IEEE Information Theory Society Newsletter

Vol. 50, No. 4, December 2000 Editor: Kimberly Wasserman ISSN 1059-2362

ISIT 2000 Plenary Talk

Control with Limited Information: the Role of Systems Theory and Information Theory*

Sanjoy K. Mitter

Department of Electrical Engineering and Computer Science and The Laboratory for Information and Decision Systems M.I.T.

Cambridge, Mass. 02139

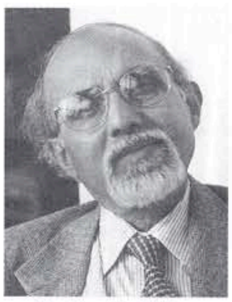
October 31, 2000

This is a somewhat expanded version of the author's plenary talk at ISIT 2000, Sorrento, Italy. It represents joint work with V. Borkar, N. Elia, A. Sahai and S. Tatikonda.

Introduction

A long-standing open conceptual problem has been the following: How does information "interact with control of a system, in particular feedback control, and what is the value of information" in achieving performance objectives for the system through the exercise of control? In answering this question we have to remember that in contrast to a variety of communications settings, the issue of time-delay is of primary importance for control problems, especially control of systems which are unstable.

The theoretical basis for modern digital communications is undoubtedly Information Theory as developed by Shannon. This theory tells us in a precise way the fundamental limitation to reliable communication over a noisy channel. The crowning achievement of this theory is the Noisy Channel Coding Theorem, which identifies the channel in terms of the invariant quantity, called capacity of the channel, and reliable communication can take place if transmission occurs at a rate below capacity and cannot if it occurs at a rate above capacity. This theorem links the input side of the communications problem via the notion of capacity with the output side, namely, the ability to decode with arbitrarily small probability of error. This theorem can be extended to the rate of distortion context as Shannon himself did! One can do no better than quote Shannon to illuminate this situation:



Sanjoy K. Mitter

> **Duality of a Source and a Channel**. There is a curious and provacative duality between the properties of a source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a "cost" associated with the difference input letters, and it is desired to find the capacity subject to the constraint that the expected cost not exceed a certain quantity. Thus input letter i might have cost a_i and we wish to find the capacity with the side condition $\Sigma_i P_i a_i \leq a_i$, say, where P_i is the probability of using input letter i. This problem amounts, mathematically, to *maximizing* a mutual information under variation of the P_i with a linear inequality as constraint. The solution of this problem leads to a capacity cost function $C(a)$ for the channel. It can be shown readily that this function is *concave* downward. Solving this prob-

Continued on page 3

*This research has been supported by the NSF-KDI Grant ECS-9873451 and by the Army Research Office under Grant DAAD 19-00-1-0466.

The founder of Information Theory Claude Elwood Shannon passed away on February 24th 2001 at the age of 84

An obituary will appear in the next newsletter and there will be a commemorative session on Sunday June 24th at ISIT 2001 in Washington.

Control with Limited Information: the Role of Systems Theory and Information Theory

Continued from page 1

lem corresponds, in a sense, to finding a source that is just right for the channel and the desired cost.

In a somewhat dual way, evaluating the rate distortion function $R(d)$ for source amounts, mathematically, to *minimizing* a mutual information under variation of the $q_i(j)$, again with a linear inequality as constraint. The solution leads to a function $R(d)$ which is *convex* downward. Solving this problem corresponds to finding a channel that is just right for the source and allowed distortion level. This duality can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it.

One of the many fundamental contributions which Shannon made which in fact renders the enunciation of the Noisy Channel Coding Theorem possible, is to think "digitally" (to use the word of a modern sage of Media technology), that is, to reduce everything to bits, a common currency in which everything can be evaluated. As we shall see later whether all bits are identical is an issue that we will have to face when dealing with the development of an Information Theory for sources which are decidedly nonstationary and non-ergodic.

A corresponding all embracing theory for control in the presence of uncertainty does not exist. The issue of fundamental limitations is far more complicated here since it is unclear that the dynamical systems which we wish to modify to behave in prescribed ways through control can be characterized through a simple invariant quantity like capacity. Even the invariants of a linear multivariable time-invariant system are the Kronecker invariants which tells us what Jordan forms we can reach through coordinate changes and linear constant feedback [1].The nearest thing to fundamental limitations of control systems analogous to Shannon theory are the Bode inequalities, the irreducible error in the Linear Quadratic Gaussian problem and characterization of performance limitations of control of linear time-invariant systems where the performance measure is sensitivity and this can be characterized through an H^∞-norm.

Nevertheless, control systems, even complex systems are being built where sensors, actuators, and controllers are being linked through noisy communications channels and a theory which unifies systems theory and a theory of information is badly needed. A caricature of such a system is shown in Fig. 1. The parts that are missing in the figure are estimators and coders between sensors and the channel, decoders between the channel and the controller. An analogous situation exists between the controller and the communication channel and between the channel and the actuator. The design problem now is to design the estimator, the coders and decoders and the controller to meet specified closed-loop design objectives. We immediately see that this is a far more complex problem than point to point communications. It is totally unclear whether the control part of the problem can be "separated" from the communications part of the problem. This problem is distributed and the issue of information structure, namely, what information is available when and where, is actually a design issue and must be understood.

The issues that I am raising are actually present in Communications problems where feedback and side information are present. In a conversation I had with Jim Massey at ETH in 1995, he pointed out that Shannon in the first Shannon lecture in 1973 had remarked that real time (time delay) issues and feedback in communication problems were questions which had received inadequate attention in Information Theory.

In light of the above discussion, I wish to raise two questions:

I. Is there a role for Information Theory in a unified theory of Control and Communications?

II. Can Systems Theory contribute to Communications and Information Theory in some non-trivial way? In my view, the answers to both questions are a Qualified Yes.

This is not the first time that these two questions have been posed. A successful interaction between Systems Theory and Coding Theory is through the work of Willems on the behavioral view of systems [2] and Forney, Massey, Trott, Loeliger, Mittelholzer on codes on Finite Groups (see e.g. [3]). There are also attempts at using rate distortion theory to obtain lower bounds on estimation error for non-linear filtering (see [4],[5]). Nevertheless we must proceed with cau-

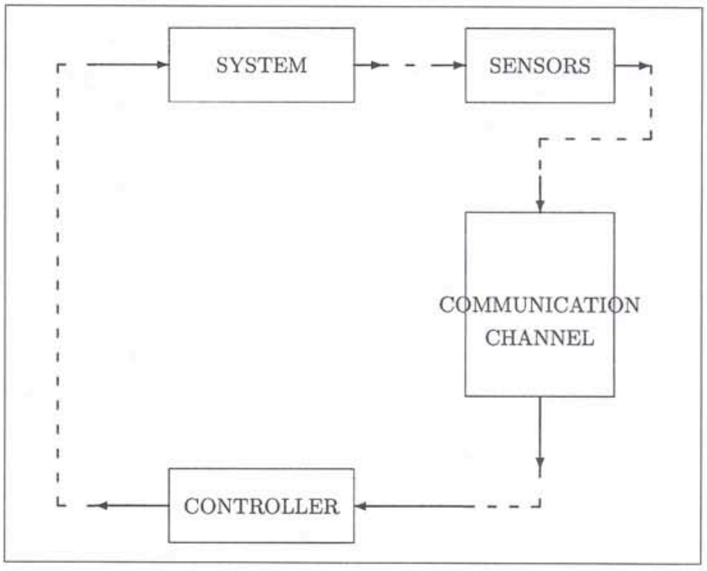


Figure 1

tion. This is best captured by quoting from Hans Witsenhausen [6] who thought deeply about these issues:

"The infimum expected cost achievable in a problem depends upon the prevailing information pattern. Changes in information produce changes in optimal cost. This suggests the idea of measuring information by its effect upon the optimal cost..."

"Such a measure of information is entirely dependent on the problem at hand and is clearly not additive. The only general property that it is known to possess is that additional information, if available free of charge can do no harm though it may well be useless. This simple montoncity property is in sharp contrast with the elaborate results of information transmission theory. The latter deals with an essentially simpler problem, because the *transmission* of information is considered independently of its *use*, long periods of transmission and use of channel are assumed and *delays* are ignored." H.S. Witsenhausen : 1971

In light of the above there is a methodological and theory formation issue which must be addressed. Simply stated, we must pose control questions in an appropriate informational sense and we must situate information theory in a dynamical framework. An elaboration of this viewpoint has been undertaken in the recently completed doctoral theses of Sekhar Tatikonda [7], Anant Sahai [8] and several papers, ([8],[9],[10],[11],[12]).

Control in an Information Setting

To make the above ideas more concrete let me consider the following question:

What is the minimal information about the current state of a single-input, discrete time, linear time-invariant unstable system needed in order to stabilize it?

The question we are asking is really about the optimal coding of the state, that is, coarsest vector quantization, to achieve stability. This problem is mathematically formulated in terms of the construction of Controlled Quadratic Lyapunov Functions (Quadratic for explicit computations). That is given

$$x(t+1) = Ax(t) + bu(t), t = 0,1,\dots \qquad (1)$$

where $x(t) \in X = R^n$ is the state of the system and $u(t) \in U = R$ is the control, A is an $n \times n$ matrix, b is an n-vector and we assume that (A, b) is a reachable pair, we are required to find the coarsest quantized feedback control which stabilizes the system. The idea of coarseness (minimal information) is captured as follows:

Given a controlled Lyapunov function

$$V(x) = (x;Px)_{R^n}, P > 0 \qquad (2)$$

find a set

$$U = \{u_i \in R | i \in Z\} \qquad (3)$$

and a quantizer

$$f: X \to U, \text{ with}$$

$$f(x) = -f(-x) \text{ and} \qquad (3a)$$
$$\Delta V(x) = V(Ax + bf(x))V(x) < 0 \qquad (3b)$$

$$\forall x \in X, x \neq 0.$$

f naturally induces a partition on the state space X and we assume that the values of f in U are ordered in the sense that $u_i < u_j, i > j, i, j \in Z$. Let $Q(V)$ = set of all quantizers which solves the stabilization problem. For $g \in Q(V)$ and $0 < \varepsilon \leq 1$, let $N(g[\varepsilon])$ denote the number of levels that g assumes in the interval $[\varepsilon, \frac{1}{\varepsilon}]$.

Define the quantization density

$$\eta_g = \lim_{\varepsilon \to 0} Sup \frac{N(g[\varepsilon])}{-\ln \varepsilon}, \text{ and} \qquad (4)$$
$$(5)$$

$$f^* = Arg.Inf_{g \in Q(V)} \eta_g \qquad (6)$$

f_* is defined to be the *coarsest*, quantizer corresponding to $V(x)$.

It turns out that the quantization problem can be confined to one preferred direction (one-dimensional), and the optimal quantization is logarithmic with the optimal scaling law ρ^* being given by

$$\rho^* = \frac{\Pi_{1 \leq i \leq k} |\lambda_i^u| - 1}{\Pi_{1 \leq i \leq k} |\lambda_i^u| + 1} \qquad (7)$$

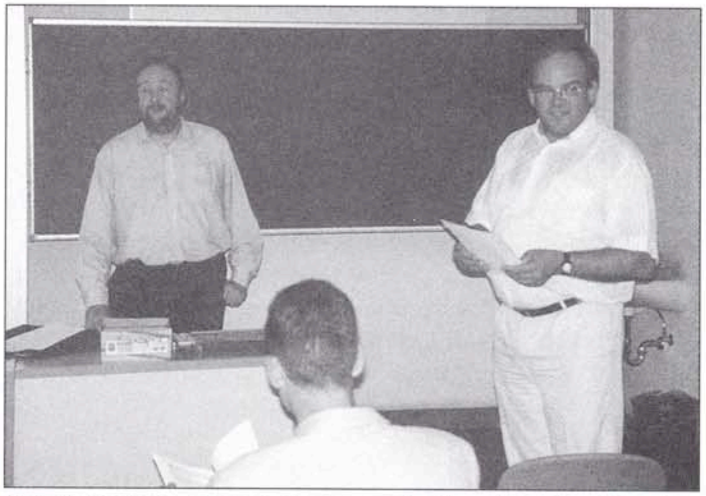
and $\lambda_i^u, 1 \leq i \leq k \leq n$ are the strictly unstable eigenvalues of A.

In the above formulation, we have allowed quantizers with a countable number of levels. An equivalent formulation of the problem leads to a method for designing finite quantizers leading to practical stability. For continuous-time systems there is a relation between the optimal sampling time T^* and the optimal quatization scaling law ρ^*:

$$T^* \sum_{i=1}^{k} \lambda_i^u(F) = \ln(1 + \sqrt{2}) \text{ and} \qquad (8)$$

$$\rho^*(T^*) = \sqrt{2} - 1. \qquad (9)$$

Continued on page 18



Peter Sweeney, invited speaker, introduced by Han Vinck.

Kees A. S. Immink, (IEM, Essen) Jin Ming, and B. Farhang-Boroujeny, (NUS Singapore): *Design technique of weakly constraint codes*

Abstract:

A general method of constructing variable length (d,k) constrained codes from arbitrary sequences is introduced. This method is then used for constructing a class of weakly constrained codes. The proposed codes are analyzed for the case of d=0 and shown to give results which are better or comparable to those of the best available codes, however at the cost of failure with some very low probability. Variable length codes can be very susceptible for error propagation. Brief results of decoding error propagation will be shown.

Hendrik Ferreira (RAU, Johannesburg): *On The Correction Of Insertion Deletion Errors*

Abstract:

We review some of our previously published results and consider the application of these results to the correction of burst errors and the use of convolutional codes

E.G.T. Jaspers and Peter de With (Philips research, University of Eindhoven, the Netherlands): *Synchronization of Base-Band Video for Multimedia systems*

Abstract:

A multimedia system is usually based on distributed computing, so that it consists of source, processing units and presentation devices in each component operates autonomously. This independency can be exploited for optimization of individual component performance, using e.g. dedicated block domains. This paper presents a Video I/O model for such a multimedia system enabling multiple video signal processing and display. This model provides an asynchronous communication interface for independent clock domains with the ability to synchronize a video display to one of the video sources. The communication model has been successfully implemented in I/O modules of an experimental multimedia system.

Samwel Martirossian, Sosina Martirossian and A.J. Han Vinck (Essen): *Optical Orthogonal Code Construction with Correlation 2*

Abstract:

Optical Orthogonal Codes (OOC) with low correlation between code words are used to allow multi user optical communication. We give a new construction for a class of (n,4,2)-codes We give the cardinality of our class of codes compared with the Johnson upper bound and conclude with computer search results.

[1] S. Bitan and T. Etzion, "Construction for Optimal Constant Weight Cyclically Permutable Codes and Difference Families," *IEEE Trans. on Information Theory*, vol. 41, pp. 77-87, Jan 1995.

Control with Limited Information: ...

Continued from Page 4

where $\lambda_i^u(F)$ are the unstable eigenvalues of the continuous-time system matrix F.

Note that $\rho^* \cdot T^*$ is an invariant of the class of single-input, continuous-time linear time-invariant systems. We may think of the quantized, stabilized feedback system as a symbolic description of the stabilized linear feedback system. It is also an example of source-coding problem with a non-standard criterion function. For details of above see [14].

The stochastic version of this problem where the quantized stabilization problem is posed for

$$x(t+1) = Ax(t) + bu(t) + w(t), t = 0,1,\ldots$$

and $w(t)$ is white Gaussian noise is even more interesting. Here one can exhibit the non-linear effects of quantization as a desirable effect as opposed to a source of noise which is undesirable and should be guarded against. A little bit later we shall see the desirable effects of quantization in a different context.

LQG and its Variants

The Linear Quadratic Gaussian problem is to control what the additive white Gaussian channel is to Information Theory. The problem is to control a linear stochastic system (single-input for simplicity)

$$X(t+1) = AX(t) + bu(t) + W(t) \qquad (10)$$

where $W(t)$ is white Gaussian noise in the presence of noisy measurements.

$$Y(t) = (c, X(t))_{R^n} + V(t) \qquad (11)$$

where $V(t)$ is white Gaussian noise and it is assumed that $V(t)$ and $W(t)$ are independent. The choice of control which is allowed to be a function of the past of the observations

$$u(t) = \varphi(Y(0), \cdots, Y(t-1))$$

has to be made so that we minimize a positive definite quadratic cost function

$$J(u;x0) = E\left[\sum_{t=0}^{T-1}(X(t), QX(t)) + (u(t), Ru(t)) + (X(t),\ SX(t))\right]$$

Infinite-time versions of this problem (average-cost, discounted cost) can be considered. Note that causality is built in here and the notion of information available at time t, namely, the σ-field generated by $(Y_0, \cdots, Y_{t-1})$ is an important quantity.

The solution to this problem is well-known and leads to the so-called separation theorem, where the optimal choice of control and the optimal choice of the state-estimator can be separated. The most important notion in here is that of the information state

$$P^u(X(t)|\, Y(0), \cdots, Y(t-1))$$

which depends on the choice of control. This dependence can be suppressed by using the innovations process. It is not hard to see that P^u is conditionally Gaussian and hence can be dynamically described by describing the evolution of the conditional mean and the conditional covariance. This leads to the Kalman filter. The control turns out to be a function of the information state (in this case only the conditional mean, since the conditional covariance is deterministic). This idea of separation between control and estimation is quite general.

A number of variants of this problem is possible. In [10] I have considered the effects of source coding delay where the source is the state of a stable system. It turns out that if instead of transmitting the coded source, if one transmits the coded "innovations" even over a noisy channel, the problem of optimized coding separates from that of optimal control. This model captures in a precise way that shorter codes correspond to low resolution and low-delays, while longer codes mean higher resolution but longer delays. We have also suggested here how the notion of successive refinement of partitions leading to tree codes is a natural one to consider here.

A more sophisticated version of the infinte-time average cost LQG problem where a communication channel is inserted between the sensor and the controller has been considered in [7],[11]. The channels that can be accomodated are digital hard rate R noiseless channel, digital average R noiseless channel, digital erasure channel and the additive white Gaussian channel with a power constraint; where the rate is $R = \frac{1}{2}\log_2(1+P)$ where P is the input power limit.

The goal is to understand the relationship between control objectives, plant dynamics and the use of the communication channel. In general, the encoder and decoder will both have internal states. Under the assumption of equi-memory, the encoder computes the next channel input symbol based only on the current state observation and the *decoder's* state. Equi-memory requires either:

1. Noiseless forward channel.
2. Explicit and noiseless feedback from the decoder to the encoder.
3. Encoder has access to the control signal and is able to "invert" that signal to recover decoder state.

Given equi-memory, the control problem separates into an optimal state-estimator (encoder/decoder) and the certainty equivalent controller. This holds regardless of the nature of the communication channel. This control problem can be solved using standard Dynamic Programming arguments. So, the *key to the problem is encoder/decoder design*.

Sequential rate distribution theory (see [15],[16] and the references cited there) has a key role to play in this problem. In our situation, we consider the process

$$X(k+1) = AX(k) + W(k)$$

with $(W(k))$ white Gaussian noise, the sequential rate distribution problem is defined as follows:

$$D_{N,Seq.}(R,M) = Inf_{\ P(\hat{X}_1^N|X_1^N)}$$
$$\frac{1}{N}E\left(\sum_{k=1}^{N}[(X(k)-\hat{X}(k), M(X(k)-\hat{X}(k)))]\right)$$

where M is positive definite, subject to the rate constraint

$$\frac{1}{N}I(\hat{X}_1^N; X_1^N) \le R$$

where the minimization is carried out over all $P(\hat{X}_1^N|X_1^N)$ which are causal (in the obvious sense).

The rate distortion function is

$$D_{Seq.}(R,M) = \lim_{N\to\infty} D_{N,Seq.}(R,M).$$

For simplicity, consider the scalar case. A surprising result is that there is a minimum rate, $R > \log_2 A$, required to stabilize the system. In this case

$$D_{\text{Seq.}}(R,M)=\frac{M\Sigma_W}{2^{2R}-A^2}$$

where Σ_W is the variance of W. For special channels, including the AWGN with equi-memory, $D_{\text{Seq.}}(R,M)$ can be achieved. The structure of the minimizing conditional law suggests that the optimal structure of the encoder is predictive.

It turns out that here also the optimal average cost separates into a sum of two costs, precisely as in the partially observed LQG problem: a cost due to control and a cost which comes from the estimation error. This second cost can be explicitly computed using sequential rate distortion theory in the scalar case. In the general case a lower bound is obtained for the average cost (see [7]).

Equi-memory is a strong condition and generally requires at least one noiseles link. One needs a way for the decoder to communicate to the encoder. Now, in a general way, we can consider the plant as a channel. Consider the scalar case and change the quadratic cost on U_k to a hard constraint: $E(U_k^2)\le P_2$. Assume that the encoder has noisy observations of U_k. Let $R_2=\frac{1}{2}\log_2(1+\frac{B^2P_2}{K_W})$. A necessary condition for well-posedness is that $R_2\ge\log_2 A$.

Returning to the original average cost problem, if there is no cost on control, then the equi-memory assumption can be dispensed with. Otherwise, there is a fundamental tradeoff between control energy and capacity required from the encoder to the decoder. Sub-optimal schemes which are optimal in the high rate regime can be designed when the equi-memory assumption cannot be justified.

A more general view of this problem where the state process is a controlled Markov Chain has been considered in [13]. In other work, we have shown how the optimal sequential quantization of Markov sources can be viewed as a partially observed stochastic control problem [12].

Toward a Dynamical View of Information Theory

The discussion in the previous section raises a new problem in Information Theory:

> How can one reliably transmit an unstable source over a noisy channel through appropriate source and channel coding and decoding at the receiving end?

More precisely, given a scalar discrete-time finite-state Markov source $(X(t))$ given by

$$X(t+1)=aX(t)+W(t),$$
$$a>1,\ t=0,1,\ldots$$

and $(W(t))_{t\ge 0}$ is additive white Gaussian noise or bounded noise with finite support and a memoryless channel (or additive white Gaussian noise channel) is it possible to design encoders and decoders within a specified finite end to end delay constraint so that the output of the decoder $(\hat{X}(t))$ achieves a desired mean-squared performance $\text{Sup}_{t\ge 0}E(X(t)-\hat{X}(t))^2\le D$?

A similar question was posed by Berger [15] for the Wiener process and an information transmission theorem for this case has been an open problem for many years. In Anant Sahai's thesis [8] a solution to this problem is presented. The solution requires a dynamical view of Information Theory, since the message, the Markov Source, is not given at time –1, but unfolds in time and a little thought will make it clear that block coding of any kind will not work in this situation. Indeed, all coding and decoding operations must be causal, causality suitably defined. In this problem, the separation of source and channel coding is no longer obvious and separation has to be inposed by a new definition of channel capacity

$$C_{anytime}(\alpha)=$$
$$\text{Sup.}\{R|\exists(K>0,\ \text{Rate}(\varepsilon,D^a)=R)$$
$$\forall d>0\,P_{\text{error}}(\varepsilon,D^a,d)\le K\cdot 2^{-\alpha D}\}$$

In the above ε denotes the encoder, D^a the anytime decoder and d the prescribed delay. It has been argued by Sahai that α should be thought of as a quality of service parameter.

This definition should be contrasted with the classical operational definition of capacity. The exponent α is related to the error exponents corresponding to block coding and convolutional coding. Appropriate source and channel coding theorems for this problem with the above definition of capacity are proved in [8]. I want to emphasize that the total end to end distortion problem has to be considered for this situation and the dynamical view which I have referred to is an essential element in the solution to this problem.

The discussions in the previous section and this section provides evidence as to why the sequential (zero delay) rate distortion problem is an important problem. Although it would be too much to expect that a Shannon like rate distortion theorem would be true in this causal situation (see [16], for example), it is still important to characterize in a precise way the gap betwen the non-causal rate distortion function and the causal rate distortion function. This has been carried out in [17].

On Information Structures: Witsenhausen Revisited

In several important papers Witsenhausen (see [6],[18]) elucidated the role of information structures, namely what information is available when and where, on performance of stochastic control problems and the complexity of the optimal solution. In a centralized totally synchronous system where the global state is available to the controller, the optimal stochastic control problem leads to a dynamic programming equation in an abstract form (see [19]):

$$V(i,x) = \text{Inf.}_{u \in U}$$
$$\left\{ g_i(x,u) + \int_X V(i+1,\xi) P_i(x, d\xi|u) \right\}$$

here $x \in X$ is the state space, $u \in U$ is the control space. When the global state is not available to a centralized controller but control is distributed, one can still do Dynamic Programming but an a priori convex problem may become nonconvex (for an elucidation of this point, see thesis of Sekhar Tatikonda [7]).

To understand the difficulty, let us consider the following problem after Witsenhausen:

Let X_1 and X_2 be discrete random variables. Assume that the first agent observes

$$Y_1(w) = X_1(w) = y_1 \quad (2)$$

and takes as an action

$$Y_1(X_1(w)) = u_1. \quad (3)$$

The second agent observes

$$Y_2(u_1, X_2(w)) = \mathcal{Y}_2 \quad (4)$$

and takes an action

$$\gamma_2(Y_2(w) = \mathcal{Y}_2) = u_2 \quad (5)$$

The problem is to choose γ_1 or γ_2 to minimize

$$E_{P_{X_1,X_2}} J(X_1, U_2) \quad (6)$$

Let J^* denote the minimum. Witsenhausen clearly recognized the communication aspects of this problem, namely, Agent 2 requires information about X_1 but can only get it through u_1 via the channel

$$(x_2, u_1) \mapsto P(Y_2|x_2, u_1).$$

This is a highly non-convex problem but it is natural to try to obtain a lower bound for J^* using an idea, apparently due to Shannon (see [20]).

Step 1

Compute the channel capacity

$$Max_{P_{U_1}} I(U_1; Y_2) = C\cdot$$

Step 2 Embedding
Minimize

$$E_{PX_1} E_{PU_2|x_1} J(X_1; U_2)$$

subject to the constraints

$$I(X_1; U_2) \le C$$
$$P_{U_2|X_1} \ge 0, \int P(U_2|X_1) du_2 = 1.$$

It is important to note that both these problems are convex programming problems. If D^* is the solution to the above minimization problem then

$$J^* \ge D^*.$$

This result follows from the fact that

$$X_1 \to U_1 = \gamma_1(X_1) \to Y_2 = U_2 = \gamma_2(U_1, X_2)$$

forms a Markov Chain.

Unfortunately, this bound is ineffective unless the minimizing channel factors in such a way that the original channel is one of the factors. One can exhibit examples to demonstrate that this problem falls into the class of NP-complete problems. Furthermore in specific situations the two convex programming problems can be explicitly solved and it turns out that the lower bound is totally ineffective. Obtaining good performance lower bounds for problems of distributed control and distributed communications is an open problem.

Towards a Theory of Feedback for Communication and Control

In recent work of Willems [2] he has outlined a theory of control where control is viewed as the interconnection of two behavioural systems . I want to suggest a stochastic analogue of this. The idea of interconnecting a message to a source and interconnecting two random variables goes back to Dobrushin [21]. In unpublished work [22], I have described the notion of a stochastic system and interconnection of stochastic systems which is similar in spirit to the ideas of Dobrushin. A more detailed investigation of these ideas have been undertaken in the doctoral thesis of Tatikonda referred to before.

The basic mathematical model of a system is a family of probability measures

$$\mathcal{M} \subseteq \mathcal{P}(Z_1, \ldots, Z_2)$$

where $\mathcal{P}(Z_1, \ldots, Z_2)$ is the space of all probability measures on the variables of interest $(Z_1, \ldots, Z_2)$. $\mathcal{M}$ represents a complete specification of the system. Plants, coders, channels, decoders are given by stochastic kernels which represent partial specifications of the joint measure. The choice of controllers, coders, decoders allow us to "complete" the joint measure. Note that controllers, coders, decoders are themselves stochastic kernels.

In a typical control situation we are given stochastic kernels $P(X_{t+1}|x_t, u_t)_{t=1,\ldots T}$ representing state transition maps corresponding to controls $(u_t)t = 1, \ldots, T\cdot$. The complete specification is the joint measure factored according to the causal ordering

$$X_1 = x_1, U_1 = u_1, X_2 = x_2, \ldots, X_T = x_T,$$

as

$$P(dX^T, dU^T) = \prod_{t=1}^{T} P(dU_t | x^t, u^{t-1})$$
$$\otimes P(dX_t | x^{t-1}, u^{t-1}).$$

Specifiying the controllers $P(dU_t | x^t, u^{t-1})$ chosen according to some criterion allows us to "complete" the measure so that the joint measure is an element of $\mathcal{M}$. The nature of the dependence of the control on the past data specifies the information structure. This viewpoint, coupled with Shannon's idea of code functions (as opposed to code words) and Massey's idea of Directed Mutual Information [23] allows one to obtain the analogue of the Noisy Channel Coding Theorem for Feedback Channels.

GOLOMB'S PUZZLE COLUMN™ CYCLES OF PERMUTATIONS SOLUTIONS

"How many of the $n!$ permutations on n symbols have a longest cycle of length k, when $\frac{n}{2} < k \le n$?"

Since $k > \frac{n}{2}$, a cycle of length k, if it occurs, is unique. There are $\binom{n}{k}$ ways to select k symbols to be on the cycle of length k, $(k-1)!$ different cyclic orders for these k symbols to be on a cycle of length k, and $(n-k)!$ ways to permute the remaining $n-k$ symbols. Thus, the number of permutations whose longest cycle has length $k > \frac{n}{2}$ is $\binom{n}{k}(k-1)!(n-k)! = \frac{n!}{k}$.

2. "Show that the sequence $\lambda_n = \frac{L_n}{n+1}$ is monotonically increasing for $n \ge 2$, where L_n is the expected length of the longest cycle in a random permutation on n symbols."

Proof. When an $(n+1)^{st}$ symbol is added to the set, which is then "shuffled" and "dealt", the probability that the new symbol lands in the longest cycle exceeds $\frac{L_n}{n}$ for $n>1$, because sometimes two or more cycles are tied for longest. Thus $L_{n+1} > L_n + 1 \cdot \frac{L_n}{n} = L_n\left(\frac{n+1}{n}\right)$, from which $\frac{L_{n+1}}{n+1} > \frac{L_n}{n}$ for all $n \ge 2$.

3. "Show that $\frac{L_n}{n}$ is a monotonically decreasing sequence, for $n \ge L$"

Proof. If we remove a symbol "at random" from a random permutation on n symbols, it had probability $> \frac{L_n}{n}$ of having been on the longest cycle. Thus $L_{n-1} < L_n - \frac{L_n}{n} = L_n\left(\frac{n-1}{n}\right)$, and $\frac{L_{n-1}}{n-1} < \frac{L_n}{n}$ for all $n \ge 1$

4. Since $\lambda_n = \frac{L_n}{n+1}$ is monotonically increasing and it is clearly bounded from above. by 1, by a theorem of Weierstrass it must have a limit λ. Since $l_n = \frac{L_n}{n} = \frac{n+1}{n} \cdot \frac{L_n}{n+1} = \left(\frac{n+1}{n}\right)\lambda_n$, and since $\lim_{n\to\infty} \frac{n+1}{n} = 1$, we have $\lim_{n\to\infty} l_n = \lim_{n\to\infty}\left(\frac{n+1}{n}\right)\lambda_n = 1 \cdot \lambda = \lambda$.

Note: λ=0.62432965...has been named "Golomb's constant" by Donald Knuth.

5. "Among the $n \cdot n!$ symbols in the set of all $n!$ permutations on n distinct symbols, how many are found in cycles of length k?"

Answer: For each k, $1 \le k \le n$, the answer is "$n!$ symbols." By problem 1, for $k > \frac{n}{2}$ there are $\frac{n!}{k}$ permutations with cycles of length k, and therefore $k \cdot \left(\frac{n!}{k}\right) = n!$ *symbols* (among the $n \cdot n!$ symbols altogether) on cycles of length k. But the restriction to $k > \frac{n}{2}$ is removable, as follows: To form a cycle of length k (for *any* k, $1 \le k \le n$), there are $\binom{n}{k}$ ways to pick the k symbols to be on the cycle, $(k-1)!$ ways to arrange these k symbols cyclically, and $(n-k)!$ ways to permute the remaining $n-k$ symbols; so these choices contribute $k\binom{n}{k}(k-1)!(n-k)! = n!$ symbols altogether on cycles of length k.

6. "Let p_n be the probability that the symbol '1' is on the longest cycle in a 'random' permutation on $A_n = \{1, 2, 3, \ldots, n\}$, Find $\lim_{n\to\infty} p_n$.

Answer: $\lim_{n\to\infty} p_n = \lambda$

Proof: The probability that any *specific* symbol is on the longest cycle in a random permutation on A_n is asymptotically $l_n = \frac{L_n}{n}$, and therefore $\lim_{n\to\infty} p_n = \lim_{n\to\infty} l_n = \lambda$.

For more details, more rigorous proofs, and related information, see:

References

[1] *Shift Register Sequences*, S. W. Golomb, Holden-Day, Inc., 1957, San Francisco; Second Edition: Aegean Park Press, 1982, Laguna Hills, CA, especially pp. 178-192.

[2] "On the number of permutations on n objects with greatest cycle length k", S. W. Golomb and P. Caal, *Advances in Applied Math*, vol. 20, 1998. pp. 98–107.

Interestingly, for Markov Channels, determining the capacity of a feedback channel leads to the maximization of Directed Mutual Information which has the formulation of a Stochastic Control Problem. For a detailed discussion of the above see [7] and the references cited there.

Conclusions

In this lecture, I have tried to demonstrate that problems of distributed control where plants, sensors, controllers and actuators are linked by communication channels lead to new problems in systems theory and information theory. A dynamical view of information theory might well be important in communication problems where delays cannot be ignored. Feedback in communications channels is best viewed from a Systems Theory perspective. Much work remains to be done and I hope Information Theorists would become interested in some of the questions raised here.

References

[1] W.M. Wonham: *Linear Multivariable Control*, Springer-Verlag, New York, Berlin.

[2] J.W. Polderman and J.C. Willems: *Introduction to Mathematical Systems Theory*, Springer, 1997.

[3] G.D. Forney and M.D. Trott: The Dynamics of Group Codes: State Spaces, Trellis Diagrams and Canonical Encoders, *IEEE Trans. on Inf. Th., 39*, 1491-1513, 1993.

[4] J. Galdos: Information and Distortion in Filtering Theory, Ph.D. Thesis, MIT, 1975.

[5] M. Zakai and J. Ziv: Lower and Upper bounds on Optimal Filtering Errors for Certain Diffusion Processes, *IEEE Inf. The. Trans.*, IT-18,1972.

[6] H.S. Witsenhausen: Separation of Estimation and Control for Discrete-time Systems, *Proc. I.E.E.E.*, 59(11), pp. 1557-1566, 1971.

[7] S. Tatikonda: Control under Communication Constraints, Doctoral Thesis, M.I.T., 2000.

[8] A. Sahai: Any-time Information Theory, Doctoral Thesis, M.I.T., 2000.

[9] S.K. Mitter and A. Sahai: Information and Control: Witesenhausen Revisited. Lecture Notes in Control and Information Sciences, 241, Springer 1998.

[10] V.S. Borkar and S.K. Mitter: LQG Control with Communications Contraints, in *Communications, Control, Computation and Signal Processing*, Kailath Festschrift, Kluwer Academic Publishers, Boston, 1997.

[11] S. Tatikonda, A. Sahai and S.K. Mitter: Control of LQG Systems under Communications Constraints, forthcoming. See also *Proc. ACC*, 1999, and *Proc. CDC*, 1998.

[12] V.S. Borkar, S.K. Mitter and S. Tatikonda: Optimal Sequential Vector Quantization of Markov Sources, revised version under consideration, *SIAM J. on Control and Optimization*.

[13] V.S. Borkar, S.K. Mitter and S. Tatikonda: Markov Control Problems with Communications Constraints, to appear, *Comm. in Information and Systems*.

[14] N. Elia and S.K. Mitter: Stabilization of Linear Systems with Quantized Controls, to appear, *IEEE Trans. on Automatic Control*.

[15] T. Berger: *Rate Distortion Theory*, Prentice Hall, Englewood Cliffs, N.J., 1971.

[16] D.L. Neuhoff and R.K. Gilbert: Causal Source Codes, *IEEE Trans. of Inf. Th.*, 28(5), pp. 701-713, 1982.

[17] V.S. Borkar, S.K. Mitter, A. Sahai and S. Tatikonda: Sequential Source Coding: An Optimization Viewpoint, pre-print, LIDS, M.I.T., 1998.

[18] H.S. Witsenhausen: A Counter-example in Stochastic Optimal Control, *SIAM J. on Control*, 6(1), pp. 131-147, 1968.

[19] D.P. Bertsekas and S. Shreve: *Stochastic Optimal Control: the Discrete-time Case*, NewYork, Academic Press, 1978.

[20] S.K. Mitter: Lecture at Allerton Conference, 1998.

[21] R.L. Dobrushin: A General Formulation of the Basic Shannon Theorem in Information Theory, *Uspekhi Math. Nauk. USSR*, 14(6), pp. 3-103, 1959.

[22] S.K. Mitter: Lecture at IMA Workshop on Codes, Graphs and Systems, August 1999.

[23] J.L. Massey: Causality, Feedback and Directed Information, Proc. ISITA-90, pp. 303-305.

SYMPOSIUM REPORT

23rd Symposium on Information Theory and its Applications (SITA2000)

October 10-13, 2000
Aso, Kumamoto, Japan

The 23rd Symposium on Information Theory and its Applications (SITA2000) took place from October 10 to October 13, 2000 in Aso, Kumamoto, Japan. The symposium was sponsored by the Society of Information Theory and its Applications (SITA) and co-sponsored by the Fundamentals on Information and Communication Sub-Society of the Institute of Electronics, Information and Communication Engineers (IEICE) and the IEEE Information Theory Society Japan Chapter.

Prof. Kyoki Imamura, Kyushu Institute of Technology, was the general chairman and Prof. Yasutada Oohama, Kyushu University, served as the chair of technical program committee. The majority of participants was from Japan but there were also papers from Germany, Korea, Saudi Arabia and China.

All 170 papers presented at SITA2000 are published in the 680 pages of the Proceedings of the 23rd Symposium on Information Theory and Its Applications. The number of participants was 270. Usually SITA has been held in late November or in early De-