The Role of Information Theory in Communication Constrained Control Systems

Sekhar C. Tatikonda
Yale University

Workshop on Networked Embedded Sensing and Control
October 17-18, 2005

Acknowledgements

- Vivek Borkar
- Nicola Elia
- Daniel Liberzon
- Sanjoy Mitter
- Anant Sahai

$\rightarrow$ A few qualifications…

Introduction

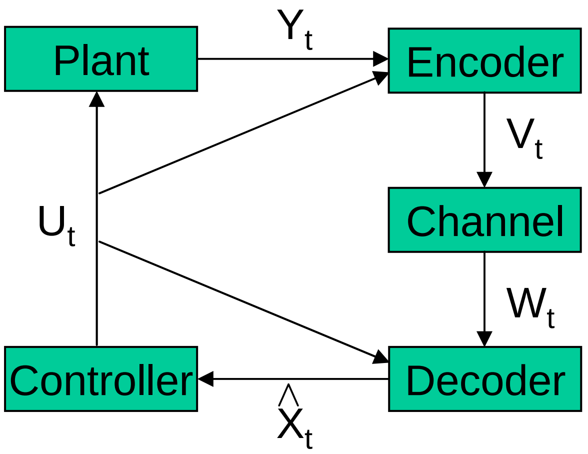
• Understanding role of communication in networked control systems

• What conditions are required on the communication network to achieve different control objectives?

• What limitations do communication networks place on control performance

$\rightarrow$ Goal: show (convince) you that tools from information theory are relevant for these problems.

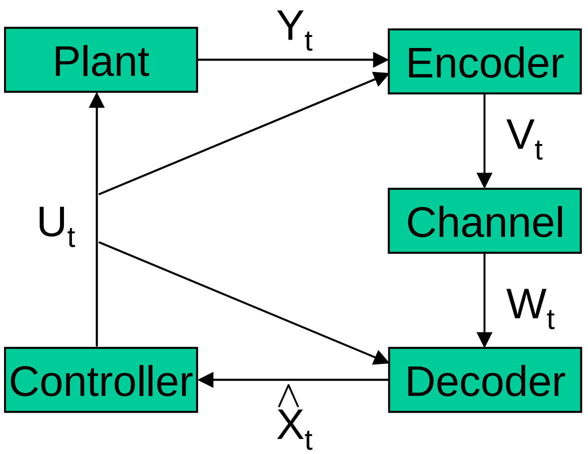
Issues



- Design encoder, decoder, and controller
- Information and policy patterns
- Noiseless versus noisy channels
- Not your father's partial observability

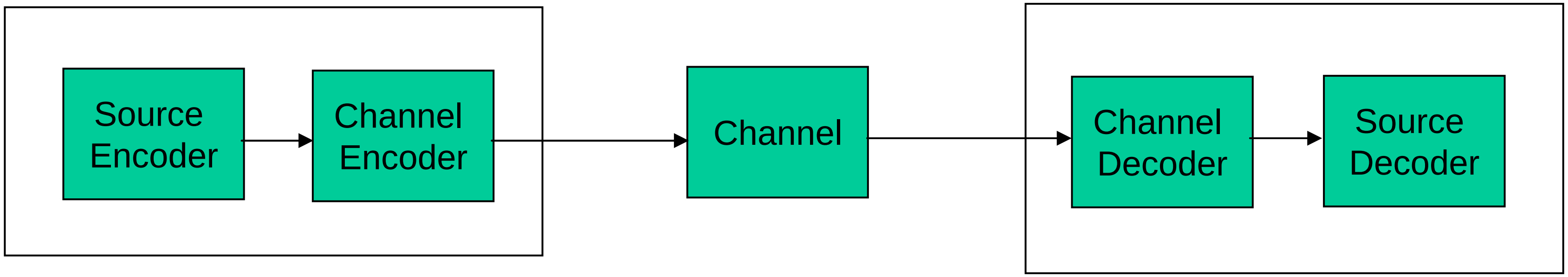
$\rightarrow$ Not as simple as it looks…

A Tale of Two Separations



- Control: Separation between estimation and control

- Information Theory: Separation between source and channel



Outline

- Control over a Noiseless Channel
  - source coding and binning
- Control over a Noisy Channel
  - (joint) source and channel coding
- Control over Multiple Noiseless Channels
  - correlated source coding
- Cooperative Control between Multiple Agents
- Conclusions

Problem Setup

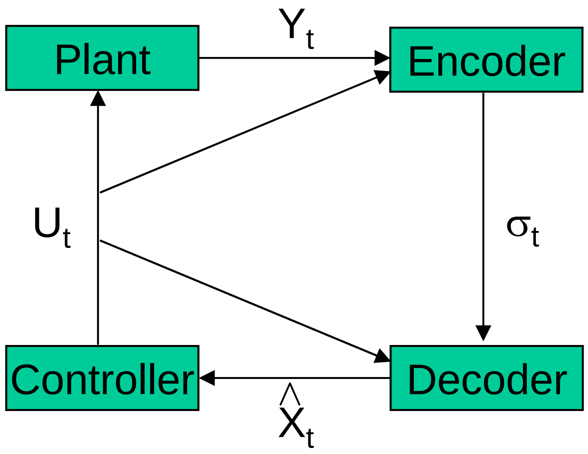
Linear time-invariant system:

$X_{t+1} = A X_t + B U_{t,}$ $Y_t = C X_t$ $\forall t \geq 0$

where $X_t \in \mathbb{R}^d$, $U_t \in \mathbb{R}^m$, and $Y_t \in \mathbb{R}^l$.

The initial position $X_0 \in \Lambda_0 \subseteq R^d$

Noiseless digital channel: R bits
(2^R symbols denoted $\sigma \in \Sigma$)



Lower Bounds on the Rate

Let the *state estimation error* be $e_t = X_t - \hat{X}_t$

At time t we can only distinguish between 2^{tR} initial positions hence the need for these definitions.

Asymptotic observablility: if there exists an encoder and decoder such that the following holds for any control sequence: $\| e_t \|_2 \to 0$.

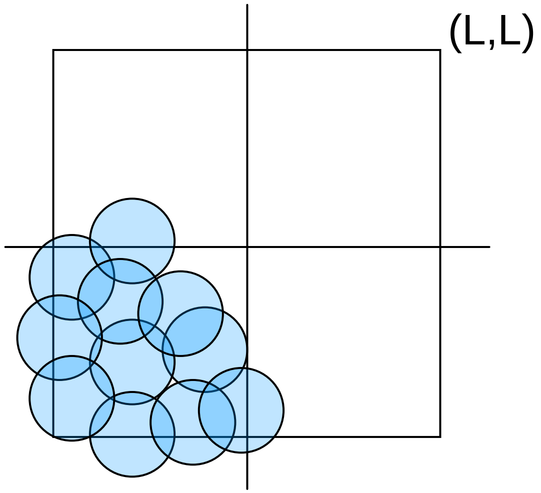
Asymptotic stabilizability: if there exists an encoder, decoder, and controller such that $\| X_t \|_2 \to 0$.

Proposition: A necessary condition for asymptotic observability and asymptotic stabilizability is $R \geq \sum_{\lambda(A)} \max \{0, \log | \lambda(A) | \}$.

Lower Bounds – Part 2

Argument for stabilizability. Assume $X_0 \in \Lambda$. We want $\| X_n \|_2 \leq \varepsilon$.

Let $\Gamma_{U_0,\ldots,U_{n-1}} = \{ X_0 : \| X_n \|_2 \leq \varepsilon \}$.



Count the number of Γ balls it takes to cover Λ.

View the control as a codeword for a given initial condition.

Note that the counting argument is independent of the encoder and decoder chosen.

Encoder Information Patterns

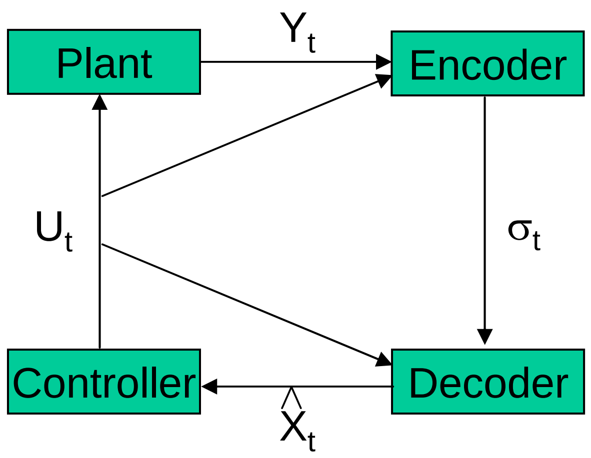
Encoder Class One: $\mathcal{E}_t$: $(Y^t, \sigma^{t-1}, U^{t-1}) \mapsto \sigma_t$.

Encoder Class Two: $\mathcal{E}_t$: $Y^t \mapsto \sigma_t$. The encoder does not know the values of the control signals.

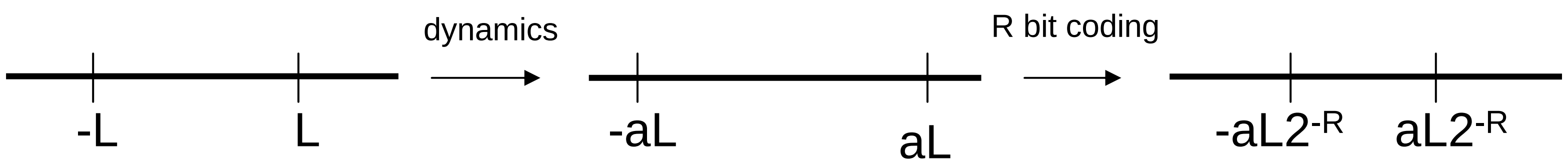
Many cases in between.

In both cases the decoder $\mathcal{D}_t$: $(\sigma^t, U^{t-1}) \mapsto \hat{X}_t$.

Controller $\mathcal{C}_t$: $\hat{X}_t \mapsto U_t$. We are assuming a separation structure between the decoder and the controller.



Key Idea



Uncontrolled system: $X_{t+1} = A X_t$.

Compute a box that bounds the set that X_{t+1} lives in given the box that X_t lives in.

Simple case: $A = \text{diag}\,[\lambda_1, . . . ,\lambda_d]$. If $X_t \in \{ [-L, L] \times ... \times [-L, L] \}$ then $X_{t+1} \in \{ [-|\lambda_1| L, |\lambda_1| L] \times ... \times [-|\lambda_d| L, |\lambda_d| L] \}$.

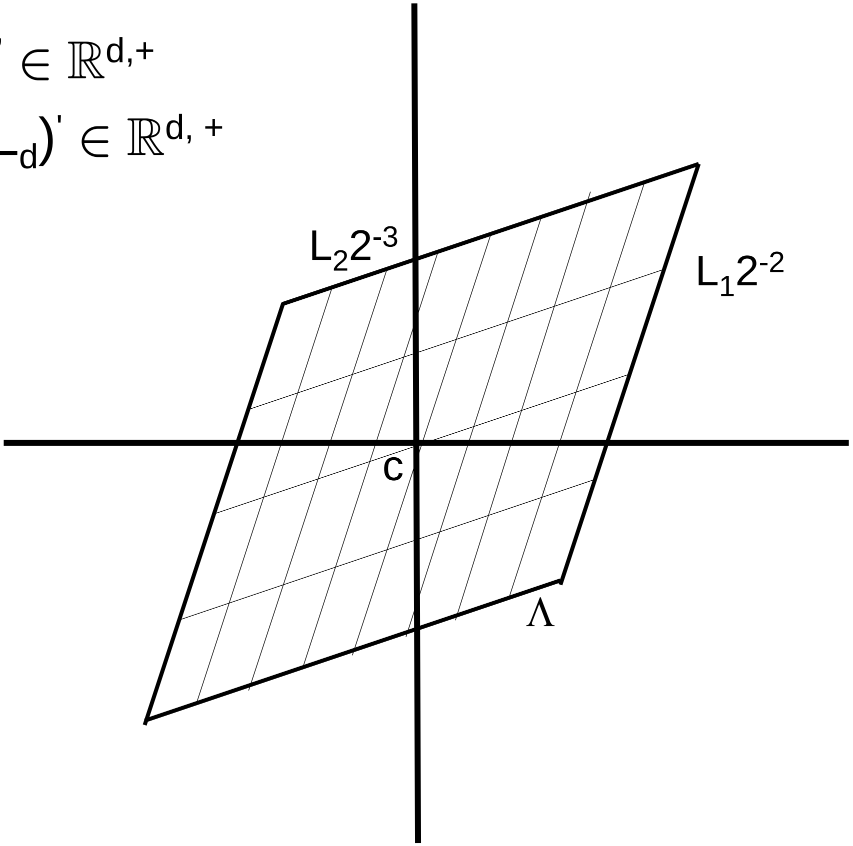
To shrink uncertainty choose $R_i > \max \{0, \log |\lambda_i| \}$

Special construction for general A using the real Jordan canonical form.

Primitive Quantizer

A *primitive quantizer* is a four-tuple (c, R, L, Φ):

- Centroid: $c \in \mathbb{R}^d$
- Rate vector: $R = (R_1, \ldots, R_d)' \in \mathbb{R}^{d,+}$
- Dynamic range: $L = (L_1, \ldots, L_d)' \in \mathbb{R}^{d,+}$
- Coordinate transformation: Φ



$R = \sum_i R_i$

Boxes? View as high-rate (low distortion) lossy source coding.

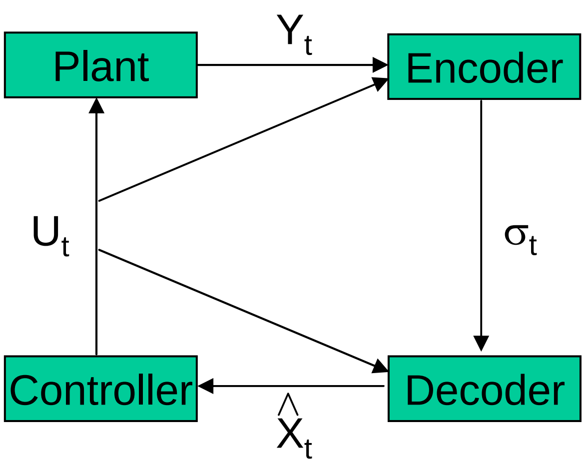
Encoder Class 1 - Sufficiency

Full observation: $Y_t = X_t$
Main idea: compute innovation at encoder. Encoder knows decoder's state estimate:

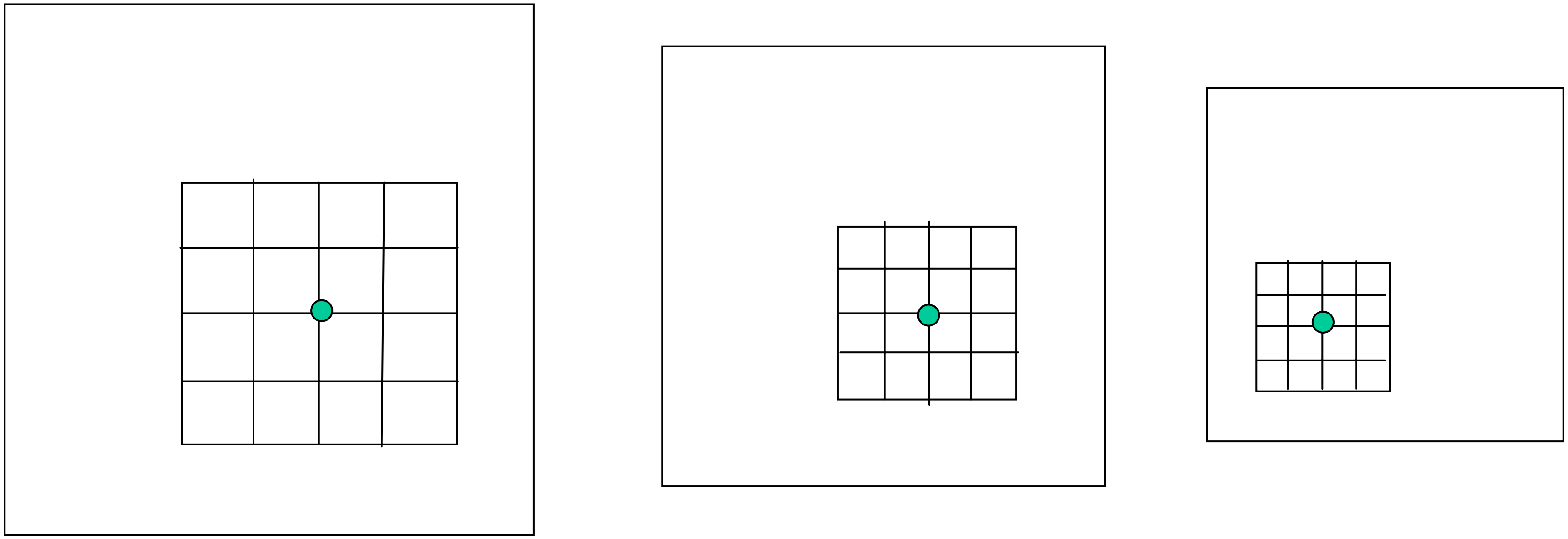
$$X_t - \hat{X}_t = A X_{t-1} + B U_{t-1} - (A \hat{X}_{t-1} + B U_{t-1}) = A e_{t-1}$$

Proposition: For encoders restricted to encoder class one, and bounded initial set Λ_0 a sufficient condition for asymptotic observability and asymptotic stabilizability is $R > \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$.

Rate of convergence: $|| e_t ||_2 \leq \kappa\, 2^{-\alpha t}$
where $\alpha = \min_i R_i - \log|\lambda_i(A)|$



Primitive Quantizer Example



Outer box: range of X_t

Inner box: range of e_t

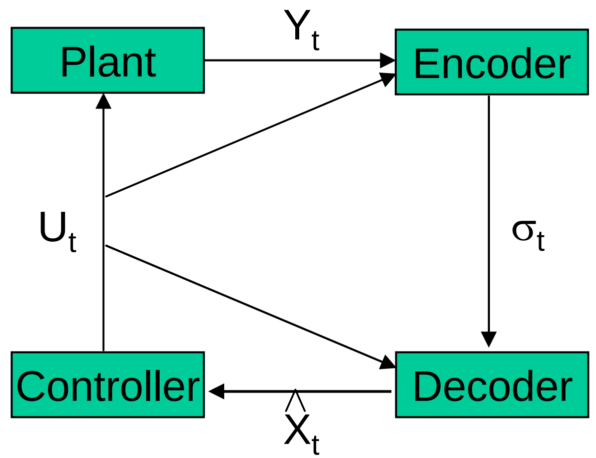
Centroid of inner box: 1-step ahead predictor, ●

Encoder Class 1 - Disturbances

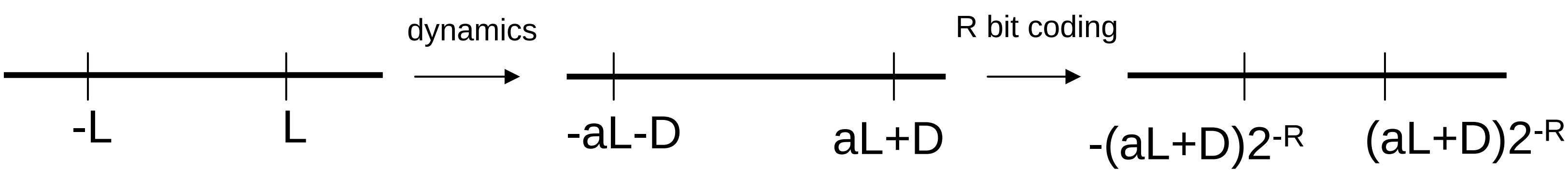
Linear time-invariant system: $X_0 \in \Lambda$,

$$X_{t+1} = A X_t + B U_t + W_t$$

where $\| W_t \|_2 \leq D$



The rate condition is sufficient to insure $\limsup_{t \to \infty} \| e_t \|_2$ is bounded.



Encoder Class 2 - Binning

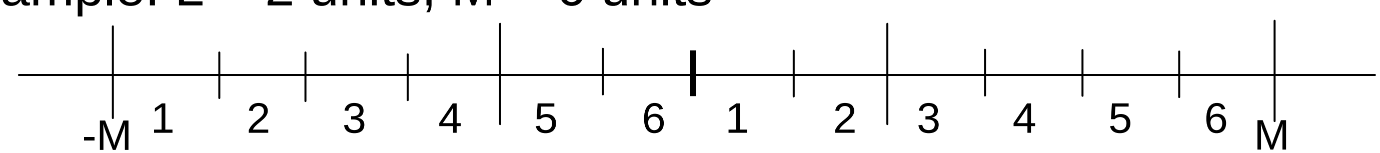
Encoder Class One: $\mathcal{E}_t$: $(Y^t, \sigma^{t-1}, U^{t-1}) \mapsto \sigma_t$
Encoder Class Two: $\mathcal{E}_t$: $Y^t \mapsto \sigma_t$

Idea: source-coding with side-information at the decoder.

Simple example: $X = z + u$ where $|z| \leq L$, $|X| \leq M$, with $L < M$

Case 1: u unknown to both Tx and Rx: $| e | \leq M\, 2^{-R}$
Case 2: u known to both Tx and Rx: $| e | \leq L\, 2^{-R}$
Case 3: u known to only Rx: $| e | \leq L\, 2^{-R}$

Example: L = 2 units, M = 6 units



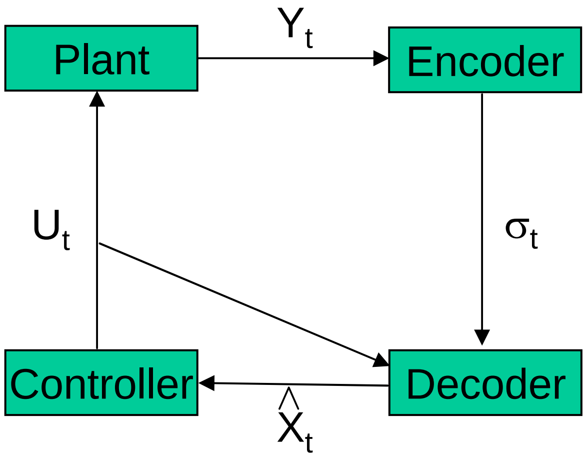
Encoder Class 2 - Sufficiency

Idea: source-coding with side-information at the decoder.

Before we quantized the innovation. Now we should *bin* $Y_t = X_t$:

$$X_t = (A \hat{X}_{t-1} + B U_{t-1}) + A e_{t-1}$$

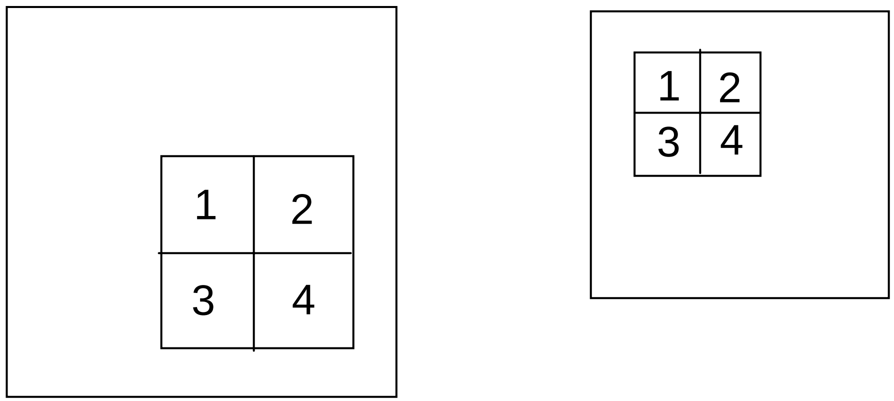
Term in parenthesis known to Rx.



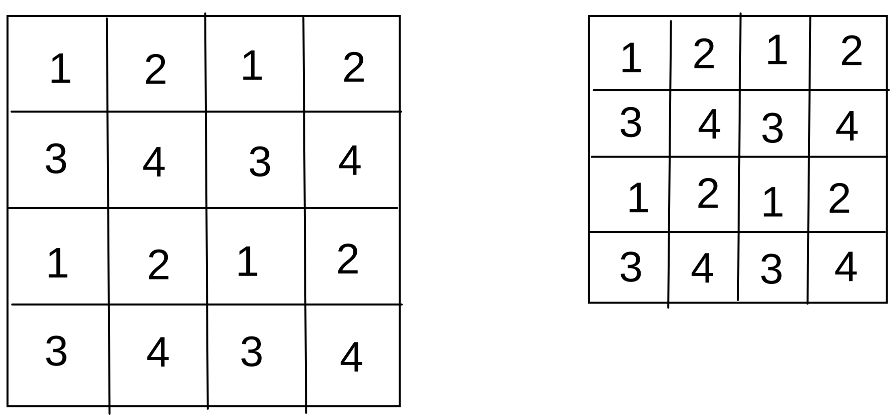
Rate condition sufficient for asymptotic observability and stabilizability.

Non-nested information patterns!

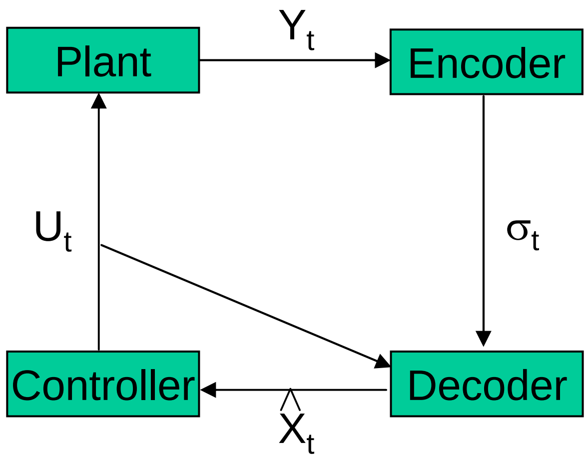
Primitive Quantizer Example – Binning



Encoder: dynamic range, primitive quantizer tiling, resolution



Encoder Class 2 - Output



$Y_t = CX_t$

(A, C) observable

$\exists$ matrices β, γ:

$$A^{d-1} \gamma [Y_{t-d+1}, \ldots, Y_t] = (A \hat{X}_{t-1} + \beta [U_{t-d+1}, \ldots, U_{t-1}]) + Ae_{t-1}$$

Term in parenthesis known to Rx. Hence bin:

$$A^{d-1} \gamma [Y_{t-d+1}, \ldots, Y_t]$$

One can also treat process disturbances

YALE UNIVERSITY Faculty of Engineering

Outline

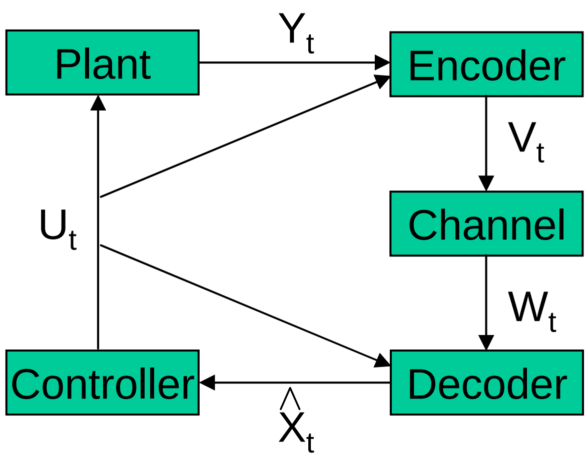
- Control over a Noiseless Channel
- ***Control over a Noisy Channel***
- Control over Multiple Noiseless Channels
- Cooperative Control between Multiple Agents
- Conclusions

Problem Formulation for Noisy Channel

$X_0 \in \Lambda_0$, $X_{t+1} = A X_t + B U_t$, $Y_t = C X_t$, $\forall t \geq 0$. Here X_0 has density $p(X_0)$ with finite differential entropy $h(X_0)$.

Channels
- Noiseless digital channel with rate R.
- Erasure channel with erasure probability α. Often used as a simple model of packet loss on Internet-like channels.
- Memoryless Gaussian channel with power ρ. Often used as a simple model of a wireless channel.

Why noisy channels?



Channel Capacity and Rate Distortion

Channel Capacity: Given a channel $P(W_t \mid v^t, w^{t-1})$, the *Shannon capacity* over a time horizon of length T is:

$$\mathcal{C}_T = \sup_{P(V^T)} I(V^T; W^T)$$

Rate-Distortion: Given a source X with distribution P(X) and d(x, y) be a distortion measure the *rate distortion* function is:

$$\mathcal{R}(D) = \inf_{P(y \mid x)} \{ I(X; Y) \text{ such that } \mathbb{E}(d(X,Y)) \leq D \}$$

Data-Processing Inequality: A necessary condition for reconstruction over T channel uses is $\mathcal{R}(D) \leq \mathcal{C}_T$.

Lower Bounds

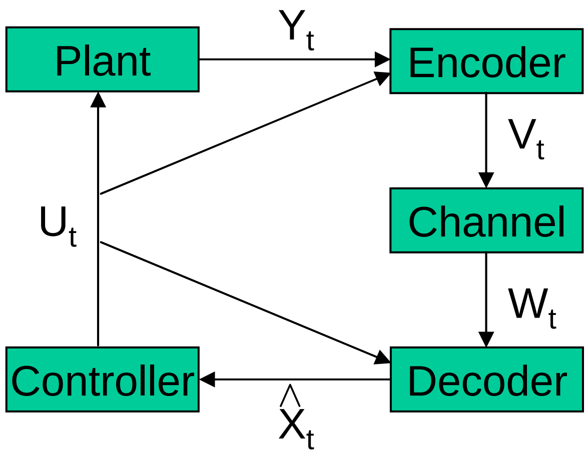
The system is *almost surely asymptotically observable* if there exists an encoder and decoder such that $|| E_t ||_2 \rightarrow 0$ almost surely. The system is *almost surely asymptotically stabilizable* if there exists an encoder, decoder, and controller such that $|| X_t ||_2 \rightarrow 0$ almost surely.

Proposition: A necessary condition for almost sure asymptotic observability and stablizability is
$\mathcal{C} \geq \sum_{\lambda(A)} \max \{0, \log | \lambda(A) | \}$

Proof: Treat state estimation error as a sequential rate distortion problem. Then use the *directed data processing inequality* (a causal version of the data processing inequality that allows for feedback.)

Information theoretic converse.

Erasure Channels



Channel drops packet with probability α

$C = (1 - \alpha) R_{packet}$

Necessary condition:
$R_{packet} \geq 1/ (1 - \alpha) \sum_{\lambda(A)} \max \{ 0, \log | \lambda(A) | \}$

Allow ACKS. Typical of TCP network protocol.

Proposition: Given an erasure channel with ACKS the packet size condition is sufficient to ensure almost sure asymptotic observability and stabilizability.

Erasure Channels – Sufficiency

Scalar case: $X_{t+1} = a X_t + b U_t$

Box dimension can be determined by the stochastic difference equation: $L_{t+1} = |a| F_t L_t$ where the F_t are IID with distribution: $P(F_t = 1) = \alpha$ and $P(F_t = 2^{-R}) = 1 - \alpha$.

The multiplicative law of large numbers shows that $L_t \to 0$ a.s. if $\mathbb{E}(\log |a| F) < 0$.

$E(\log |a| F) = \alpha \log |a| + (1 - \alpha) \log |a| 2^{-R} = \log |a| - (1 - \alpha) R$

Generalizes to vector case.

Plant Disturbances and Noisy Channels

Scalar case: $X_{t+1} = a X_t + bU_t + W_t$ where $| W_t | \leq D$.

Consider erasure case. Box dynamics: $L_{t+1} = F_t (|a| L_t + D)$ where F_t as before.

There is no rate that can insure the $\{L_t\}$ are almost surely bounded. For R_{packet} satisfying the rate condition one can show the $\{ L_t \}$ are uniformly tight and converge to L_∞.

Idea: always the chance for a run of erasures

$\rightarrow$ Need a weaker notion of stability, e.g. moment stability.
Capacity hit with weaker performance

LQG Formulation

$X_{t+1} = F X_t + G U_t + W_t$
where $\{W_t\}$ is an IID sequence $\sim N(0, K_W)$ and $X_0 \sim N(0, K_0)$
and (F, G) is a controllable pair.

Memoryless vector Gaussian channel with power P.

LQG cost: $\limsup_{T \to \infty} 1/T \, \mathbb{E}[\sum X_t' Q X_t + U_t' S U_t]$
with Q, S positive definite.

The addition of a communication channel converts the fully observed LQG control problem above into a partially observed LQ control problem.

Certainty Equivalence

Certainty Equivalence: $U_t = L\, Y_t$ where $Y_t = \mathbb{E}(X_t \mid B^t, U^{t-1})$ is the decoder's estimate of the state of the plant.

Lemma: For the memoryless vector Gaussian channel the certainty equivalent control law is optimal.

Idea: show that there is no dual effect. Roughly, the estimation error is independent of the control applied.

Reduce to fully observed LQ model. Then optimal cost decomposes into a full state observation cost and a cost that depends only on Λ the steady state estimation error covariance: $\mathrm{tr}(PK_W) + \mathrm{tr}((F'PF - P + Q)\,\Lambda)$

(P satisfies the Riccati equation and L is the optimal gain.)

Sequential Rate Distortion Problem

Rate distortion is non-causal and has delays

The *sequential rate distortion function*

$$\mathcal{R}_T^{SRD}(D) = \inf_{\mathcal{F}} 1/T \; I(X^T ; Y^T)$$

where $\mathcal{F} = \{ \; P(Y_t \mid x^t, y^{t-1}) \; : \; \mathbb{E}[\, d(X_t, Y_t) \,] \leq D, \quad t = 1,\ldots,T \; \}$.

A necessary condition on the capacity of a given channel to achieve a given distortion is: $\mathcal{R}_T^{SRD}(D) \leq \mathcal{C}_T$.

This is computable for Gauss-Markov sources with quadratic distortion (i.e. single-letter characterization)

A Computation

For the scalar memoryless Gaussian channel with capacity C the LQG cost is:

$$PK_W + K_W(F^2 P - P + Q) / (2^{2C} - F^2)$$

The optimal cost decomposes into two terms: a full state observation cost and a sequential rate distortion cost.

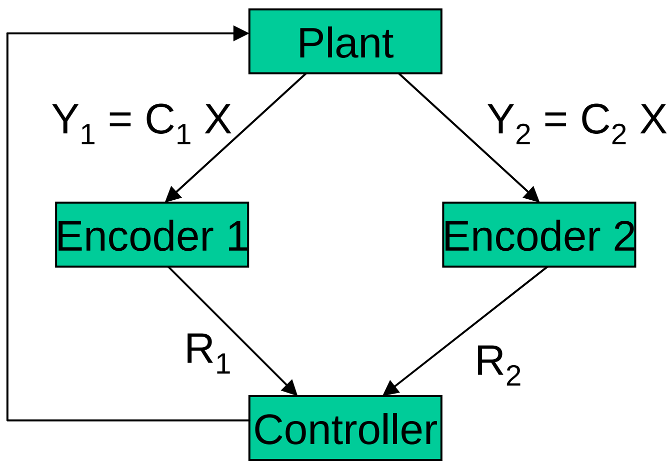
Formulas for vector case are known, though messy.

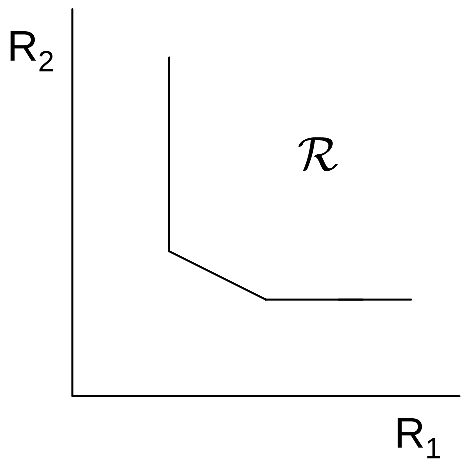
Open question: encoder class 2? Sequential binning over a noisy channel?

Outline

- Control over a Noiseless Channel
- Control over a Noisy Channel
- ***Control over Multiple Noiseless Channels***
- Cooperative Control between Multiple Agents
- Conclusions

Multiple Sensors





M sensors and one controller:
$X_{t+1} = A X_t + BU_t$, $Y_{m,t} = C_m X_t$, m=1,...,M

The system is jointly observable but each individual (A,C_m) may not be observable.

What rates are needed? First examine encoder class one.

Multiple Sensors - Continued

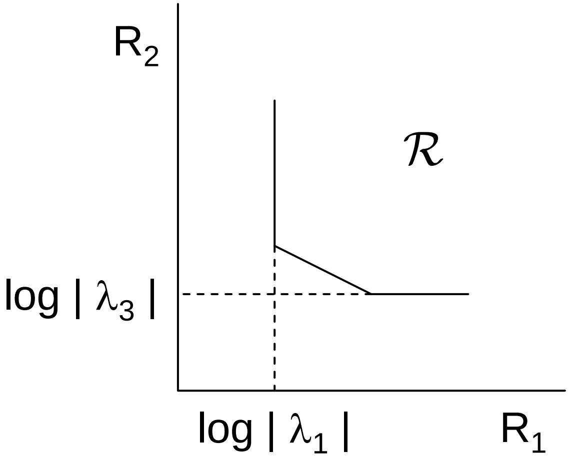
For each m let $\mathcal{O}_m$ be the observable subspace (quotient space corresponding to A-invariant unobservable subspace)

$\Lambda_m = \{ \lambda(A) :$ those eigenvalues of A corresponding to the subspace $\mathcal{O}_m \}$.

Example: $A = \text{diag}[\lambda_1, \lambda_2, \lambda_3]$. Let $C_1 = [1, 1, 0]$ and $C_2 = [0, 1, 1]$.

Hence $\Lambda_1 = \{ \lambda_1, \lambda_2 \}$ and $\Lambda_2 = \{ \lambda_2, \lambda_3 \}$. Then:

- $R_1 + R_2 > \log | \lambda_1 | + \log | \lambda_2 | + \log | \lambda_3 |$
- $R_1 > \log | \lambda_1 |$
- $R_2 > \log | \lambda_3 |$



Rate Region

For each m define the rate vector to be $\underline{R}_m = (R_{m,1}, \ldots, R_{m, \dim m})$.

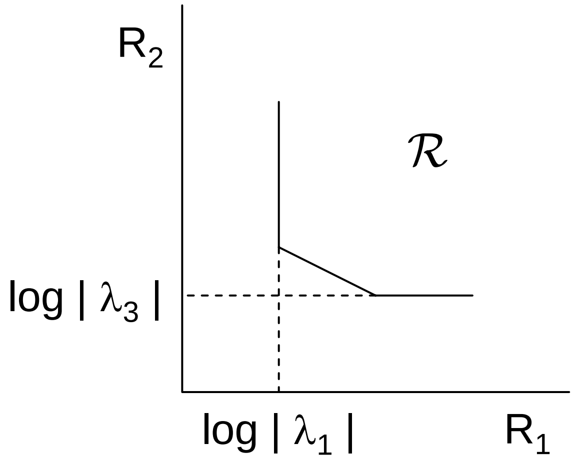
Let

$\mathcal{R} = \{(\underline{R}_1,\ldots,\underline{R}_M) : \sum_{m:\, \lambda(A) \in \Lambda} R_{m,\, \lambda(A)} > \max\{0, \log|\lambda(A)|\},\ \forall\ \lambda(A)\}$

where $R_{m,\, \lambda(A)}$ is the rate assigned by sensor m to mode $\lambda(A)$.

Proposition: A necessary and sufficient condition on the rate vector for asymptotic observability and stabilizability is $(\underline{R}_1,\ldots,\underline{R}_M) \in \mathcal{R}$.

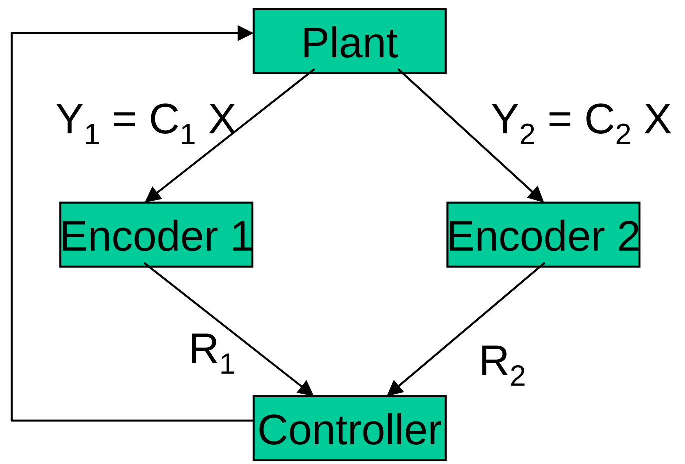
Connection to Slepian-Wolf Coding

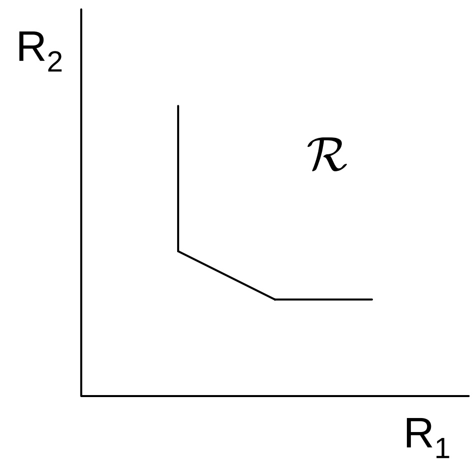


Slepian-Wolf coding region: $p(Z_1, Z_2)$

- $R_1 + R_2 \geq H(Z_1, Z_2) \leftrightarrow \sum_{\lambda} \max\{0, \log|\lambda|\}$
- $R_1 \geq H(Z_1 \mid Z_2) \leftrightarrow \sum_{\lambda \in \Lambda_1 \setminus (\Lambda_1 \cap \Lambda_2)} \max\{0, \log|\lambda|\}$
- $R_2 \geq H(Z_2 \mid Z_1) \leftrightarrow \sum_{\lambda \in \Lambda_2 \setminus (\Lambda_1 \cap \Lambda_2)} \max\{0, \log|\lambda|\}$

Multiple Sensors – Encoder Class 2



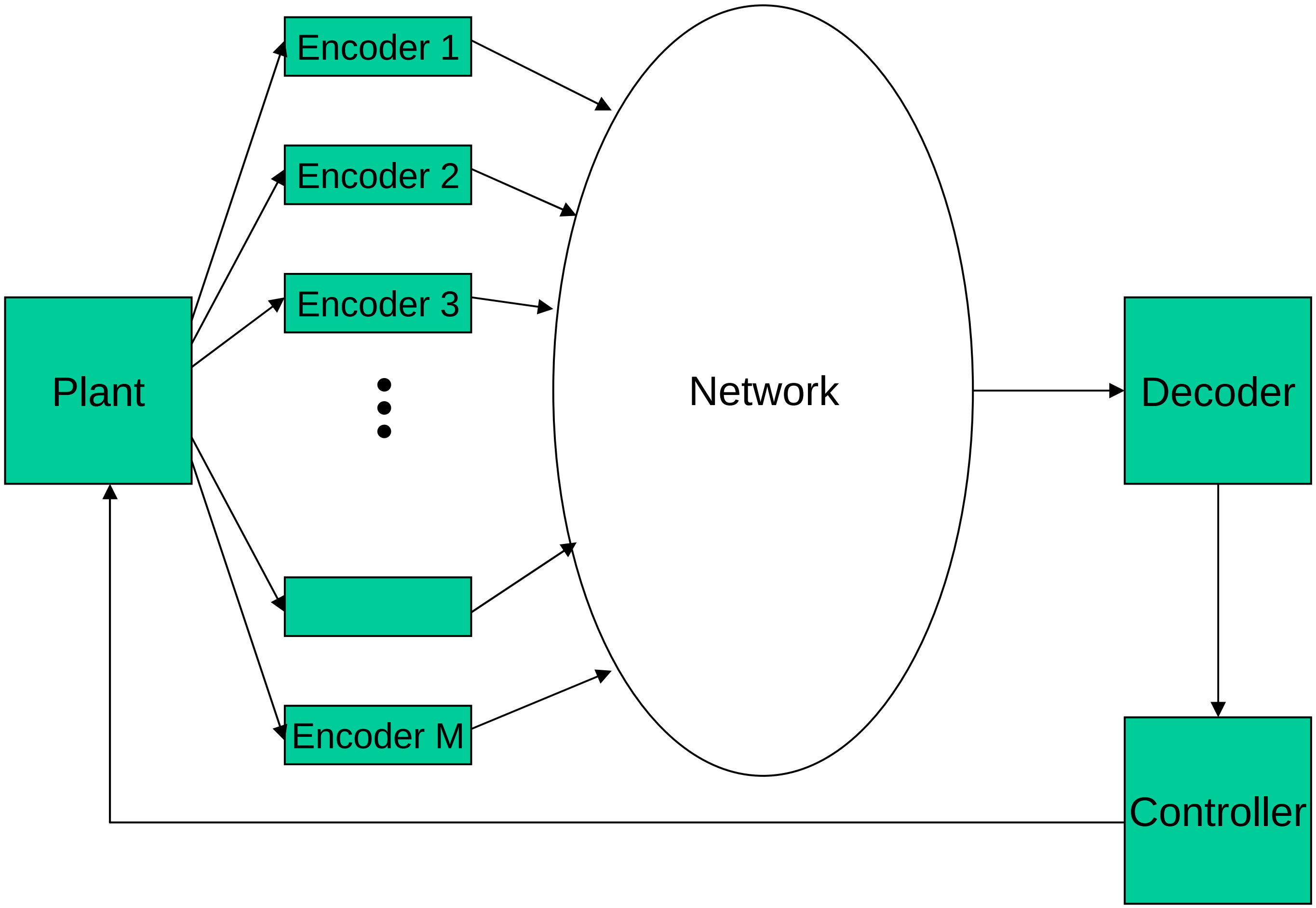


What do we do if controls not available to encoders?

Combine Slepian-Wolf coding with binning (source-coding with side-information) technique.

Why is result useful?

Sensor Network Example



Outline

- Control over a Noiseless Channel
- Control over a Noisy Channel
- Control over Multiple Noiseless Channels
- ***Cooperative Control between Multiple Agents***
- Conclusions

Cooperative Control Problem

G = (V, E), static (no mobility) undirected graph

Decoupled dynamics: $X_i(t+1) = A_i X_i(t) + B_i U_i(t) \quad \forall\, i \in V$

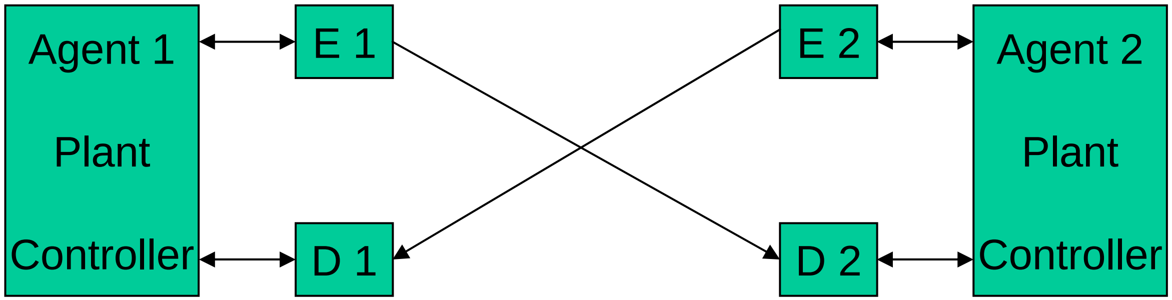
Coupled control: $U_i(t) = \sum_{j \in \mathcal{N}_i \cup i} K_{ij} X_j(t) \quad \forall\, i \in V$

Need to communicate $\{X_j(t) : j \in \mathcal{N}_i\}$ to agent i.

Lower bound: $R_{ji} \geq \sum_{\lambda(A)} \max \{ 0, \log |\lambda(A_j)| \} \quad \forall\, j \in \mathcal{N}_i$.

Achievable? Open question...

Cooperative Control Example

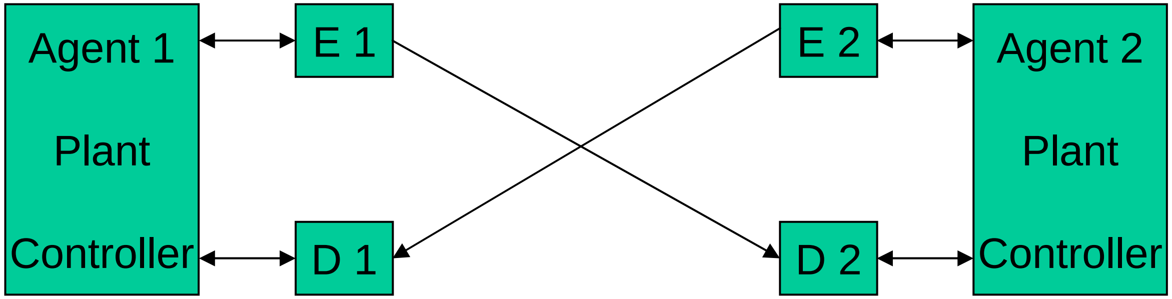


Agent 2 transmits information about $X_2(t)$ to agent 1.

Control availability:
- Before: Encoder: no Decoder: yes
- Now: Encoder: yes Decoder: no, unless...

Agent 1's decoder has access to $U_2(t)$ if either $U_2(t)$ is transmitted or it can be computed locally at agent 1.

Cooperative Control Example – Part 2



Agent 1's decoder has access to $U_2(t)$ if either $U_2(t)$ is transmitted or it can be computed locally at agent 1.

The latter is possible if $U_2(t) = K_{21}\hat{X}_1(t) + K_{22}\hat{X}_2(t)$

as opposed to $U_2(t) = K_{21}\hat{X}_1(t) + K_{22}X_2(t)$.

(Note we assume the agents know each other's control policies.)

In this way both agents 1 and 2 know the state estimate of the other and hence can compute each other's control.

Cooperative Control Example – Part 3

Note that the control is based on global knowledge of a state estimate. The coding is based on local knowledge.

This scheme requires every node to communicate (broadcast) with every other node even if they are not connected by an edge in the underlying graph. In order for agent i to compute the control of a neighbor j, $U_j(t)$, it needs to know the states of j's neighbors.

Spatial mixing? Can we get away with knowledge of agents only a few hops away?

YALE UNIVERSITY Faculty of Engineering

Conclusions and Challenges

- Information pattern and policy pattern (cooperation)
- Non-classical information patterns and binning
- Real-time information theory
- Optimality of separation
- Cooperative control with communication constraints:
  - local versus global state
  - spatial mixing

More info: http://www.pantheon.yale.edu/~sct29