SYSTEMS THEORY
A Retrospective and Prospective Look

Sanjoy K. Mitter

Laboratory for Information and Decision Systems

Massachusetts Institute of Technology

November 14, 2009

“What can be said at all can be said clearly; and whereof one cannot speak thereof one must be silent.”

Ludwig Wittgenstein,

Tractatus Logico-Philosophicus

Agenda of Systems Theory

- Models and their Structure
- Fundamental Limitations (Laws)
- Uncertainty and Robustness

  Robustness of performance uncertainty at different levels of granularity
- Interconnections, Architecture and Algorithms

  Architecture = organization of distributed algorithms and their implementation in hardware

Agenda of Systems Theory (cont.)

- Resource Management (Energy, Time, Space, . . .)

  A broad vision of Systems Theory aids in providing a unified conceptual framework for problems in different fields (Control, Communication, Signal Processing, Operations Research)

- Structure

- Action

- and their Interaction

History of Science in the Sense of Kuhn:

Incommensurability

Thomas Kuhn in his book *The Structure of Scientific Revolutions* distinguished between Normal Science and Revolutionary Science.

Revolutionary Science (e.g., Quantum Mechanics) arises when:

Existing Theories fail to explain phenomena

A new "paradigm" is needed to reconcile theory and experiment

With the new paradigm, a new language is needed

Something like that happened in the late fifties and early sixties in the Systems and Control field.

Earlier revolution (1948):
Shannon Information Theory and
Invention of the Transistor

“The Double Big Bang,” to quote Viterbi

I want to suggest that in the Systems and Control field, there was a crisis in the field in the fifties. Let me suggest as pointers three manifestations of that crises.

1. Internal Stability: Feedback Control Systems designed from an external (input/output) point of view failed to recognize the presence of these internal instabilities.

2. The approach to design of multi-input/multi-output systems was essentially a reduction to a single-input/single-output system through a decoupling procedure.

3. The attempts to deal with the Wiener filtering problem in the nonstationary situation (Zadeh–Regazzini) leading to some analog of the Wiener–Hopf equation was not very successful (no procedure analogous to Spectral Factorization was available).

It is also worth mentioning that the Mathematics that was prevalent in Linear Systems Theory at the time was Complex Function Theory and Transform Theory.

New Element

Computation and the Concept of a Solution

Solution not necessarily an analytical expression

Theories leading to Algorithms

Advent of State Space Theory

(New Paradigm)

- New Language: Algebra, Differential Equations
- Concept of State
- State Space Representation=

$$\begin{cases} \frac{\mathrm{d}x}{\mathrm{d}t} &= Fx(t) + Gu(t) \\ y(t) &= Hx(t) \end{cases}$$

$u =$ input, $x =$ state, $y =$ output

Extends to time-varying and nonlinear systems

Advent of State Space Theory

(New Paradigm cont.)

$$y(t) = He^{(t-t_0)F}x(t_0) + \int_{t_0}^{t} He^{(t-s)F}Gu(s)\mathrm{d}s$$

Reconciliation of Input-Output and Internal (State) Point-of-view through introduction of concepts of reachability (controllability) and observability

Natural Connection to Stability and Optimality (Calculus of Variations)

Minimize

$$J(u,x) = \int_{t_0}^{t_1} [(x(t), Qx(t)) + (u(t), Ru(t))] \mathrm{d}t$$

$$Q \geq 0 \quad , \qquad R > 0$$

Behavior of optimal control

$$u(t) = K(t)x(t) \quad \text{as} \quad t_1 \to \infty$$

Role of Controllability and Observability

Deeper Aspects of Structure

Actions of semi-direct product

$$GL(n) \times \mathcal{F} \times GL(m)$$

on (F, G) controllable

$$(F, G) \mapsto (T^{-1}(F + GK)T, GL)$$

Kronecker Invariants

Transporting the algebraic variety structure of (F, G) to the quotient

Implications in System Identification

How should we think about Graphs beyond thinking about them as (V, E)?

How should we think about Systems of Coupled Differential Equations evolving over Graphs?

What are these invariants?

We should be able to distinguish between differential equations evolving over trees from differential equations evolving over graphs with loops

We need Canonical Problems

Pattern Recognition (Vision)

"Tranformation Group" acting on the space of objects is not given but needs to be identified!!

See the section on Pattern Recognition in Minsky's paper:

"Steps Towards Artificial Intelligence," *Proc. IEEE*, 1961.

Influence of Systems Theory in Coding Theory and Signal Processing

(Intersection with Behavioral View of Systems: Willems)

Linear Systems taking values in Finite Groups (Forney–Trott)

Minimality, Controllability and Observality, Duality in Signal Processing

State Space Viewpoint: Influence on Algorithms exploiting structure

Adaptive Filtering

Filtering and Stochastic Control:

Separation Principle

$$\begin{cases} \mathrm{d}X(t) &= FX(t)\mathrm{d}t + Gu(t) + J\mathrm{d}W(t) \\ \mathrm{d}Y(t) &= Hx(t)\mathrm{d} + \mathrm{d}V(t) \end{cases}$$

Choose $u(t) = \varphi(\Pi_t Y)$ to minimize

$$J(u,x) = \mathbb{E}\left[\int_{t_0}^{t_1} [(X(t), QX(t)) + (u(t), Ru(t))]\mathrm{d}t\right]$$

Solution

$$u^*(t) = K(t)\hat{X}(t)$$
$$\hat{X}(t) = \mathbb{E}(X(t)|\mathcal{F}_t^Y)$$

Separation into estimation and deterministic control

- Infinite-time

(Controllability, Observability, Stability)

- Non-linear

Smoothing (Decoding)

Compute: $\mathbb{P}(X_s, t_0 \leq s \leq t_1)|\mathcal{F}_{t_1}^Y)$

Bayesian Inference and Statistical Mechanics

Estimate U from Y (Observations), both belonging to separable metric spaces

Given P_{UY} with marginal P_U and P_Y,

Let $Q(u,y)$ be the Likelihood Function and $H(u,y) = -\log Q(u,y)$ Hamiltonian

Let

$$
\begin{aligned}
h(\tilde{P}|P_U) &:= \int_{\mathbb{U}} \log\left(\frac{d\tilde{P}}{dP_U}(u)\right)\tilde{P}(du) \text{ if } \tilde{P} \ll P_U \\
&\quad\ +\infty \qquad\qquad\qquad\qquad \text{otherwise,} \\
\langle \tilde{H}, \tilde{P}\rangle &:= \int_{\mathbb{U}} \tilde{H}(u)\tilde{P}(du) \text{ if the integral exists} \\
&\quad\ +\infty \qquad\qquad\qquad\qquad \text{otherwise,} \\
i(\tilde{H}) &:= -\log \int_{\mathbb{U}} \exp(-\tilde{H}(u))P_U(du) \\
&\qquad\qquad\qquad\qquad \text{if the integral is nonzero} \\
&\quad\ -\infty \qquad\qquad\qquad\qquad \text{otherwise,}
\end{aligned}
$$

The following proposition characterizes $h(P_{U|Y}(\,\cdot\,,y)|P_U)$ in terms of $i(H(\,\cdot\,,y))$ and vice versa

Proposition 1.1.

(i) $$i(H(\,\cdot\,,y)) = \min_{\tilde{P}\in\mathcal{P}(\mathcal{U})}\{h(\tilde{P}|P_U) + \langle H(\,\cdot\,,y),\tilde{P}\rangle\};$$

(ii)

$$h(P_{U|Y}(\,\cdot\,,y)|P_U) = \max_{\tilde{H}\in\mathcal{M}(\mathbb{U})}\{i(\tilde{H}) - \langle \tilde{H}, P_{U|Y}(\,\cdot\,,y)\rangle\};$$

(iii) $P_{U|Y}(\,\cdot\,,y)$ is the unique minimizer;

(iv) If $\hat{H}$ is a maximizer, then there exists a real constant c such that $\mathbb{P}(\hat{H}(U) = H(U,y) + c) = 1$.

Bayesian Inference (cont.)

- Filtering, Smoothing has interpretation as Free Energy Minimization
- Information-theoretic Interpretation
- Analogue of Dissipation Inequality, in Information Quantities
- Proof of the Noisy Channel Coding Theorem as a Limiting Bayesian Inference Problem
  (Gibbs Variational Principle: Characterization of Translation Invariance Gibbs Measures as Minimization of Specific Free Energy)

(See: "Variational Bayes and the Noisy Channel Coding Theorem," Parts 1 and 2
Newton and Mitter

Uncertainty and Robustness

Process and Measurement Uncertainty

vs.

Model Uncertainty

Approximation of Input-Output Maps

vs.

Approximation at the State Space Representation

Two input-output maps may be close to each other but the dimensions of their corresponding state spaces may be far apart

(See: "The Legacy of George Zames," Mitter and Tannenbaum, *IEEE Trans. on Auto. Control*)

Fundamental Problem of Control: Design of Control Systems whose performance is robust against uncertainties

For linear time-invariant, bounded, causal maps from $L^2(\mathbb{R}) \to L^2(\mathbb{R})$, which, from the Segal–Foures theorem, is in one-to-one correspondence with operators which are multiplication operators by H^∞-functions

Uncertainty in model represented by a ball in H^∞

Feedback: reduction of complexity

Deep connections to Operator Theory, in particular the work of Krein

Recent work of Y.H. Kim:

Feedback Capacity of Stationary Gaussian Channels

The computation of feedback capacity is posed as an Infinite Dimensional Variational Problem and uses Systems Theory for its solution

Interestingly, Keynes viewed the representation of “uncertainty” and how to deal with uncertainty as one of the fundamental problems of Macroeconomics

He also questioned the use of probability for certain uncertain situations (prospect of a European war is uncertain, the price of copper, rate of interest twenty years hence)

Indeed, for systems which are distributed, modeling and representation of uncertainty remains a fundamental issue

Towards a Unified View
of
Communication and Control

Feedback communication problem

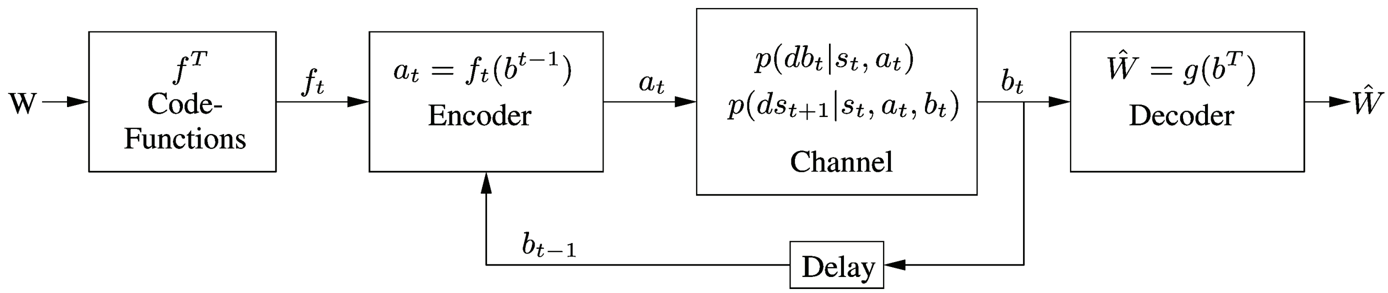


Figure 1. Interconnection

Choose encoder and decoder to transmit message over the channel to minimize the probability of error

Channel at time t: $P(db_t|a^t, b^{t-1})$ stochastic kernel

$$a^t = (a_0, \dots, a_t)$$

$$\textit{Channel} = \text{Sequence of } \left. P(db_t|a^t, b^{t-1})\right|_{t=1}^{t}$$

Time ordering: Message $= W, A_1, B_1, \quad , A_T, B_T, \hat{W} =$ Decoded message

$$W = (1, 2, \dots, M)$$

Code function:

$$\mathcal{F}_t = \{f_t : B^{t-1} \to A : \text{measurable}\}$$
$$\mathcal{F}_T = \prod_{t=1}^{T} \mathcal{F}_t$$

Channel code function: $f^T = (f_1, \ldots, f_t)$

Distribution on code functions:
$P(df_t|f^{t-1})\Big|_{t=1}^{T}$

Channel code = list of M channel code functions

Code functions are introduced to reduce the feedback communication problem to a no feedback communication problem.

Average Measure of Dependence

Mutual Information

$$
\begin{aligned}
I(A^T; B^T) &= \mathbb{E}_{P_{A^T,B^T}} \log \left(\frac{P_{A^T,B^T}}{P_{A^T} P_{B^T}} \right) \\
&= \mathbb{E}_{P_{A^T,B^T}} \log \left(\frac{P_{B^T|A^T}}{P_{B^T}} \right)
\end{aligned}
$$

$$
I(A^T; B^T) = \sum_{t=1}^{T} I(A^T; B_t | B^{t-1})
$$

Information transmitted to the receiver depends on future $(A_{t+1}, \ldots, A_T)$.

Directed Mutual Information (Causal)

$$
I(A^T \to B^T) = \sum_{t=1}^{T} I(A^t; B_t | B^{t-1})
$$

To compute Mutual Information (Directed Mutual Information), need joint distribution

$$\mathbb{P}_{A^T,B^T}(da^T, db^T)$$

This can be done if we are given the channel

$$P(db^t|a^t, b^{t-1})\Big|_{t=1}^{T}$$

and channel input distributions

$$\mathcal{D}_t := \mathbb{P}(da_t|a^{t-1}, b^{t-1})\Big|_{t=1}^{T}$$

Interconnection of channel input to channel

Channel Capacity

$$C_T = \sup_{\mathcal{D}_T} \frac{1}{T} I(A^T \to B^T)$$

(Note: Optimization over original input codes, not on space of code functions.)

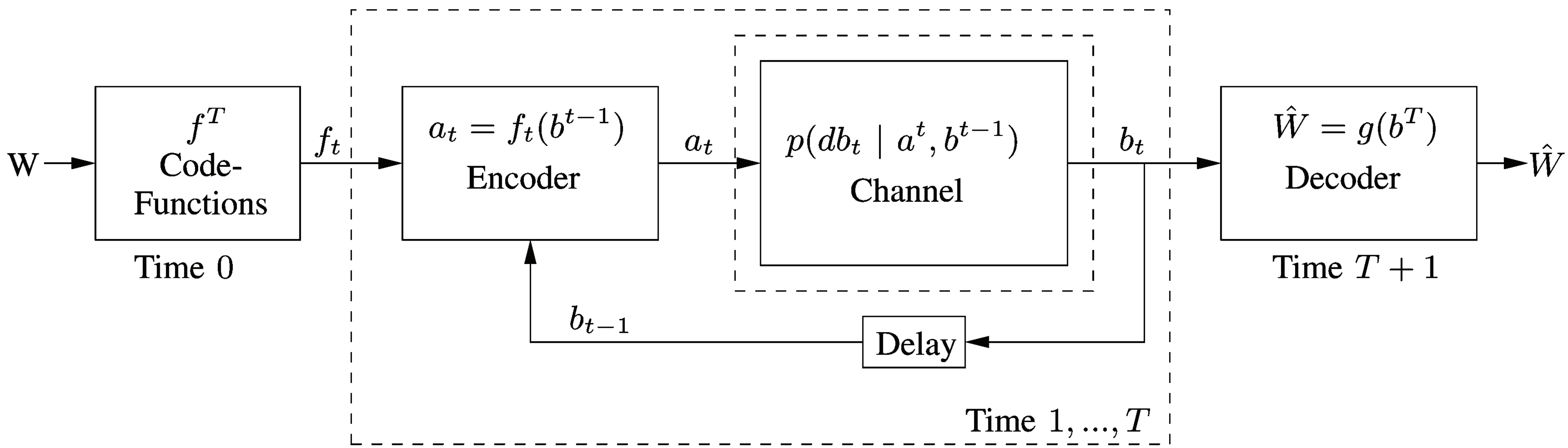


Figure 2: Markov Channels

Markov Channel

$$P(ds_{t+1}|s_t, a_t, b_t)\Big|_{t=1}^{T} \quad : \quad \text{state transition}$$

$$P(db_t|s_t, a_t)\Big|_{t=1}^{T} \quad : \quad \text{channel output}$$

Capacity of Markov Channels

$$(1) \qquad \sup_{\mathcal{D}_\infty} \lim_{T\to\infty} \frac{1}{T} I(A^T \to B^T)$$

It turns out that by appropriately defining sufficient statistics (π_t) (conditional distributions of the state given information from encoder to decoder) and controls $u_t(da_t|\pi_t)$, and state $X_t = (\pi_{t-1}, A_{t-1}, B_{t-1})$ and instantaneous cost $c(x_t, u_t, u_{t+1})$, (1) can be formulated as a Partially Observed Stochastic Control Problem.

In turn, this can be reformulated as a fully-observable stochastic control problem.

This problem is more like a *dual* control problem since the choice of the channel input can help the decoder identify the channel.

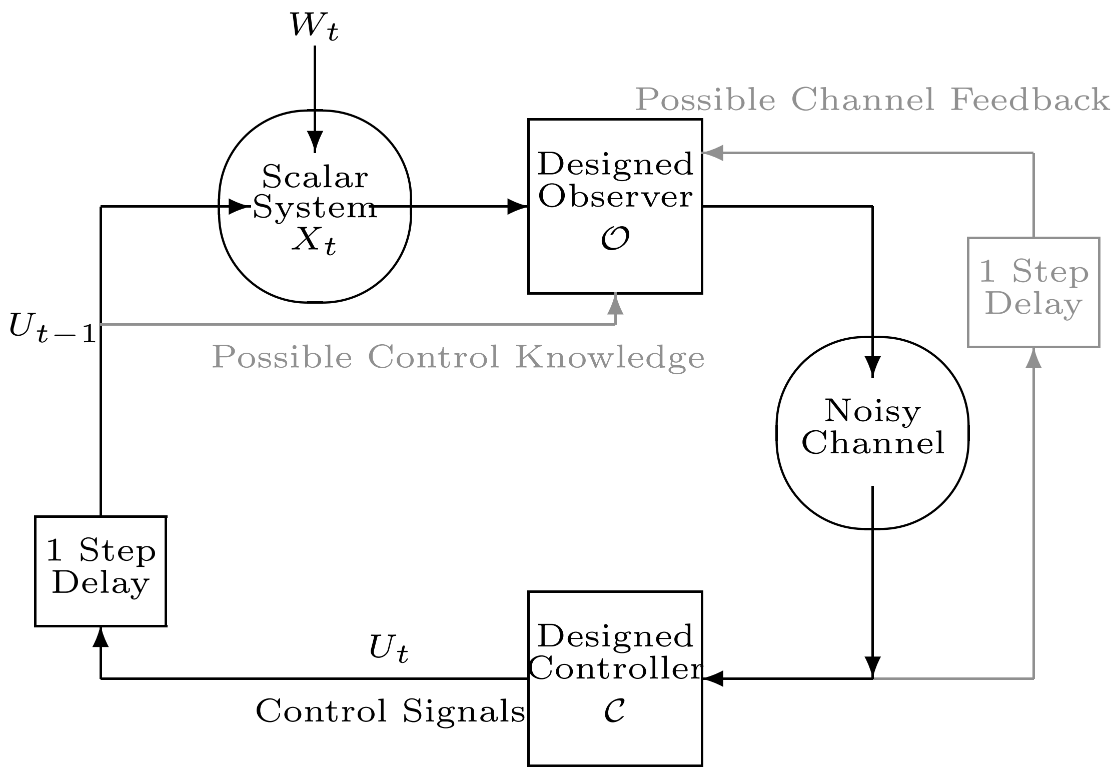
This is also an example where the *information pattern is nested*: The encoder has more information than the decoder.

Communication

and

Control

A simple scalar distributed control problem



$$X_{t+1} = \lambda X_t + U_t + W_t$$

- Unstable $\lambda > 1$, bounded initial condition and disturbance W.
- Goal: Stability $= \sup_{t>0} E[|X_t|^\eta] \le K$ for some $K < \infty$.

Stabilization equivalent to reliable
Communication through the loop

Signaling through the loop

Open Problem

Existence of Channel Linking

Controller and Actuator

Asymmetry in Information Transfer

Problems for the Future

- Distributed Estimation and Control

Signalling: Controllers, Estimators have to communicate their actions (estimates) through the plant. There is a role for Information Theory here.

(See recent work of Sahai on Witsenhausen problem)

See: **Michael Spence (Nobel lecture)**

Signalling in Retrospect and the Information Structure of Markets

- Games as Multiple Feedback Loops

(Witsenhausen)

Related to Distributed Control

Problems for the Future (cont.)

• Connections to Statistical Mechanics and Field Theory

Information Theory of Message Passing Algorithms

(See for example: Cramer's Rule and Loop Ensembles: A. Abdesselam and D.C. Brydges)

• Interconnections and Interactions

Optimal Transportation Theory

• What is the Nature of Experimental Work in our Field?

Theory vs. Experiment

Problems for the Future (cont.)

- Systems View (Dynamical) of Economic

Classifying Equilibria

(See: Global Trade and Conflicting National Interests: Ralph E. Gomory and William J. Baumol, MIT)

Concluding Remarks