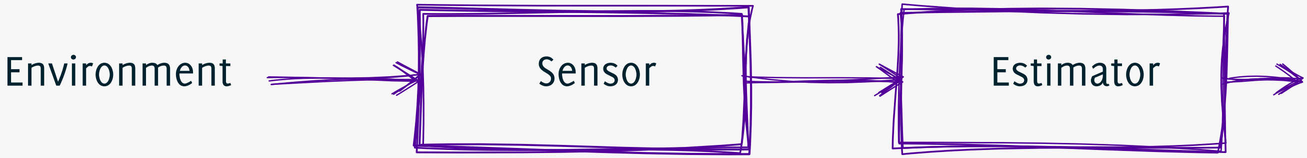
Structure of optimal policies in active sensing

Aditya Mahajan

McGill University

March 28, IEEE International Conference on Accoustics, Speech, and Signal Processing

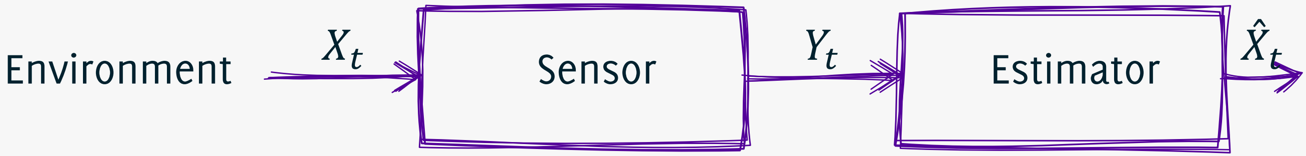
Sensor network over packetized network



Sensor transmits packet to the estimator

- Measurement's are (almost) free, but . . .
- . . . each transmission incurs a constant energy
  (energy required to switch on the radio and transmit a packet).
- The sensor decides when and how to transmit
  (as opposed to the estimator scheduling the transmissions
  . . . or the sensor communicating without coding)
- Trade-off transmit energy and estimation quality

Simplest model to capture the trade-off



Environment

First order Markov process $\{X_t, t = 1, 2, \ldots\}$, $X_t \in \mathbb{X}$

Simplest model to capture the trade-off

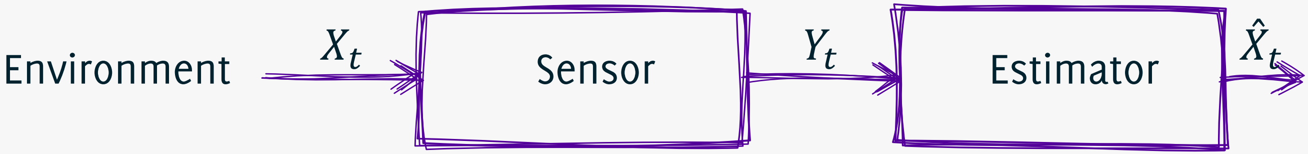
Environment

First order Markov process $\{X_t, t = 1, 2, \ldots\}$, $X_t \in \mathbb{X}$

Sensor

$$Y_t = f_t(X_{1:t}, Y_{1:t-1}), \qquad Y_t \in \mathbb{Y} := \mathbb{X} \cup \{\mathbb{b}\}$$

Simplest model to capture the trade-off



Environment

First order Markov process $\{X_t, t = 1, 2, \dots\}$, $X_t \in \mathbb{X}$

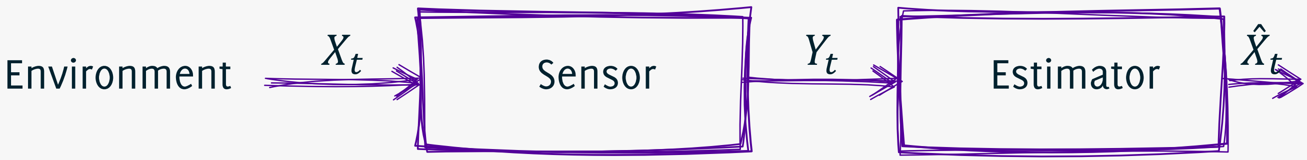
Sensor

$$Y_t = f_t(X_{1:t}, Y_{1:t-1}), \qquad Y_t \in \mathbb{Y} := \mathbb{X} \cup \{\mathbb{b}\}$$

Estimator

$$\hat{X}_t = g_t(Y_{1:t}), \qquad \hat{X}_t \in \mathbb{X}$$

Simplest model to capture the trade-off



Environment

First order Markov process $\{X_t, t = 1, 2, \dots\}$, $X_t \in \mathbb{X}$

Sensor

$$Y_t = f_t(X_{1:t}, Y_{1:t-1}), \qquad Y_t \in \mathbb{Y} := \mathbb{X} \cup \{\mathbb{b}\}$$

Estimator

$$\hat{X}_t = g_t(Y_{1:t}), \qquad \hat{X}_t \in \mathbb{X}$$

Cost $c(X_t, Y_t, \hat{X}_t) = p(Y_t) + \lambda d(X_t, \hat{X}_t)$

- $p(y) = \begin{cases} 0, & y = \mathbb{b}, \\ p^*, & y \in \mathbb{X} \end{cases}$
- d is a metric on $\mathbb{X}$

Problem Formulation

Transmission policy and estimation policies

$$\mathbf{f} = (f_1, f_2, \dots, f_T) \qquad \mathbf{g} = (g_1, g_2, \dots, g_T)$$

Performance of a policy

$$\mathcal{J}(\mathbf{f}, \mathbf{g}) \coloneqq \mathbb{E}^{\mathbf{f},\mathbf{g}}\Big[\sum_{t=1}^{T} p(Y_t) + \lambda d(X_t, \hat{X}_t) \Big]$$

History Dependent Policies $(\mathcal{F}, \mathcal{G})$

$$\mathcal{F} = \prod_{t=1}^{T} F_t, \quad F_t = \big(\mathbb{X}^t \times \mathbb{Y}^{t-1} \mapsto \mathbb{Y}\big)$$

$$\mathcal{G} = \prod_{t=1}^{T} G_t, \quad G_t = \big(\mathbb{Y}^t \mapsto \mathbb{X}\big)$$

Problem Formulation

Transmission policy and estimation policies

$$\mathbf{f} = (f_1, f_2, \ldots, f_T) \qquad \mathbf{g} = (g_1, g_2, \ldots, g_T)$$

Performance of a policy

$$\mathcal{J}(\mathbf{f}, \mathbf{g}) := \mathbb{E}^{\mathbf{f},\mathbf{g}}\Big[\sum_{t=1}^{T} p(Y_t) + \lambda d(X_t, \hat{X}_t) \Big]$$

History Dependent Policies $(\mathcal{F}, \mathcal{G})$

(P) Choose $(\mathbf{f}, \mathbf{g}) \in (\mathcal{F}, \mathcal{G})$ to minimize $\mathcal{J}(\mathbf{f}, \mathbf{g})$

Salient Features

Data increasing with time

$$Y_t = f_t(X_{1:t}, Y_{1:t-1}) \qquad \hat{X}_t = g_t(Y_{1:t})$$

Salient Features

Data increasing with time

$$Y_t = f_t(X_{1:t}, Y_{1:t-1}) \qquad \hat{X}_t = g_t(Y_{1:t})$$

Decentralized control problem

- Two decision makers, the sensor and the estimator, that have access to different information.
- . . . non-classical information structure: in general, not possible to solve using dynamic programming.
- Can be transformed to a system with partial information sharing information structure (Nayyar, Mahajan, and Teneketzis, -2012) which allows us to use dynamic programming

Literature overview

Active sensing/Sensor “censoring”

- Huge literature. Most relevant results:

Literature overview

Active sensing/Sensor "censoring"

- Huge literature. Most relevant results:
- Imer and Basar, 2005, Lipsa and Martins, 2011
  - Assume $\{X_t, t = 1, 2, \ldots\}$ is a Gauss-Markov process
  - Structure of optimal policy: Communicate if

$$|X_t - \text{prediction if sensor does not transmit}| \geq \tau_t$$

  - Cyclic definition? In general, estimation depends on the transmission policy.

    Also show that the estimation policy is the same as prediction of an unobserved Markov source and, as such, does not depend on the transmission policy.

Literature overview

Active sensing/Sensor "censoring"

- Huge literature. Most relevant results:
- Imer and Basar, 2005, Lipsa and Martins, 2011
  - Assume $\{X_t, t = 1, 2, \dots\}$ is a Gauss-Markov process
  - Structure of optimal policy: Communicate if

$$|X_t - \text{prediction if sensor does not transmit}| \geq \tau_t$$

- Nayyar, Basar, Teneketzis, Veeravalli, 2012
  - Assume $\{X_t, t = 1, 2, \dots\}$ is a symmetric, unimodal, and countably supported Markov chain
  - Similar structure of transmission and estimation policies.

Literature overview

Implication of the results

- (For specific models of Markov sources) optimal policies are easy to search and easy to implement.
- The proof relies on symmetry of the Markov process

Literature overview

Implication of the results

- (For specific models of Markov sources) optimal policies are easy to search and easy to implement.
- The proof relies on symmetry of the Markov process

Non-symmetric finitely supported Markov chains

For example, $\mathbb{X} = \{\text{dry}, \text{low moisture}, \text{saturated}\}$ and

$$P = \begin{bmatrix} 0.8 & 0.0 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.0 & 0.1 & 0.9 \end{bmatrix}$$

Literature overview

Implication of the results

- (For specific models of Markov sources) optimal policies are easy to search and easy to implement.
- The proof relies on symmetry of the Markov process

Non-symmetric finitely supported Markov chains

For example, $\mathbb{X} = \{\text{dry}, \text{low moisture}, \text{saturated}\}$ and

$$P = \begin{bmatrix} 0.8 & 0.0 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.0 & 0.1 & 0.9 \end{bmatrix}$$

Are optimal policies easy to find and easy to implement?

Main Results: Notation

Notation

Info state : $\pi_t = \mathbb{P}(X_t \mid Y_{1:t-1})$

Silence set : $S_t = \{x \in \mathbb{X} : \varphi_t(x) = \mathbb{b}\}$

Optimal estimate : $\hat{x}_t^* = \gamma_t(\mathbb{b})$

Main Results: Structure of optimal policy

Recursively define $V_{T+1}(\pi) = 0$ and

$$V_t(\pi_t) = \min_{(S_t, \hat{x}_t)} \Big\{ \mathbb{E}[c(X_t, \hat{X}_t, Y_t) + V_{t+1}(\pi_{t+1}) \mid \pi_t)] \Big\}$$

Main Results: Structure of optimal policy

Recursively define $V_{T+1}(\pi) = 0$ and

$$V_t(\pi_t) = \min_{(S_t, \hat{x}_t)} \Big\{ \mathbb{E}[c(X_t, \hat{X}_t, Y_t) + V_{t+1}(\pi_{t+1}) \mid \pi_t)] \Big\}$$

Let $S^*_t(\pi_t)$ and $\hat{x}^*_t(\pi_t)$ denote the arg-min of the RHS.

Optimal transmission policy

$$f^*_t(x_t, \pi_t) = \begin{cases} \mathbb{b} & \text{if } x_t \in S^*_t(\pi_t) \\ x_t & \text{otherwise} \end{cases}$$

Optimal estimation policy

$$g^*_t(y_t, \pi_t) = \begin{cases} \hat{x}^*_t(\pi_t) & \text{if } y_t = \mathbb{b} \\ y_t & \text{otherwise} \end{cases}$$

Infinite horizon system

The dynamic program reduces to a fixed-point equation

$$V(\pi) = \min_{(S,\hat{x})} \Big\{ \mathbb{E}[c(X, \hat{X}, Y) + \beta V(\pi_+ \mid \pi)] \Big\}$$

Infinite horizon system

The dynamic program reduces to a fixed-point equation

$$V(\pi) = \min_{(S,\hat{x})} \Big\{ \mathbb{E}[c(X, \hat{X}, Y) + \beta V(\pi_+ \mid \pi)] \Big\}$$

The mapping $\pi \mapsto (S, \hat{x})$ is time-invariant. $\implies$ stationary policies are optimal.

Optimal transmission policy

$$f^*(x, \pi) = \begin{cases} \mathbb{b} & \text{if } x \in S^*(\pi) \\ x & \text{otherwise} \end{cases}$$

Optimal estimation policy

$$g^*(y, \pi) = \begin{cases} \hat{x}^*(\pi) & \text{if } y = \mathbb{b} \\ y & \text{otherwise} \end{cases}$$

Infinite horizon system

The dynamic program reduces to a fixed-point equation

$$V(\pi) = \min_{(S,\hat{x})} \Big\{ \mathbb{E}[c(X, \hat{X}, Y) + \beta V(\pi_+ \mid \pi)] \Big\}$$

The mapping $\pi \mapsto (S, \hat{x})$ is time-invariant. $\implies$ stationary policies are optimal.

In general, stationary policies need not be optimal for decentralized control problem.

Time-homogeneous systems with partial sharing information structure are one instance where stationary policies are optimal.

y otherwise

Numerical solution

Let $\tau_t = \max\{s < t : Y_s \neq \mathbb{b}\}$ be the last time a non-blank symbol was transmitted. Then,

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}) = \mathbb{P}(X_t \mid Y_{\tau:t}) \equiv (Y_\tau, t - \tau)$$

Numerical solution

Let $\tau_t = \max\{s < t : Y_s \neq \mathbb{b}\}$ be the last time a non-blank symbol was transmitted. Then,

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}) = \mathbb{P}(X_t \mid Y_{\tau:t}) \equiv (Y_\tau, t - \tau)$$

Simplification of the dynamic program

The above transformation converts the POMDP to a countable state MDP.

$$V(y,n) = \min_{S_n, \hat{x}_n} \Big\{ \sum_{x \notin S_n} \pi_n(x)\big[p^* + \beta V(x,1)\big] + \sum_{x \in S_n} \pi_n(x)\big[d(x,\hat{x}_n) + \beta V(x, n+1)\big]\Big\}$$

where $\pi_n(\cdot) = \mathbb{P}(X_n = \cdot \mid X_1 = y, X_2 \in S_2, \dots, X_{n-1} \in S_{n-1})$

Numerical solution

Let $\tau_t = \max\{s < t : Y_s \neq \mathbb{b}\}$ be the last time a non-blank symbol was transmitted. Then,

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}) = \mathbb{P}(X_t \mid Y_{\tau:t}) \equiv (Y_\tau, t - \tau)$$

Simplification of the dynamic program

The above transformation converts the POMDP to a countable state MDP.

$$V(y, n) = \min_{S_n, \hat{x}_n} \Big\{ \sum_{x \notin S_n} \pi_n(x) \big[p^* + \beta V(x, 1) \big] + \sum_{x \in S_n} \pi_n(x) \big[d(x, \hat{x}_n) + \beta V(x, n+1) \big] \Big\}$$

If we can bound $t - \tau$ (either due to the structure of the cost and transition matrix or due to implementation constraints), then the countable state MDP reduces to finite state MDP.

Implementation

Define, the array of indexes:

$$A(t, t-\tau) = \{x, x' \in \mathbb{X} : f_t(x, x', t-\tau) = \mathbb{b}\}$$

and the vector

$$B(t, t-\tau)[x'] = g_t(\mathbb{b}, x', t-\tau)$$

Implementation

Define, the array of indexes:

$$A(t, t-\tau) = \{x, x' \in \mathbb{X} : f_t(x, x', t-\tau) = \mathbb{b}\}$$

and the vector

$$B(t, t-\tau)[x'] = g_t(\mathbb{b}, x', t-\tau)$$

Store the arrays $A(t, n)$ and the vectors $B(t, n)$, $t = 1, \dots, T$ and $n < t$, at the transmitter and the receiver, respectively.

Keep track of τ and y_τ. Then,

$$y_t = \begin{cases} \mathbb{b} & \text{if } (x_t, y_\tau) \in A(t, t-\tau) \\ x_t & \text{otherwise} \end{cases}$$

$$\hat{x}_t = \begin{cases} B(t, t-\tau)[y_\tau] & \text{if } y_t = \mathbb{b} \\ y_t & \text{otherwise} \end{cases}$$

Implementation

Define, the array of indexes:

$$A(t, t-\tau) = \{x, x' \in \mathbb{X} : f_t(x, x', t-\tau) = \mathbb{b}\}$$

and the vector

$$B(t, t-\tau)[x'] = g_t(\mathbb{b}, x', t-\tau)$$

d $n < t$, at

For infinite horizon, A and B
depend on t only through $t - \tau$

$$y_t = \begin{cases} \mathbb{b} & \text{if } (x_t, y_\tau) \in A(t, t-\tau) \\ x_t & \text{otherwise} \end{cases}$$

$$\hat{x}_t = \begin{cases} B(t, t-\tau)[y_\tau] & \text{if } y_t = \mathbb{b} \\ y_t & \text{otherwise} \end{cases}$$

Comparison with previous results

Symmetric, unimodal, countably or uncountably supported MC

- Optimal transmission policy:

$$(x_t, y_\tau) \in A(t, t-\tau) \iff |x_t - B(t, t-\tau)[y_\tau]| < d_t$$

- Optimal estimation policy:

$$B(t, t-\tau)[y_\tau] = \text{best } t-\tau \text{ step predictor of MC starting at } y_\tau$$

Four stage proof outline

Four stage proof outline

Stage 1 The sensor may ignore the history of past observations and use a transmitting policy of the form

$$y_t = f_t(x_t, y_{1:t1})$$

without any loss of optimality.

Four stage proof outline

Stage 1 The sensor may ignore the history of past observations and use a transmitting policy of the form

$$y_t = f_t(x_t, y_{1:t1})$$

without any loss of optimality.

Proof: The process $r_t = (x_t, y_{1:t-1})$ is a controlled MC with control action y_t, i.e.,

- $\mathbb{P}(r_{t+1} \mid r_{1:t}, y_{1:t}) = \mathbb{P}(r_{t+1} \mid r_t, y_t)$
- $\mathbb{E}[c(x_t, y_t, \hat{x}_t) \mid r_{1:t}, y_{1:t})] = \mathbb{E}[c(x_t, y_t, \hat{x}_t) \mid r_t, y_t]$

Four stage proof outline

Stage 2 Causal real-time coding does not improve performance, i.e., if a sensor transmits, it must transmit the current observation.

Therefore, if the estimator receives a non-blank, it chooses that symbol as its estimate.

Four stage proof outline

Stage 2 Causal real-time coding does not improve performance, i.e., if a sensor transmits, it must transmit the current observation.

Therefore, if the estimator receives a non-blank, it chooses that symbol as its estimate.

Proof: Based on a backward interchange argument. If transmitters and estimators at time $t + 1 : T$ have the desired structure, we can construct a transmitter and estimator at time t that has the desired structure and performs as well as the original policy.

Four stage proof outline

Stage 3 $\pi_t(\cdot) = \mathbb{P}(X_t = \cdot | y_{1:t-1})$ is a sufficient statistic for $y_{1:t-1}$ at both the transmitter and receiver. Thus, restricting attention to transmitters and estimators of the form

$$y_t = f_t(x_t, \pi_t) \qquad \hat{x}_t = g_t(y_t, \pi_t)$$

is without any loss.

Four stage proof outline

Stage 3 $\pi_t(\cdot) = \mathbb{P}(X_t = \cdot | y_{1:t-1})$ is a sufficient statistic for $y_{1:t-1}$ at both the transmitter and receiver. Thus, restricting attention to transmitters and estimators of the form

$$y_t = f_t(x_t, \pi_t) \qquad \hat{x}_t = g_t(y_t, \pi_t)$$

is without any loss.

Proof: common data $y_{1:t-1}$ and private data x_t and y_t. Thus, the system has partial sharing information structure (Nayyar, Mahajan, Teneketzis, 2012).

The sensor and receiver choose their actions as

$$f_t(x_t, y_{1:t-1}) = \varphi_t(y_{1:t-1})(x_t) \qquad g_t(y_t, y_{1:t-1}) = \gamma_t(y_{1:t-1})(y_t).$$

Four stage proof outline

Stage 3 $\pi_t(\cdot) = \mathbb{P}(X_t = \cdot | y_{1:t-1})$ is a sufficient statistic for $y_{1:t-1}$ at both the transmitter and receiver. Thus, restricting attention to transmitters and estimators of the form

$$y_t = f_t(x_t, \pi_t) \qquad \hat{x}_t = g_t(y_t, \pi_t)$$

is without any loss.

Proof:

Consider a coordinator that observes common data $y_{1:t-1}$ and chooses

$$\varphi_t : \mathbb{X} \to \mathbb{Y} \qquad \gamma_t : \mathbb{Y} \to \mathbb{X}$$

The sensor and receiver choose their actions as

$$f_t(x_t, y_{1:t-1}) = \varphi_t(y_{1:t-1})(x_t) \qquad g_t(y_t, y_{1:t-1}) = \gamma_t(y_{1:t-1})(y_t).$$

Four stage proof outline

Stage 3 $\pi_t(\cdot) = \mathbb{P}(X_t = \cdot | y_{1:t-1})$ is a sufficient statistic for $y_{1:t-1}$ at both the transmitter and receiver. Thus, restricting attention to transmitters and estimators of the form

$$y_t = f_t(x_t, \pi_t) \qquad \hat{x}_t = g_t(y_t, \pi_t)$$

is without any loss.

Proof:

The sensor and receiver choose their actions as

$$f_t(x_t, y_{1:t-1}) = \varphi_t(y_{1:t-1})(x_t) \qquad g_t(y_t, y_{1:t-1}) = \gamma_t(y_{1:t-1})(y_t).$$

Coordinated system is a PODMP that is equivalent to the original decentralized system $\implies$ DP decomposition

$$V_t(\pi_t) = \min_{(\varphi_t, \gamma_t)} \Big\{ \mathbb{E}[c(X_t, \hat{X}_t, Y_t) + V_{t+1}(\pi_{t+1}) \mid \pi_t)] \Big\}$$

Four stage proof outline

Stage 4 Let $S_t = \{x \in \mathbb{X} : \varphi_t(x) = \mathbb{b}\}$ and $\hat{x}_t^* = \gamma_t(\mathbb{b})$. Then

$$\varphi_t(\cdot) \equiv S_t \qquad \gamma_t(\cdot) \equiv \hat{x}_t^*$$

This simplifies the form of the DP at the coordinator.

Four stage proof outline

Stage 4 Let $S_t = \{x \in \mathbb{X} : \varphi_t(x) = \mathbb{b}\}$ and $\hat{x}_t^* = \gamma_t(\mathbb{b})$. Then

$$\varphi_t(\cdot) \equiv S_t \qquad \gamma_t(\cdot) \equiv \hat{x}_t^*$$

This simplifies the form of the DP at the coordinator.

Proof: Follows from the result of Stage 2 that real-time coding does not improve performance.

Conclusion

Main result Identified structure of optimal policy

Optimal transmission policy

$$f_t^*(x_t, y_\tau, t-\tau) = \begin{cases} \mathbb{b} & \text{if } x_t \in S_t^*(y_\tau, t-\tau) \\ x_t & \text{otherwise} \end{cases}$$

Optimal estimation policy

$$g_t^*(y_t, y_\tau, t-\tau) = \begin{cases} \hat{x}_t^*(y_\tau, t-\tau) & \text{if } y_t = \mathbb{b} \\ y_t & \text{otherwise} \end{cases}$$

where $(S_t^*, \hat{x}_t^*)$ are given by the solution of a countable state MDP.

Optimal policy is easy to implement.

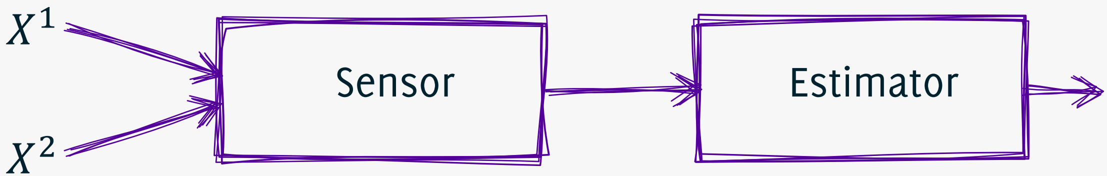
Future work

Generalization to arbitrarily connected sensor networks .

Future work

Generalization to arbitrarily connected sensor networks .

- Multi-dimensional symmetric, unimodal Markov Processes.

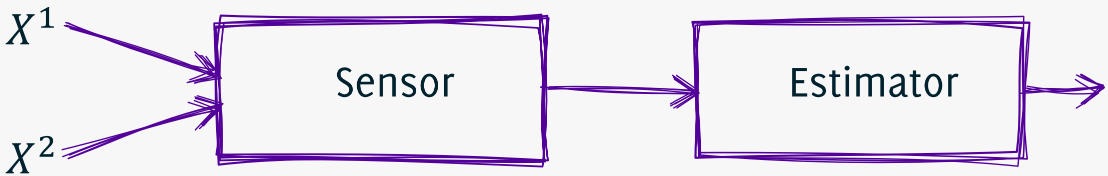


The results of Lipsa and Martins, 2011 and of Nayyar, Basar, Teneketzis, Veeravalli, 2012 do not apply to multi-dimensional Markov processes.

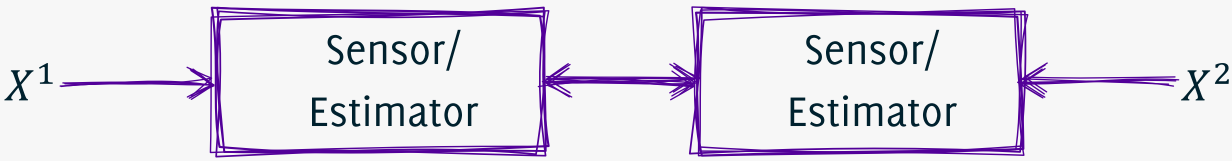
Future work

Generalization to arbitrarily connected sensor networks .

- Multi-dimensional symmetric, unimodal Markov Processes.



- Two-node sensor/estimator system



Steps 1 and 2 of our approach fail. Real-time coding may help.

Thank you.

Questions?