"Signaling and uncertainty: a case study."

by

David A. Castanon
Nils R. Sandell

Abstract:

This paper studies the well-known counter example of Witsenhausen when the initial uncertainty is small. Using an asymptotic approach, it is established that linear strategies are asymptotically optimal over a large class of nonlinear strategies. This serves as a guideline for optimal solutions of non-classical problems with very noisy communication channels.

Acknowledgements

DOE-E(49-18)-2087, ONR-N0014-C-0346
Electronic Systems Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 Room 35-308

*Submitted to 1978 IEEE Conference on Decision and Control, and for publication in the IEEE Trans. on automatic control
1. Introduction

The key theoretical result in classical discrete time stochastic control is the so-called nonlinear separation theorem\[1,2\]. This theorem implies that the optimal control laws depend on the past observations and controls only through the conditional probability distribution of the present state, given these variables, and that this distribution is obtained by solution of a nonlinear filtering problem independent of the choice of control laws. The most important practical result in classical stochastic control is the solution of the linear-quadratic-Gaussian problem. This solution is a linear feedback of Kalman filter state estimates which is easily computed and implemented, and thus, has served as a basis for practical control law synthesis \[3\].

One of the key differences in decentralized control is the presence of decision makers with different available information. Decision problems with "classical" information patterns are characterized by the presence of one central decision maker endowed with perfect recall. In finite-state games, the absence of perfect recall gives rise to signalling information sets. Signalling is the act of transferring information from one decision maker to another through the use of decision variables. Perhaps the best known example of signalling in decentralized control was presented by Witsenhausen \[5\]. This example is discussed in detail in this paper. The control-sharing team problem discussed by Sandell and Athans \[6\] is another example of how control actions are used to convey information.
2. The Witsenhausen Counterexample (5)

Let $x_0, \nu$ be zero-mean independent random variables with variances $\sigma^2$ and 1 respectively. Consider the following stochastic control problem

State equations
\begin{align*}
x_1 &= x_0 + u_0 \\
x_2 &= x_1 - u_1
\end{align*} \hspace{1cm} (1)

Observation equations
\begin{align*}
y_0 &= x_0 \\
y_1 &= x_1 + \nu
\end{align*} \hspace{1cm} (2)

Cost function
\[ J = k^2 u_0^2 + x_2^2 \] \hspace{1cm} (3)

Admissible Controllers

\begin{align*}
u_0 &= \gamma_0(y_0) \\
u_1 &= \gamma_1(y_1)
\end{align*} \hspace{1cm} (4)

where $\gamma_1, \gamma_2$ are Borel functions.

This information pattern is nonclassical, since the value of $y_0$ is not known at stage 1. Hence, there is information to signal, and the observation $y_1$ provides a vehicle of communication. Figure 1 provides a diagram of the decision problem involved.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Problem: Minimize $E\{(x_1 - u_1)^2 + k^2 u_0^2\}$}
\end{figure}
Note the similarities between the diagram in Figure 1 and classical communications problems. If one denoted Controller 1 as the encoder, Controller 2 as the decoder, a classical communications problem would be

Minimize \( E \left\{ (u_1 - x_0)^2 \right\} \)

subject to \( E \{ u_0^2 \} < C \).

Thus, the Witsenhausen Counterexample is somewhat different in that one does not try to reconstruct the source; instead one seeks to reconstruct the output of the encoder.

Witsenhausen established that the optimal decoder was the conditional estimate of \( x_1 \) given \( y_1 \). However, the statistics of \( x_1 \) and \( y_1 \) depend on the form of the encoder. Assuming that \( \nu \) is Gaussian, the encoder problem can be reduced to the following equivalent problem:

Let \( h(x) = (2\pi e^x)^{-1/2} \); define

\[
D_f(y) = \int h(y-f(x))dF(x) \tag{5}
\]

\[
N_f(y) = \int f(x)h(y-f(x))dF(x) \tag{6}
\]

\[
g_f^*(y) = \frac{N_f(y)}{D_f(y)} \tag{7}
\]

\[
J_2^*(f) = J(f, g_f^*) \tag{8}
\]

where \( F(x) \) is the distribution of \( x_0 \). Witsenhausen showed that \( g_f^*(y) \) was the conditional expectation of \( f(x) \) with respect to \( y \). The resulting problem for the encoder was to minimize the expected value of \( J_2^*(f) \) with the choice of \( f(x) = x + y_0(x) \). Furthermore,

\[
E[J_2^*(f)] = 1 - I(D_f) + k^2 E \left\{ (x-f(x))^2 \right\} \]

\[
= 1 + k^2 E \left\{ (x - f(x))^2 \right\} \int_{-\infty}^{\infty} \left( \frac{d}{dy} \frac{D_f(y)}{D_f(y)} \right)^2 dy \tag{9}
\]
where $I(D_f)$ is the Fisher information content of the density $D_f$ (the density of $y$). Hence, the problem of choosing the optimal encoder consists of a tradeoff between maximizing the information available (signalling) and minimizing the use of control at the first stage.

Witsenhausen also proved the following results:

a. There exists an $f(x)$ which minimizes $E \{J_2^*(f)\}$ for arbitrary distributions $F(x)$.

b. The linear strategy $f(x) = \lambda x$ is a stationary point for a particular value of $\lambda$ when $F(x)$ is Gaussian.

c. For large values of $\sigma$, the strategy $f(x) = \lambda x$ is not optimal.

There exist numerous stationary points, and the optimal strategies $Y_0$, $Y_1$ are nonlinear.

When $\sigma$ is large, there is a lot of uncertainty in the system. In this case, it becomes optimal to use signalling to reduce the cost, as indicated by c above. However, it is not clear that signalling would be so evident if channel capacity was limited or the distortion noise $\nu$ had a large variance. Ideally, one would like to recognize problems where signalling would be present in terms of the parameters of the problem.

The next section studies the Witsenhausen counterexample when the variance $\sigma^2$ is small.

3. Asymptotic Analysis of the Witsenhausen Counterexample

Assume that $\sigma$ is very small, and that $F(x)$ is Gaussian. Then, one can obtain approximate expressions for the elements defined in equations (5)-(8), representing them as power series in $\sigma$. Consider a function $I(\sigma, y)$ defined by the integral equation
\[ I(\sigma, y) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} h(x, y) e^{-x^2/2\sigma^2} \, dx \quad (10) \]

Assume \( h(x, y) \) can be represented as
\[ h(x, y) = \sum_{i=0}^{m} a_i(y) \frac{x^i}{i!} + R_m(x, y) \]
where \( \lim_{x \to 0} \left| \frac{R_m(x, y)}{x^{m+1}} \right| < M(y) \) for all \( y \), and there exists \( b > 0 \) such that for all \( y \),
\[ \int_{-\infty}^{\infty} \left| R_m(x, y) e^{-x^2/2b^2} \right| dx < C(y). \]

**Definition 1** [7]: The series \( \sum_{i=0}^{m} a_i(y) \frac{x^i}{i!} \) is an asymptotic expansion of \( h(x, y) \) for \( x \) small if
\[ \lim_{x \to 0} \left| \frac{h(x, y) - \sum_{i=0}^{m} a_i(y) \frac{x^i}{i!}}{x^m} \right| = 0 \]
for each \( y \). The expansion is uniform if the limit is uniform in \( y \).

**Lemma 1**: The series \( \sum_{i=0}^{m/2} a_{2k}(y) \frac{2k}{2^k k!} \) is a uniform asymptotic expansion of \( I(\sigma, y) \) for small \( \sigma \), and it is uniform if \( M \) and \( C \) are independent of \( y \).

**Proof**: See Appendix.

Assume \( f(x) \) can be expanded in a series about \( x=0 \), as
\[ f(x) = \sum_{i=1}^{m} a_i(x) \frac{x^i}{i!} + R_m(x) \]
where \( \lim_{x \to 0} \frac{R_m(x)}{x^m} = 0 \). Then, using Lemma 1 and equations (5)-(11) one gets the following result (after considerable algebra):
Theorem 1: An asymptotic expansion of $J_2^*(f)$ to sixth order in $\sigma$ for small
is given by:

$$
J_2^*(f) = k^2 a_0 + \sigma^2 (k^2 a_1 - 1)^2 + k^2 a_0 a_2 + a_1^2 + \frac{\sigma^4}{4} (3k^2 a_2 + k^2 a_0 a_4 + 4k^2 a_3 (a_1 - 1) + 4a_1 a_3 - 4 a_1^4 + 2a_1^2) + \frac{\sigma^6}{24} (k a_0 a_6 + 10k^2 a_3^2 + 15k^2 a_2 a_4 + 6k^2 (a_1 - 1) a_5 + 6a_1 a_5 + 12a_2 a_4 + 10a_3^2 - 24a_2^2 - 48a_1^3 a_3 + 24a_1^6) + o(\sigma^6).
$$

Proof: See Appendix.

The asymptotic expansion in Theorem 1 gives an approximate cost; there are many strategies $f(x)$, differing slightly, which yield the same cost. The difficulty in optimizing the expansion in Theorem 1 with regards to the parameters $a_1$ of $f$ is that the expansion is not valid for arbitrary $f$. This leads to the next theorem.

Theorem 2: Suppose $f$ is restricted to belong to a set $\Gamma$ on which the expansion of Theorem 1 is uniformly valid. Then, the strategy

$$
f(x) = \frac{k^2}{1 + k^2} \left( 1 + 2\sigma^2 k^4 \frac{x^4}{(k^2 + 1)^3} \right)
$$

minimizes the asymptotic cost up to 6th order.

Proof: See Appendix.

Typical conditions for obtaining the set $\Gamma$ are uniform restrictions on the magnitude of the coefficients $a_1$. No such representation is given here because it is unnecessarily restrictive.
4. DISCUSSION

The asymptotic analysis of section 3 implies that, as the uncertainty in the initial state decreases, linear controllers are optimal relative to a class of nonlinear controllers. This class includes polynomial controllers with bounded coefficients. This suggests the existence of a region of values for $\sigma$ where linear controllers are optimal in general. Unfortunately, the results are not as strong as one would like due to their asymptotic nature and to the restrictions that must be placed on the class of admissible control laws to obtain uniformly valid expansions.

Problems with nonclassical information pattern are of great importance, while analytical results for these problems are virtually nonexistent. This paper suggests that linear controllers are approximately optimal in limiting cases, such as when there is little information to signal, or when the channel is very noisy or has limited capacity.
Appendix

Proof of Lemma 1:

Consider the difference $I(G, y) - I(a, y)$. From (10),

$$I(G, y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{k=0}^{m} \frac{a_k(y)}{k!} x^k e^{-x^2/2\sigma^2} dx$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} R_m(x, y) e^{-x^2/2\sigma^2} dx$$

$$= \frac{\hat{I}(G, y)}{\sigma^m} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} R_m(x, y) e^{-x^2/2\sigma^2} dx$$

Hence

$$\left| I(G, y) - I(a, y) \right| \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} R_m(x, y) e^{-x^2/2\sigma^2} dx$$

Now

$$\lim_{x \to 0} \frac{R_m(x, y)}{x^{m+1}} < M(y), \text{ so there exists } \varepsilon(y) > 0$$

such that, for all $|x| < \varepsilon(y)$,

$$|R_m(x, y)| < M(y) x^{m+1}$$

Thus

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} R_m(x, y) e^{-x^2/2\sigma^2} dx \leq \frac{2}{\sqrt{2\pi}} \int_{0}^{\varepsilon(y)} M(y) x \ e^{-x^2/2\sigma^2} dx$$

$$+ \frac{2}{\sqrt{2\pi} \varepsilon(y)} \int_{\varepsilon(y)}^{\infty} R_m(x, y) e^{-x^2/2\sigma^2} dx$$

$$\leq 2 M(y) \sigma^{m+1} + \frac{2C(y)}{\sqrt{2\pi}} \sigma e^{-\frac{1}{b^2} - \frac{1}{b^2}} \frac{\varepsilon^2(y)}{2}$$
Thus,
\[
\lim_{\sigma \to 0} \frac{I(\sigma, y) - \alpha I(\sigma, y)}{\sigma^m} = 0
\]
and the limit is uniform if \(C\) and \(M, \) (thus \(E(y)\)) are independent of \(y.\)

q.e.d.

**Proof of Theorem 1:**

Both \(D_f(y)\) and \(N_f(y)\) are integrals of the form \(I(\sigma, y)\) described in Lemma 1. Expansion of \(f(x)\) up to fifth order and straight substitution yields

\[
g_f(y) \approx f(o) + \frac{\sigma^2}{2} R_2(y) + \frac{\sigma^4}{8} \left( R_4(y) - 2R_2(y) \frac{d^2}{dx^2} h(x,y) \right)_{x=o} + \frac{\sigma^6}{48} \left( R_6(y) - 3R_4(y) \frac{d^2}{dx^2} h(x,y) \right)_{x=o} + \left. 6R_2(y) \left( \frac{d^2}{dx^2} h(x,y) \right)_{x=o}^2 - 3R_2(y) \frac{d^4}{dx^2} h(x,y) \right)_{x=o} + o(\sigma^6)
\]

where

\[
R_{2n}(y) = \left. \frac{d^{2n}}{dx^{2n}} (f(x) - f(o)) h(x,y) \right|_{x=o}
\]

and

\[
h(x,y) = e^{-1/2(f(x) - y)^2}
\]

Now, from (5), (6) and (7), it is established that

\[
\frac{d}{dy} D_f(y) = g_f(y) - y
\]

\[
\frac{D_f(y)}{D_f(y)}
\]

-10-
Using an asymptotic expansion to 6th order for the integral yields

\[
I(D_f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(o,y) \left( z^2 + \frac{\sigma^2}{2} \left( z^2 \frac{d^2}{dx^2} h(x,y) \bigg|_0 - 2R_2(y)z \right) \right) + \frac{\sigma^4}{8} \left( z^2 \frac{d^4}{dx^4} h(x,y) \bigg|_0 + 2R_2^2(y) - 2R_4(y)z \right) + \frac{\sigma^6}{48} \left( z^2 \frac{d^6}{dx^6} h(x,y) \bigg|_0 - 6R_2^2(y) \frac{d^2 h(x,y)}{dx^2} + 6R_2(y)R_4(y) - 2R_6(y)z \right) + o(\sigma^6) dy; \ z = y - f(o)
\]

The integrals defining \(I(D_f)\) are easily computed in terms of the \(a_i\) coefficients of \(f(x)\), yielding

\[
I(D_f) = 1 - a_1^2 \sigma^2 + \frac{\sigma^4}{2} (-2a_1a_3 + 2a_1^4 - a_2^4) + \frac{\sigma^6}{12} (-3a_1a_5 - 6a_2a_4 + 5a_3^2 + 12a_1a_2^2 + 24a_1a_3a_2 - 12a_1^6) + o(\sigma^6)
\]

Similarly, \(E \{x-f(x)\}\)

\[
\approx a_0^2 + \sigma^2 \left( (a_1-1)^2 + a_0a_2 + \frac{\sigma^4}{4} (3a_2^2 + a_0a_4 + 4a_3(a_1-1)) \right) + \frac{\sigma^4}{4} \left( 3a_2^2 + a_0a_4 + 4a_3(a_1-1) + \frac{\sigma^6}{24} (a_0a_6 + 10a_3^2 + 15a_2a_4) \right) + 6(a_1-1)a_5 + o(\sigma^6)
\]

Combining these two formulas yields Theorem 1.

\(\sigma.e.d.\)

**Proof of Theorem 2.**

Theorem 1 gives an asymptotic approximation to the cost \(J_2^*(f)\), as 
\[
J_2(a_0, \ldots, a_5).
\]
\(J_2(a_0, \ldots, a_5)\) can be written as

\[
J_2(\cdot) \approx k^2(a_0 + a_2 + a_4 + \frac{\sigma}{2} a_6)^2 + \frac{k^2}{2} (\sigma^2 a_2 + \sigma^4 a_4)^2 + \frac{\sigma^4}{2} (a_2 (1 - 2\sigma^2 a_1^2) + a_4 \sigma^2)^2 \frac{1 - 2\sigma^2 a_1^2}{1 - \sigma^2 a_1^2}
\]

\[
+ J_2 (a_1, a_3, a_5) + o(\sigma^6)
\]

where \(J_2 (a_1, a_3, a_5) = \sigma^2 (k^2(a_1 - 1)^2 + a_1^2) + k^2 a_3 (a_1 - 1) \sigma^4 + 4 a_1 a_3 - \sigma^4 a_1^2 + \frac{6}{12} (5 k^2 a_3^2 + 3 k^2 (a_1 - 1) a_5 + 3 a_5 - 24 a_1 a_3 + 12 a_1^3

The first 3 terms are positive, and achieve a minimum at \(a_0 = a_2 = a_4 = a_6 = 0\) independent of \(a_1, a_3, a_5\). Hence, the minimization reduces to

\[
\text{Min } J_2(a_1, a_3, a_5).
\]

Taking partial derivatives yields the necessary conditions

\[
\frac{\partial J_2}{\partial a_1} = 2 k^2 \sigma^2 (a_1 - 1) + 2 a_1 \sigma^2 + k^2 a_3 \sigma^4 - 4 a_1 \sigma^4 + k^2 a_5 \sigma^6
\]

\[
+ \left( \frac{a_5}{4} - 6 a_1 a_3 + 6 a_1^5 \right) \sigma^6 \approx 0
\]

\[
\frac{\partial J_2}{\partial a_3} = \sigma^4 (a_1 + k^2 (a_1 - 1)) + \frac{\sigma^6}{6} \left( 5 k^2 a_3 + 5 a_3 - 12 a_1 \right) \approx 0
\]

Because of the order of the approximation, these two equations are equivalent to

\[
2 k^2 (a_1 - 1) + 2 a_1 + k^2 a_3 \sigma^2 + a_3 \sigma^2 - 4 a_1^3 \sigma^2 \approx 0 \sigma^4
\]

\[
2 k^2 (a_1 - 1) + 2 a_1 + 5 a_3 \sigma^2 (a_1^2 + 5 a_3 \sigma^2 - 4 a_1^3 \sigma^2 \approx 0 \sigma^4
\]

Combining yields

\[
\frac{2}{3} a_3 \sigma^2 (1 + k^2) = 0 \sigma^4
\]

which implies \(a_3 \approx 0 \sigma^2\)

and \(k^2 (a_1 - 1) + a_1 - 2 \sigma^2 a_1^3 \approx 0 \sigma^4\)
Thus, \( a_3 \) is essentially zero, to the order of the expansion, since no value of \( a_3 \) of order \( \sigma^2 \) can contribute to the cost in the expansion up to sixth order. This implies

\[
a_1 \approx \frac{k^2}{1+k^2} \left( 1 + \frac{2\sigma^2 k^4}{(1+k^2)^3} \right)
\]

up to contributing costs of sixth order, and

\[
a_3 \approx 0.
\]

q.e.d.