Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Review of The Witsenhausen problem

Yu-Chi Ho, Harvard University & Tsinghua University, *Life Fellow, IEEE*

Abstract—**We review the status of the famous Witsenhausen problem after 40 years.**

INTRODUCTION

A celebrated problem in system and control is the so-called Witsenhausen problem. In 1968, H. S. Witsenhausen posed an innocent looking problem of the simplest kind. It consists of a scalar linear dynamic discrete time system of two time stages (thus involving two decisions at time stages one and two). The first decision is to be made at time one with perfect knowledge of the state, and there is a quadratic cost associated with the decision variable. The second decision can only be made based on noisy Gaussian observation of the state at time stage two, however, there is no control cost associated with the decision. The performance criterion is to minimize the quadratic terminal state after the two decisions. Thus, it represents the simplest possible Linear-Quadratic-Gaussian (LQG) control problem except for one small detail:

"***Instead of the usual assumption of one centralized decision maker who remembers at time stage two what s/he knows at time stage one, we do not have perfect memory or recall.***"

In fact, we have a decentralized team problem with two decision makers (DMs), DM1 and DM2 who do not have complete knowledge of what the other knows. Here the possibility for optimization is clear. DM1 knows the state of the system perfectly. S/he can simply use his/her control variable to cancel the state perfectly and leave DM2 nothing more to do. However, his/her action entails a cost. On the other hand, DM2 has no cost to act, but, without perfect memory, s/he has no perfect knowledge of the state of the dynamic systems at time stage two. A simple approach would be to strike a compromise using linear feedback control law for each decision maker, which is also known to be optimal under the traditional centralized LQG system theory for problems with perfect memory (i.e., the later decision maker at time stage 2 knows what the earlier decision maker knows at time stage 1). In fact, it is easy to prove that such a solution is a person-by-person optimal solution in equilibrium, i.e., if DM1 fixes his/her linear feedback control law, the best response by DM2 is a linear feedback control law and vice versa. However, Witsenhausen demonstrated that, without perfect memory, there exists a nonlinear control law for both DM1 and DM2, which involves signaling by DM1 to DM2 using its control action that outperforms the linear person-by-person optimal control law. In other words, the Witsenhausen problem presents a remarkable counterexample which shows that the optimal control law of LQG problems may not always be linear when there is imperfect memory. At the time, this was totally surprising since the problem seemed to possess all the right mathematical assumptions to permit an easy optimal solution. However, the globally optimal control law for such a simple LQG problem (or team decision problem) was unknown. The discrete version of the problem was known to be NP-complete (Papadimitriou - Tsitsiklis (1986)). Many attempts and papers on the problem followed in the next thirty and more years before the problem was understood and a numerical

Yu-Chi Ho is Professor emeritus of Systems Engineering and Applied Mathematics at Harvard University and a part time chair professor at Tsinghhua University, Beijing China.

solution of the globally optimal control law obtained in (*Lee et al 2001*).

The difficulty of the problem constitutes the essence of information structure (who knows what and when) in decentralized control, which is a subject worthy of a separate book. We shall not go into the matters here. Suffice is to say that the difficulties arises from the following intuitive explanation:

The information of the DM2 depends on the action of DM1 which is determined by the strategy or control law used by DM1 (see Eqs. (1) and (2) below). Since there is no prior reason that DM1 must use a linear control law, consideration of the possibility of a nonlinear control law will destroy any nice mathematical properties assumed in the original problem statement. Consequently. things such linearity, convexity, and Gaussianess all disappear. The simplest problem becomes the hardest problem. For a quick reference see (Ho 1980).

THE PROBLEM STATEMENT

Mathematically, the Witsenhausen problem can be described as follows. It is a two-stage decision making problem. At stage 1, we observe the initial state of the system x. Then we have to choose a control $u_1=\gamma_1(x)$ and the new state will be determined as $x_1=x+u_1=x+\gamma_1(x)$. At stage 2, we cannot observe x_1 directly. Instead, we can only observe $y=x_1+v$, where v is the additive noise. Then we have to choose a control $u_2=\gamma_2(y)$ and the system state stops at $x_2=x_1-u_2$. The cost function is $E[k^2(u_1)^2+(x_2)^2]$ with $k^2>0$ as a constant. The problem is to find a pair of control functions (γ_1, γ_2) which minimize the cost function. The trade off is between the costly control of γ_1 which has perfect information and the costless control γ_2 which has noisy information. We consider the famous benchmark case when $x\sim N(0,\sigma^2)$ and $v\sim N(0,1)$ with $\sigma=5$ and $k=0.2$.

Witsenhausen made a transformation from (γ_1, γ_2) to (f, g), where $f(x)=x+\gamma_1(x)$ and $g(y)=\gamma_2(y)$. Then the problem is to find a pair of functions (f, g) to minimize $J(f, g)$ where

$$J(f,g)=E\left[k^2\left(f(x)-x\right)^2+\left(f(x)-g\left(f(x)+v\right)\right)^2\right] \tag{1}$$

The first term in Eq. (1), $E[k^2(f(x)-x)^2]$, represents the cost shouldered by player one in the first time stage, so it is also called the stage one cost. The second term, $E[(f(x)-g(f(x)+v))^2]$, represents the cost shouldered by player two in the second time stage, so it is also called the stage two cost. Witsenhausen (Witsenhausen1968) proved that:

1) For any $k^2>0$, the problem has an optimal solution.

2) For any $k^2<0.25$ and $\sigma=k^{-1}$, the optimal solution in linear control class with $f(x)=\lambda x$ and $g(y)=\mu y$ has $J^*_{\text{linear}}=1-k^2$, and $\lambda=\mu=0.5\left(1+\sqrt{1-4k^2}\right)$. In the benchmark case that we consider, $k=0.2$, and $J^*_{\text{linear}}=0.96$.

3) There exist k and σ such that J^*, the optimal cost, is less than J^*_{linear}, the optimal cost achievable in the class of linear controls. Witsenhausen gave the following example. Consider the design: $f_W(x)=\sigma\,\mathrm{sgn}(x)$, $g_W(y)=\sigma\tanh(\sigma y)$, where $\mathrm{sgn}(\bullet)$ is the sign function, then the cost function J is $J_W=0.4042$. 4) For given $f(x)$ satisfying $E[f(x)]=0$ and $\mathrm{var}[f(x)]\le 4\sigma^2$, which are the conditions that the optimal $f^*(x)$ should satisfy, the optimal g^*_f associated with function f is

$$g^*_f=\frac{E\left[f(x)\varphi\left(y-f(x)\right)\right]}{E\left[\varphi\left(y-f(x)\right)\right]}, \tag{2}$$

where $\varphi(\bullet)$ is the standard Gaussian density function. Now the problem becomes that of searching for a single function f to minimize $J\left(f,g^*_f\right)$. Although the problem looks simple, no analytical method is available yet to determine the optimal f^*. The numerical optimal solution only came after over thirty years later and after many attempts (*Lee at al 2001 ibid*).

THE BENCHMARK CASE

Historically, many authors over a period of 30+ years attempted to find increasingly better solutions to the above benchmark case. Fig.1 shows a pictorial representation of their efforts where the subscripts stand for:

W: the original Witsenhausen solution (1968)
BB: *Banal R, Basar T (1987)*
DH: *Deng M, Ho YC (1999)*
LLH: *Lee JT, Lau EL, Ho YC (2001)*

There are other attempts on this benchmark case which were not known to the author at the time of making Fig.1. However, it was argued and believed that all features were explored and no further improvement can be made over that of J_{LLH}.

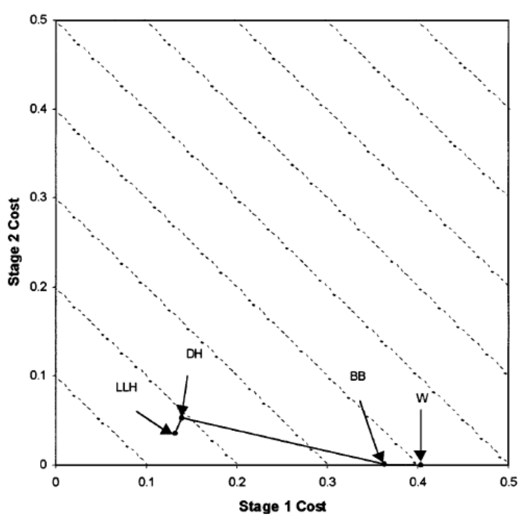

Fig. 1. Historical improvements on the Witsenhausen problem (benchmark: k=0.2 and σ=5) (*Lee et al. 2001*)

Finally, in 2007 in the book on Ordinal Optimization (*Ho-Zhao-Jia, Ordinal Optimization, Springer 2007*), they showed that using the methodology of ordinal optimization and limit the search of "*f*", the feedback strategy, to fixed number of bits for implementation ease they were able to systematically find a strategy that is 30 fold simpler (in terms of number of bits required) with only a performance degradation of no more than 5% over the best J_{LLH}.

REFERENCES

Witsenhausen HS (1968) "A counterexample in stochastic optimum control". *SIAM Journal on Control* 6(1):131-147.

Lee JT, Lau EL, Ho YC (2001) "The Witsenhausen counterexample: A hierarchical search approach for nonconvex optimization problems". *IEEE Transactions on Automatic Control* 46(3):382-397.

Deng M, Ho YC (1999) "Sampling-selection method for stochastic optimization problems." *Automatica* 35(2):331-338.

Banal R, Basar T (1987) "Stochastic teams with nonclassical information revisited: When is an affine law optimal?". *IEEE Transactions on Automatic Control* 32:554-559.

Ho YC, Zhao QC, Jia QS (2007), *Ordinal Optimization,* Springer-Kluwer Publishers.

Ho YC, (1980)"Team Decision theory and Information Structures" *Proceedings of the IEEE*, Vol. 68, No. 6, June 1980, pp 644-654.

Papadimitriou CH, Tsitsiklis JN (1986) Intractable problems in control theory. *SIAM Journal on Control and Optimization* 24(4): 639-654.