m = 0$ and consider just the evaluation of $R_v$.

Let $J^*A = \sum_{n=1}^{\infty} g_n \Theta g_n$, $g_n \in E_1^i$. Then, for $f \in E_1$,

$$\int_{E_1} \langle f, x \rangle^2 \langle AJ(x), J(x) \rangle \, du(x) = \sum_{n=1}^{\infty} \int_{E_1} \langle f, x \rangle^2 \langle g_n, x \rangle^2 \, du(x),$$

so that we can reduce to evaluating integrals of the form

$$\int_{E_1} \langle f, x \rangle^2 \langle g, x \rangle^2 \, du(x).$$

Choose $h_n \in E_1^i$ such that $j^*(h_n), n \geq 1$ is a CONS for $H$. Define

$$\pi_k = \sum_{n=1}^{k} \langle h_n, x \rangle R_h, \quad x \in E_1.$$

Then, by Tien (1978, Lemma 2), $\pi_k x$ converges a.s. $[u]$ to $x$. But

$$\int_{E_1} \langle f, \pi_k x \rangle^2 \langle g, \pi_k x \rangle^2 \, du(x) \leq \left( \int_{E_1} \langle f, \pi_k x \rangle^8 \, du(x) \right)^{\frac{1}{4}} \left( \int_{E_1} \langle g, \pi_k x \rangle^8 \, du(x) \right)^{\frac{1}{2}} \leq 105 \langle f, f \rangle^2 \langle g, g \rangle^2,$$

since $\langle f, \pi_k x \rangle$ is a $N(0, \sum_{n=1}^{k} \langle R_h, f \rangle^2)$ random variable and

$$\sum_{n=1}^{k} \langle R_h, f \rangle^2 \leq \langle R_f, f \rangle.$$ 

It follows that $\{\langle f, \pi_k x \rangle^2 \langle g, \pi_k x \rangle^2, k \geq 1\}$ is uniformly integrable and the Lebesgue convergence theorem can be applied.

The integral $\int_{E_1} \langle f, \pi_k x \rangle^2 \langle g, \pi_k x \rangle^2 \, du(x)$ can be calculated using the fact
that \( \langle h_n, x \rangle \), \( n \geq 1 \) is an i.i.d. \( N(0,1) \) sequence of random variables with respect to \( m \).

The next proposition shows that the posterior is a QN-law if either the prior is Gaussian and the noise is a QN-law or the prior is a QN-law and the noise is Gaussian. Let \( \mu_N = N(0,R_N) \), \( \mu_X = N(m_X,R_X) \) as in Section 2 and let \( \mu_X|_Y \) denote the corresponding posterior distribution given in Proposition 2.2.

**Proposition 3.2.** (i) If the prior is \( \mu_X = N(m_X,R_X) \) and the noise is \( \nu_N = QN((a,A),\mu_N) \) then the posterior is \( \nu_X|_Y = QN((j_N,y-a,A),\mu_X|_Y) \).

(ii) If the prior is \( \nu_X = QN((a,A),\mu_X) \) and the noise is \( \nu_N = N(0,R_N) \) then the posterior is \( \nu_X|_Y = QN((a,A),\mu_X|_Y) \).

The proof of this proposition uses the following consequence of Lemma 2.1.

**Lemma 3.3.** Let \( \mu_X|_Y \) and \( \nu_X|_Y \) be probability measures on \( S \times T \) such that

(a) \( \mu_X \ll \nu_X \) and \( \nu_Y \ll \nu_Y \);

(b) \( \nu_Y|_X \) and \( \nu_Y|_X \) exist and \( \mu_Y|_X \ll \nu_Y|_X \) a.e. \( du_X(x) \);

(c) the maps \( (x,y) \mapsto \frac{du_Y|_X}{du_X|_Y}(x,y), (x,y) \mapsto \frac{du_Y|_X}{du_Y(y)} \) are \( S \times T \) measurable. Then \( \nu_X|_Y \) exists, \( \nu_X|_Y \ll \nu_X|_Y \) a.e. \( du_Y(y) \) and

\[
\frac{du_X|_Y(x)}{du_X|_Y} = \frac{du_Y}{du_Y|_X}(y) \frac{du_Y|_X(x)}{du_Y|_X} \frac{du_X(x)}{du_Y|_X} \text{ a.e. } du_X \otimes du_Y(x,y).
\]

**Proof.** Using (a) and (b) get

\[
\frac{du_Y|_X(x)}{du_Y|_X} = \frac{du_Y|_X(x)}{du_Y|_X} \frac{du_Y(x)}{du_Y(y)} \frac{du_Y(y)}{du_Y(y)} \text{ a.e. } du_X \otimes du_Y(x,y).
\]
so that, by (c), the function \((x,y) \mapsto \frac{dv_Y|_x}{dv_Y(y)}\) is \(S \times T\) measurable and \(\nu_X|_y\) exists by Lemma 2.1. The proof is completed by applying Bayes formula.

Proof of Proposition 3.2. (i) \(\nu_Y \omega_Y\) since \(\nu_Y \omega_Y \omega_X\) for all \(x \in H_N\).

\[
\frac{dv_Y|x}{dv_Y|_x}(y) = c_N(1 + \langle A(y-a-j_N x), y-a-j_N x \rangle),
\]

so that the map \((x,y) \mapsto \frac{dv_Y|x}{dv_Y|_x}(y)\) is \(\sigma(H_N) \times \sigma(E')\) measurable. The map \((x,y) \mapsto \frac{dv_Y|x}{dv_Y}(y)\) is \(\sigma(H_N) \times \sigma(E')\) measurable from the proof of Proposition 2.2. Thus, by Lemma 3.3 \(\nu_X|_y\) exists and

\[
\frac{dv_X|y(x)}{dv_X|_y}(y) = \frac{dv_Y}{dv_Y}(y) c_N(1 + \langle A(j_N x-(y-a)), j_N x-(y-a) \rangle),
\]

which shows that \(\nu_X|_y = QN((j_N,y-a,A),\nu_X|_y)\). The proof of (ii) is similar. 

4. Bayesian Robustness.

Let \(\delta\) denote a decision rule for estimating the true signal \(x \in H_N\). \(\delta\) is a measurable function from the observation space \(E\) into the parameter space \(H_N\). For prior \(\nu_X\) and noise \(\nu_N\) the mean square error of \(\delta\) is given by

\[
r(\nu_X,\nu_N,\delta) = \int_{H_N \times E \times H_N} \| x-\delta(y) \| ^2 \, dv_X(y,x),
\]

The following functions of \(\nu_X\) and \(\nu_N\) will be used to measure the robustness of a decision rule \(\delta_o\): the increase in the mean square error in using \(\delta_o\) over the minimum possible mean square error,

\[
\Delta(\nu_X,\nu_N,\delta_o) = r(\nu_X,\nu_N,\delta_o) - \inf_{\delta} r(\nu_X,\nu_N,\delta),
\]

and the ratio of the mean square error using \(\delta_o\) to the minimum possible mean square error,
Let \( \delta_0 \) be the optimal (in the mean square sense) estimator for Gaussian prior \( \nu_X = N(m_X, R_X) \) and Gaussian noise \( \nu_N = N(0, R_N) \). Then \( \delta_0(y) = m_X|y' \), the posterior mean given in Proposition 2.2. The results of this section give some upper bounds on \( \Delta(\nu_X, \nu_N, \delta_0) \) and \( \phi(\nu_X, \nu_N, \delta_0) \) for \( \nu_X \) and \( \nu_N \) as QN-law contaminations of \( \mu_X \) and \( \mu_N \) respectively. First we evaluate the mean square error of \( \delta_0 \) under contaminated prior or contaminated noise.

Denote \( R_1 = R_X|y = R_X(I + R_X)^{-1} \).

**Lemma 4.1.** (i) Let \( \nu_X = QN((a, A), \mu_X) \). Then

\[
\frac{\text{tr}(R_1) + 2c_X\text{tr}(AR_1^2)}{\text{tr}(R_1)} = \frac{\text{tr}(R_1) + 2c_X\text{tr}(AR_1^2)}{\text{tr}(R_1)},
\]

where \( c_X^{-1} = 1 + \text{tr}AR_X + <A(m_X-a), m_X-a> \).

(ii) Let \( \nu_N = QN((a, A, \mu_N) \). Suppose that \( E \) is a separable Banach space and \( A \) has a separable RKHS. Then

\[
\frac{\text{tr}(R_1) + 2c_N\text{tr}(AR_1^2)}{\text{tr}(R_1)} = \frac{\text{tr}(R_1) + 2c_N\text{tr}(AR_1^2)}{\text{tr}(R_1)},
\]

where \( A = \sum_n A_n \) and \( c_N^{-1} = 1 + \text{tr}(A_n) + <A_n, a> \).

**Proof.** (i) \( \frac{\text{tr}(R_1) + 2c_X\text{tr}(AR_1^2)}{\text{tr}(R_1)} = \int_H \int_E \| m_X|y - x \|^2 \text{d}u_X|y'| \text{d}u_X(x) \). But

\[
m_X|y - x = \sum_{n \geq 1} \frac{\tau_n}{1 + \tau_n} \{ [u_N^{-1}(e_n)](y) - <x, e_n> - \frac{<x - m_X, e_n>}{\tau_n} e_n \},
\]

so that

\[
\int_E \| m_X|y - x \|^2 \text{d}u_Y|y'(y) = \sum_{n \geq 1} \left( \frac{\tau_n}{1 + \tau_n} \right)^2 \int_E ([u_N^{-1}(e_n)](y) - <x, e_n> - \frac{<x - m_X, e_n>}{\tau_n} \right)^2 \text{d}u_Y|y'(y)
\]

\[
= \sum_{n \geq 1} \left( \frac{\tau_n}{1 + \tau_n} \right)^2 \left( 1 + \frac{<x - m_X, e_n>^2}{\tau_n^2} \right),
\]
since \([u^{-1}_n(e_n)](y) \sim \langle e_n, x \rangle\) is a \(N(0, 1)\) random variable under \(\mu_{Y\mid X}\).

By Lemma 3.1

\[
\int_{\mathcal{H}_N} e_n^T(x - \mu_x) e_n^2 \, d\nu_X(x) = \tau_n + 2c_\chi \tau_n^2 \langle A e_n, e_n \rangle,
\]

so that

\[
\tau(v_X, \mu_N, \delta_a) = \sum_{n=1}^{\tau(n)} \left( \frac{\tau_n}{1 + \tau_n} \right)^2 (1 + \frac{1}{\tau_n} + 2c_\chi \langle A e_n, e_n \rangle)
\]

\[
= \text{tr}(R_X(I + R_X)^{-1}) + 2c_\chi \text{tr}(AR_X^2(I + R_X)^{-2}).
\]

(ii) is proved in a similar way.

The following theorem gives an upper bound on the increase in the mean square error of \(\delta_a\) over the minimum possible mean square error under a contaminated prior distribution.

**Theorem 4.2.** Let \(v_X = QN((a, A), \mu_X)\). Then

\[
\Delta(v_X, \mu_N, \delta_a) \leq 4c_1^2 \|R_1 A\|^2 \left[ \text{tr}(R_X R_1 + 2c_\chi \text{tr}(AR_X^2)(1 + 4c_\chi \|A R_X\|_2)) \right] \|m_X - a\|^2,
\]

where \(c_1^{-1} = 1 + \text{tr}(AR_X^2)\).

**Proof.** It is easily checked that \(\Delta(v_X, \mu_N, \delta_a) = \int_E \|m_X^Q \mid y - m_X^Q \mid y\|^2 \, d\nu_Y(y)\).

By Proposition 3.2 and Lemma 3.1, \(m_X^Q \mid y = m_X^Q \mid y + 2c_\chi \|R_X \mid y\| A(m_X \mid y - a)\), so that

\[
\Delta(v_X, \mu_N, \delta_a) \leq 4c_1^2 \|R_1 A\|^2 \int_E \|m_X^Q \mid y - a\|^2 \, d\nu_Y(y).
\]

Now consider
\[ \int_E \| m_X \|_{y-a} \|^2 \, d\nu_Y(y) = \int_{\mathcal{H}_N} \int_E \| m_X \|_{y-a} \|^2 \, d\nu_Y|_x(y) \, d\nu_X(x). \]

\[ \int_E \| m_X \|_{y-a} \|^2 \, d\nu_Y|_x(y) = \sum_{n \geq 1} \left( \frac{\tau_n}{1+\tau_n} \right)^2 \{ [U_N^{-1}(e_n)](y) - \langle e_n, x \rangle \}
\]

\[ + \langle e_n, x-a \rangle \left( 1 + \left( \frac{\langle e_n, x-a \rangle}{\tau_n} \right)^2 \right). \]

Use Lemma 3.1 to get

\[ \int_{\mathcal{H}_N} \{ \langle e_n, x-a \rangle \left( \frac{\langle e_n, m_X-a \rangle}{\tau_n} \right) \}^2 \, d\nu_X(x) = \]

\[ \tau_n + 2c_X \langle R_X A X e_n, e_n \rangle + 4c_X \left( \frac{\tau_n}{\tau_n} \right) \langle e_n, R_X A (m_X-a) e_n \rangle + \left( \frac{\tau_n}{\tau_n} \right)^2 \langle e_n, m_X-a \rangle^2. \]

This yields

\[ \int_E \| m_X \|_{y-a} \|^2 \, d\nu_Y(y) = \sum_{n \geq 1} \left\{ \tau_n^2 \langle 1+\tau_n \rangle^{-1} + 2c_X \left( \frac{\tau_n}{1+\tau_n} \right)^2 \langle R_X A X e_n, e_n \rangle \right\}
\]

\[ + 4c_X \langle R_X (I+R_X)^{-1} e_n, m_X-a \rangle + \langle e_n, m_X-a \rangle^2 \}
\]

\[ \leq \tau n^2 \langle 1+R_X \rangle^{-1} + 2c_X \tau n^2 \langle R_X \rangle (I+R_X)^{-1}^2 + 4c_X \| A R_X (I+R_X)^{-1} \| \| m_X-a \| \]

\[ + \| m_X-a \|^2 \],

and the result follows.

It is now possible to give an upper bound on \( \Theta(v_X, \mu_N, \delta_a) \), and since we are mainly interested in the effects of small amounts of contamination, we state it in the following form.
Corollary 4.3. Let \( \nu_X = QN((a, eA), \mu_X) \), where \( e > 0 \). Then

\[
\phi(\nu_X, \nu_N, \delta) \leq 1 + \frac{4 \| R_1 A \|^2 \left[ tr(R_1 R_1) + \| X - a \|^2 \right]}{tr(R_1)} (1 + o(1)) e^2, \quad as \quad e \to 0.
\]

In particular, \( \phi(\nu_X, \nu_N, \delta) = 1 + O(e^2), \quad e \to 0. \)

**Proof.** The result follows from Proposition 4.1, Theorem 4.2 and the identity

\[
\phi(\nu_X, \nu_N, \delta) = 1 + \frac{\Delta(\nu_X, \nu_N, \delta)}{\Delta(\nu_X, \nu_N, \delta)}.
\]

The next theorem gives an upper bound on the increase in the mean square error of \( \delta \) over the minimum possible mean square error under a contaminated noise distribution. In order to use the known formulae (Lemma 3.1) for the mean and covariance operator of a QN-law on \( E \) it is assumed for the remainder of this section that \( E \) is a separable Banach space and \( A \) has a separable RKHS.

**Theorem 4.4.** Let \( \nu_N \equiv QN((a, A), \mu_N) \). Then

\[
\Delta(\mu_X, \nu_N, \delta) \leq 8 c_2^2 \left[ \| R_1 A_N \|^2 \left[ tr(R_1 R_1) + 2 c_N tr A N \right]^2 \right.
\]

\[
\left. + \| A_N^2 A_N \| + 2 c_N^2 A_N^2 \right] + (1 + 4 c_N^2 \| A_N \|) < A_N, a >,
\]

where \( A_N = J_N^* A_N \) and \( c_2^{-1} = 1 + tr(A N R_1). \)

**Proof.** By Proposition 3.2, \( \nu_X \mid y = QN((J_N y - a, A), \mu_X \mid y) \), and by Lemma 5.1,

\[
\begin{array}{c}
\mu_{X \mid y} = \mu_X \mid y + 2 X \mid y R_1 J_N^* A(J_N \mu_X \mid y - a). \quad Thus
\end{array}
\]
\[ \Delta(u_X, v_N, \delta) = \int_E \| m_X|y - m_X|y \|^2 \, dv_Y(y) \]
\[ \leq 4c_2^2 \int_E \| R_1 j_N^{\sigma A}(j_N m_X|y - y + a) \|^2 \, dv_Y(y) \]
\[ \leq 8c_2^2 \left[ \| R_1 A_N \|^2 \int_E \| m_X - m_X \|^2 \, dv_Y(y) \right. \\
\left. + \int_E \| R_1 j_N^{\sigma A}(j_N m_X|y + a) \|^2 \, dv_Y(y) \right]. \]

It is easily checked that
\[ \int_E \| m_X|y - m_X \|^2 \, dv_Y(y) = \text{tr}(R_X R_1) + 2c_N \text{tr}(A N R_1^2). \]

Note that \( m_N^Q = j_N m_X + u \) and \( R_N^Q = j_N R_N j_N^* + R_N + 2c_N R_N A_N - u \bar{w} u \),
where \( u = -2c_N R_N A_N(a) \). Hence
\[ \int_E \| R_1 j_N^{\sigma A}(j_N m_X|y + a) \|^2 \, dv_Y(y) \]
\[ = \text{tr}(R_1^2 j_N^{\sigma A} A_N^* A_N) + \| R_1 j_N^{\sigma A}(a - u) \|^2 \]
\[ = \text{tr}(R_1^2 (A_N R_N A_N + A_N^2 + 2c_N A_N^3) - \| R_1 j_N^{\sigma A}(u) \|^2 \]
\[ + \| R_1 j_N^{\sigma A}(a - u) \|^2 \]
\[ = \text{tr}(R_1^2 (A_N R_N A_N + A_N^2 + 2c_N A_N^3) + \| R_1 j_N^{\sigma A}(a) \|^2 \]
\[ + 4c_N \langle R_1 j_N^{\sigma A}(a), R_N A_N j_N^{\sigma A}(a) \rangle \]
\[ \leq \text{tr}(R_1^2 (A_N R_N A_N + A_N^2 + 2c_N A_N^3) + (1 + 4c_N \| A_N \|) \langle A_N A_N, a \rangle). \]

The result follows immediately. \( \square \)

**Corollary 4.5.** Let \( v_N = QN((a, e A), \mu_N) \), where \( \epsilon > 0 \). Then
\[ \theta(u_X, v_N, \delta) \leq 1 + \frac{8[\| R_1 A_N \|^2 \text{tr}(R X R_1) + \text{tr}(R_1^2 (A_N R_N A_N + A_N^2)) + \langle A_N A_N, a \rangle]}{\text{tr}(R_1^2)}(1 + o(1)) \epsilon^2, \]
as \( \epsilon \to 0 \). In particular, \( \theta(u_X, v_N, \delta) = 1 + o(\epsilon^2), \epsilon \to 0 \).
Examples

1. **The one-dimensional case with contaminated prior.** Let $X$ and $N$ be independent random variables with distributions $\nu_X = \mathcal{Q}N((\mu_X, \epsilon), \mu_X)$ and $\nu_N = N(0, \sigma_N^2)$ respectively, where $\nu_X = N(\mu_X, \sigma_X^2)$. Then $Y = X + N$, $A = \epsilon \sigma_N$, and $R_X = \sigma_X^2 / \sigma_N^2 \equiv \rho$, the signal to noise ratio. By Corollary 4.3

\[
\frac{E(Y^2)}{\inf E(X^2)} \leq 1 + \frac{4\sigma_X^4}{(1+\rho)^2}(1+o(1))\epsilon^2
\]

2. **The one-dimensional case with contaminated noise.** Let $X$ and $N$ be independent random variables with distributions $\nu_X = N(\mu_X, \sigma_X^2)$ and $\nu_N = \mathcal{Q}N((0, \epsilon), \mu_N)$ respectively, where $\nu_N = N(0, \sigma_N^2)$. Then $A_N = \epsilon \sigma_N$, $R_X = \rho$ and by Corollary 4.5

\[
\frac{E(Y^2)}{\inf E(X^2)} \leq 1 + 8\sigma_N^4[1 + (\frac{\rho}{1+\rho})^2](1+o(1))\epsilon^2.
\]

3. **Kalman filtering in the presence of contamination.** Let the signal process $X_t$ and the observation process $Y_t$ be given by the stochastic differential equations

\[
\begin{align*}
\frac{dX_t}{dt} &= -\beta X_t dt + dW^1_t \\
\frac{dY_t}{dt} &= X_t dt + dW^2_t
\end{align*}
\]

(0 \leq t \leq 1), where $W^1$ and $W^2$ are independent Wiener processes, $\beta > 0$, and $X_0$ is a $N(0, \frac{1}{2\beta})$ random variable which is independent of $W^1$ and $W^2$. Then $E = \mathcal{C}[0,1]$, $H_N = L^2[0,1]$, $J_N : H_N \to E$ is defined by

\[
J_N(f)(t) = \int_0^t f(s) ds, \text{ for } f \in H_N, \ t \in [0,1],
\]

$R_N$ is the integral operator with kernel $\min(s,t)$ and $R_X$ is the integral operator on $L^2[0,1]$ with kernel $\frac{1}{2\beta} e^{-\beta|s-t|}$. It can be expressed as the solution of a stochastic
differential equation for the interpolation of a Gaussian process (see
Liptser and Shiryaev, 1978).

a) Contaminated signal. Let $A$ be the identity operator on $L^2[0,1]$
and let $v_X = \text{QN}((0, \varepsilon A), \mu_X)$, where $\mu_X = \text{N}(0, R_X)$. By Corollary 4.3
$\Phi(v_X, \mu_N, \delta_0) \leq 1 + 4 \text{tr}(R_X)(1 + o(1))\varepsilon^2$. But $\text{tr}(R_X) = 1/2B$ so that

$$\Phi(v_X, \mu_N, \delta_0) \leq 1 + \frac{2}{B}(1 + o(1))\varepsilon^2.$$ 

b) Contaminated noise. Let $A$ be the natural injection of $C[0,1]$ into $C^* [0,1]$ and let $v_N = \text{QN}((0, \varepsilon A), \mu_W)$, where $\mu_W$ is Wiener measure on $C[0,1]$. Thus $\frac{dv_N}{du_N}(x) = c_N (1 + \varepsilon \int_0^x 1^2 dt)$, where $c_N$ is a constant. By Corollary
4.5, $\Phi(v_X, v_N, \delta_0) \leq 1 + 24 \text{tr}(R_X)(\text{tr}(A_N))^2(1 + o(1))\varepsilon^2$. But $A_N$ is the integral
operator on $L^2[0,1]$ with kernel $\min(s, t)$. Thus $\text{tr}(A_N) = \frac{1}{2}$ and it
follows that

$$\Phi(v_X, v_N, \delta_0) \leq 1 + \frac{3}{B}(1 + o(1))\varepsilon^2.$$ 

Acknowledgements. The author would like to thank Debabrata Basu, Roger Berger,
Dennis Lindley and Jayaram Sethuraman for some helpful conversations.
REFERENCES


