Nonstochastic Information Theory for Feedback Control

Girish Nair

Department of Electrical and Electronic Engineering
University of Melbourne
Australia

Dutch Institute of Systems and Control Summer School
Zandvoort aan Zee
The Netherlands
June 2015
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1. Basics of Shannon Information Theory
2. Overview of Capacity-Limited State Estimation/Control
3. Motivation for Nonstochastic Control
4. Uncertain Variables, Unrelatedness and Markovness
5. Nonstochastic Information
6. Channels and Coding Theorems
7. State Estimation and Control via Noisy Channels
Outline

1 Basics of Shannon Information Theory
2 Overview of Capacity-Limited State Estimation/Control
3 Motivation for Nonstochastic Control
4 Uncertain Variables, Unrelatedness and Markovness
5 Nonstochastic Information
6 Channels and Coding Theorems
7 State Estimation and Control via Noisy Channels
What is Information?

- Wikipedia:
  
  *Information is that which informs*
  
i.e. that from which data and knowledge can be derived (as data represents values attributed to parameters, while knowledge is acquired through understanding of real things or abstract concepts, in any particular field of study).

- In [Shannon BSTJ 1948], information was somewhat more concretely defined, within a probability space.
What is Information?

Wikipedia:

*Information is that which informs*
*i.e. that from which data and knowledge can be derived*
(as data represents values attributed to parameters, while knowledge is acquired through understanding of real things or abstract concepts, in any particular field of study).

In [Shannon BSTJ 1948], information was somewhat more concretely defined, within a probability space.
What is Information?

- Wikipedia:

  *Information is that which informs i.e. that from which data and knowledge can be derived (as data represents values attributed to parameters, while knowledge is acquired through understanding of real things or abstract concepts, in any particular field of study).*

- In [Shannon BSTJ 1948], information was somewhat more concretely defined, within a probability space.
What is Information?

- Wikipedia:
  
  Information is that which informs
  i.e. that from which data and knowledge can be derived
  (as data represents values attributed to parameters, while knowledge is acquired through understanding of real things or abstract concepts, in any particular field of study).

- In [Shannon BSTJ 1948], information was somewhat more concretely defined, within a probability space.
Shannon Entropy

- Prior uncertainty or *entropy* of a discrete random variable (rv) \( X \sim p_X \)

\[
H[X] := E \left[ \log_2 \left( \frac{1}{p_X(X)} \right) \right] = -\sum_x p_X(x) \log_2 p_X(x) \geq 0.
\]

- Minimum expected no. yes/no questions sufficient to determine \( X \).
- Joint (discrete) entropy \( H[X, Y] \) defined by replacing \( p_X \) with \( p_{X,Y} \).
- *Conditional entropy* of \( X \) given \( Y \)
  is average uncertainty in \( X \) given \( Y \):

\[
H[X|Y] := E \left[ \log_2 \left( \frac{1}{p_{X|Y}(X|Y)} \right) \right] = H[X, Y] - H[Y] \ ( \geq 0 ).
\]
Shannon Entropy

- Prior uncertainty or entropy of a discrete random variable (rv) $X \sim p_X$

$$H[X] := E \left[ \log_2 \left( \frac{1}{p_X(X)} \right) \right] = -\sum_x p_X(x) \log_2 p_X(x) \geq 0.$$  

- Minimum expected no. yes/no questions sufficient to determine $X$.

- Joint (discrete) entropy $H[X, Y]$ defined by replacing $p_X$ with $p_{X,Y}$.

- Conditional entropy of $X$ given $Y$

is average uncertainty in $X$ given $Y$:

$$H[X|Y] := E \left[ \log_2 \left( \frac{1}{p_{X|Y}(X|Y)} \right) \right] = H[X, Y] - H[Y] \ (\geq 0).$$
Shannon Entropy

- Prior uncertainty or *entropy* of a discrete random variable (rv) $X \sim p_X$

$$H[X] := E \left[ \log_2 \left( \frac{1}{p_X(X)} \right) \right] = - \sum_x p_X(x) \log_2 p_X(x) \geq 0.$$ 

- Minimum expected no. yes/no questions sufficient to determine $X$.

- Joint (discrete) entropy $H[X, Y]$ defined by replacing $p_X$ with $p_{X,Y}$.

- *Conditional entropy* of $X$ given $Y$
  - is average uncertainty in $X$ given $Y$:

$$H[X|Y] := E \left[ \log_2 \left( \frac{1}{p_{X|Y}(X|Y)} \right) \right] \equiv H[X, Y] - H[Y] \ (\geq 0).$$
Shannon Information

- Information gained about $X$ from $Y$ := Reduction in uncertainty:
  \[
  I[X; Y] := H[X] - H[X|Y].
  \]

- Called \textit{mutual information} since symmetric:
  \[
  I[X; Y] = -\sum_{x,y} p_{X,Y}(x,y) \log_2 \left( \frac{p_X(x)p_Y(y)}{p_{X,Y}(x,y)} \right)
  \equiv H[X] + H[Y] - H[X,Y].
  \]

- For continuous $X, Y$, replace pmf’s $p$ with corresponding pdf’s $f$. 
Shannon Information

- Information gained about $X$ from $Y$ := Reduction in uncertainty:

$$I[X; Y] := H[X] - H[X|Y].$$

- Called *mutual information* since symmetric:

$$I[X; Y] = - \sum_{x,y} p_{X,Y}(x,y) \log_2 \left( \frac{p_X(x)p_Y(y)}{p_{X,Y}(x,y)} \right),$$

$$\equiv H[X] + H[Y] - H[X, Y].$$

- For continuous $X, Y$, replace pmf’s $p$ with corresponding pdf’s $f$. 
Shannon Information

- Information gained about $X$ from $Y :=$ Reduction in uncertainty:
  \[ I[X; Y] := H[X] - H[X \mid Y]. \]

- Called *mutual information* since symmetric:
  \[
  I[X; Y] = - \sum_{x,y} p_{X,Y}(x,y) \log_2 \left( \frac{p_X(x)p_Y(y)}{p_{X,Y}(x,y)} \right) \\
  \equiv H[X] + H[Y] - H[X, Y].
  \]

- For continuous $X, Y$, replace pmf’s $p$ with corresponding pdf’s $f$. 
Codes for Stationary Memoryless Random Channels

\[ p_{Y_k|X_k}(y|x) = q(y|x) \]

\[ p_{Y_{0:n}|X_{0:n}}(y_{0:n}|x_{0:n}) = \prod_{k=0}^{n} q(y_k|x_k) \]

A block code is defined by

- an error tolerance \( \varepsilon > 0 \), block length \( n + 1 \in \mathbb{N} \) and message-set cardinality \( \mu \geq 1 \);
- an encoder mapping \( \gamma \) s.t. for any independent, uniformly distributed message \( M \in \{ m^1, \ldots, m^\mu \} \),
  \[ X_{0:n} = \gamma(i) \text{ if } M = m^i; \]
- and a decoder \( \hat{M} = \delta(Y_{0:n}) \) s.t. \( \Pr[\hat{M} \neq M] \leq \varepsilon \).
A block code is defined by

- an error tolerance \( \varepsilon > 0 \), block length \( n + 1 \in \mathbb{N} \) and message-set cardinality \( \mu \geq 1 \);
- an encoder mapping \( \gamma \) s.t. for any independent, uniformly distributed message \( M \in \{m^1, \ldots, m^\mu\} \),
  \[
  X_{0:n} = \gamma(i) \text{ if } M = m^i;
  \]
- and a decoder \( \hat{M} = \delta(Y_{0:n}) \) s.t. \( \Pr[\hat{M} \neq M] \leq \varepsilon \).
A block code is defined by
- an error tolerance $\varepsilon > 0$, block length $n + 1 \in \mathbb{N}$ and message-set cardinality $\mu \geq 1$;
- an encoder mapping $\gamma$ s.t. for any independent, uniformly distributed message $M \in \{ m^1, \ldots, m^\mu \}$,
\[
X_{0:n} = \gamma(i) \text{ if } M = m^i;
\]
- and a decoder $\hat{M} = \delta(Y_{0:n})$ s.t. $\Pr[\hat{M} \neq M] \leq \varepsilon$. 

Codes for Stationary Memoryless Random Channels

\[
p_{Y_k|X_k}(y|x) = q(y|x)
\]
\[
p_{Y_{0:n}|X_{0:n}}(y_{0:n}|x_{0:n}) = \prod_{k=0}^{n} q(y_k|x_k)
\]
A block code is defined by
- an error tolerance $\epsilon > 0$, block length $n + 1 \in \mathbb{N}$ and message-set cardinality $\mu \geq 1$;
- an encoder mapping $\gamma$ s.t. for any independent, uniformly distributed message $M \in \{m^1, \ldots, m^\mu\}$,
  $$X_{0:n} = \gamma(i) \text{ if } M = m^i;$$
- and a decoder $\hat{M} = \delta(Y_{0:n})$ s.t. $\Pr[\hat{M} \neq M] \leq \epsilon$. 
Define \textit{(ordinary) capacity} $C$ operationally as the highest block-coding rate that yields vanishing error probability:

$$C := \lim_{\epsilon \to 0} \sup_{n, \mu \in \mathbb{N}, \gamma, \delta} \frac{\log_2 \mu}{n + 1} = \lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{\mu \in \mathbb{N}, \gamma, \delta} \frac{\log_2 \mu}{n + 1}.$$

Shannon showed that capacity can also be thought of \textit{intrinsically}, as the maximum information rate across channel:

\textbf{Theorem (Shannon BSTJ 1948)}

$$C = \sup_{n \geq 0, p_{X^n}} \frac{I[X_0^n; Y_0^n]}{n + 1} \left( = \sup_{p_X} I[X; Y] \right).$$
Define (ordinary) capacity $C$ operationally as the highest block-coding rate that yields vanishing error probability:

$$C := \lim_{\varepsilon \to 0} \sup_{n, \mu \in \mathbb{N}, \gamma, \delta} \log_2 \mu \frac{1}{n+1} = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sup_{\mu \in \mathbb{N}, \gamma, \delta} \log_2 \mu \frac{1}{n+1} .$$

Shannon showed that capacity can also be thought of intrinsically, as the maximum information rate across channel:

**Theorem (Shannon BSTJ 1948)**

$$C = \sup_{n \geq 0, p_{X_0:n}} \frac{l[X_0:n; Y_0:n]}{n+1} \left( = \sup_{p_X} l[X; Y] \right)$$
Networked State Estimation/Control

- Classical assumption: controllers and estimators knew plant outputs perfectly.
- Since the 60’s this assumption has been challenged:
  - Delays, due to latency and intermittent channel access, in control area networks.
  - Quantisation errors in digital control,
  - Finite communication capacity per sensor in long-range radar surveillance networks
- Focus here on limited quantiser resolution and capacity.
Estimation/Control over Communication Channels

\[ Y_k = GX_k + W_k, \]
\[ X_{k+1} = AX_k + BU_k + V_k \]

Noise \( V_k, W_k \)

Decoder/Estimator

Channel

Decoder/Controller

Quantiser/Coder

Quantiser/Coder

Controller

Channel

Noise \( V_k, W_k \)
Additive Noise Model

- Early work considered errorless digital channels and static quantisers, with uniform quantiser errors modelled as additive, uncorrelated noise \([e.g.\ Curry\ 1970]\) with variance \(\propto 2^{-2R}\) \((R = \text{bit rate})\).

- Good approximation for stable plants and high \(R\), and allows linear stochastic estimation/control theory to be applied.

- However, for unstable plants it leads to conclusions that are wrong, \(e.g.\):
  - if plant is noiseless and unstable, then states/estimation errors cannot converge to 0;
  - and if plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.
Additive Noise Model

- Early work considered errorless digital channels and static quantisers, with uniform quantiser errors modelled as additive, uncorrelated noise [e.g. Curry 1970] with variance $\sim 2^{-2R}$ ($R =$ bit rate).
- Good approximation for stable plants and high $R$, and allows linear stochastic estimation/control theory to be applied.
- However, for unstable plants it leads to conclusions that are wrong, e.g.
  - if plant is noiseless and unstable, then states/estimation errors cannot converge to 0;
  - and if plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.
Additive Noise Model

- Early work considered errorless digital channels and static quantisers, with uniform quantiser errors modelled as additive, uncorrelated noise [e.g. Curry 1970] with variance $\sigma^2 = 2^{-2R}$ ($R =$ bit rate).

- Good approximation for stable plants and high $R$, and allows linear stochastic estimation/control theory to be applied.

- However, for unstable plants it leads to conclusions that are wrong, e.g.
  - if plant is noiseless and unstable, then states/estimation errors cannot converge to 0;
  - and if plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.
Additive Noise Model

- Early work considered errorless digital channels and static quantisers, with uniform quantiser errors modelled as additive, uncorrelated noise [e.g. Curry 1970] with variance \( \sigma^2 \) \( R \) (bit rate).

- Good approximation for stable plants and high \( R \), and allows linear stochastic estimation/control theory to be applied.

- However, for unstable plants it leads to conclusions that are wrong, e.g.
  - if plant is noiseless and unstable, then states/estimation errors cannot converge to 0;
  - and if plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.
In fact, coding-based analyses reveal that stable state estimation/control possible iff

\[ R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

Holds under various assumptions and stability notions:

- Random initial state, noiseless plant; mean \( r \)th power convergence to 0. [N.-Evans, Auto.03]
- Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
- Random plant noise; mean-square boundedness. [N.-Evans, SICON04]
- Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.

Sufficiency via explicit construction.
In fact, coding-based analyses reveal that stable state estimation/control possible iff

$$R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|,$$

where $$\lambda_1, \ldots, \lambda_n =$$ eigenvalues of plant matrix $$A$$.

Holds under various assumptions and stability notions:

- Random initial state, noiseless plant; mean $$r$$th power convergence to 0. [N.-Evans, Auto.03]
- Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
- Random plant noise; mean-square boundedness. [N.-Evans, SICON04]
- Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.

Sufficiency via explicit construction.
Errorless Channels - Data Rate Theorem

- In fact, coding-based analyses reveal that stable state estimation/control possible iff

\[ R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

- Holds under various assumptions and stability notions:
  - Random initial state, noiseless plant; mean \( r \)th power convergence to 0.[N.-Evans, Auto.03]
  - Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
  - Random plant noise; mean-square boundedness.[N.-Evans, SICON04]
  - Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

- Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.

- Sufficiency via explicit construction.
Errorless Channels - Data Rate Theorem

- In fact, coding-based analyses reveal that stable state estimation/control possible iff
  \[ R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

  where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

- Holds under various assumptions and stability notions:
  - Random initial state, noiseless plant; mean \( r \)th power convergence to 0. [N.-Evans, Auto.03]
  - Bounded initial state, noiseless plant; uniform convergence to 0
    [Tatikonda-Mitter, TAC04]
  - Random plant noise; mean-square boundedness. [N.-Evans, SICON04]
  - Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.
- Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.
- Sufficiency via explicit construction.
Errorless Channels - Data Rate Theorem

- In fact, coding-based analyses reveal that stable state estimation/control possible iff
  \[ R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]
  where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

- Holds under various assumptions and stability notions:
  - Random initial state, noiseless plant; mean \( r \)th power convergence to 0. [N.-Evans, Auto.03]
  - Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
  - Random plant noise; mean-square boundedness [N.-Evans, SICON04]
  - Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.
- Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.
- Sufficiency via explicit construction.
In fact, coding-based analyses reveal that stable state estimation/control possible iff

$$R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|,$$

where $\lambda_1, \ldots, \lambda_n = \text{eigenvalues of plant matrix } A$.

Holds under various assumptions and stability notions:

- Random initial state, noiseless plant; mean $r$th power convergence to 0. [N.-Evans, Auto.03]
- Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
- Random plant noise; mean-square boundedness. [N.-Evans, SICON04]
- Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.

Sufficiency via explicit construction.
Errorless Channels - Data Rate Theorem

- In fact, coding-based analyses reveal that stable state estimation/control possible iff

\[ R > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

- Holds under various assumptions and stability notions:
  - Random initial state, noiseless plant; mean \( r \)th power convergence to 0. [N.-Evans, Auto.03]
  - Bounded initial state, noiseless plant; uniform convergence to 0 [Tatikonda-Mitter, TAC04]
  - Random plant noise; mean-square boundedness. [N.-Evans, SICON04]
  - Bounded plant noise; uniform boundedness [Tatikonda-Mitter, TAC04]

- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

- Necessity typically proved using differential entropy power, quantisation theory or volume partitioning bounds.

- Sufficiency via explicit construction.
Noisy Channels

- ‘Stable’ states/estimation errors possible iff a suitable channel figure-of-merit (FoM) satisfies

\[
\text{FoM} > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|,
\]

where \(\lambda_1, \ldots, \lambda_n\) = eigenvalues of plant matrix \(A\).

- Unlike noiseless channel case, FoM depends critically on stability notion and noise model:
  - FoM = \(C - \) states/est. errors \(\rightarrow 0\) almost surely (a.s.) [Matveev-Savkin SIAM07], or mean-square bounded (MSB) states over AWGN channel [Braslavsky et al. TAC07]
  - FoM = \(C_{\text{any}} -\) MSB states over random discrete memoryless channels [Sahai-Mitter TIT06]
  - FoM = \(C_{0f}\) for control or \(C_0\) for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]

As \(C \geq C_{\text{any}} \geq C_{0f} \geq C_0\), these criteria generally do not coincide.
Noisy Channels

‘Stable’ states/estimation errors possible iff a suitable channel figure-of-merit (FoM) satisfies

$$\text{FoM} > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|,$$

where $\lambda_1, \ldots, \lambda_n$ = eigenvalues of plant matrix $A$.

Unlike noiseless channel case, FoM depends critically on stability notion and noise model:

- FoM = $C$ - states/est. errors $\to 0$ almost surely (a.s.) [Matveev-Savkin SIAM07], or mean-square bounded (MSB) states over AWGN channel [Braslavsky et al. TAC07]
- FoM = $C_{\text{any}}$ - MSB states over random discrete memoryless channels [Sahai-Mitter TIT06]
- FoM = $C_{0f}$ for control or $C_0$ for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]

As $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$, these criteria generally do not coincide.
Noisy Channels

- ‘Stable’ states/estimation errors possible iff a suitable channel figure-of-merit (FoM) satisfies

\[ \text{FoM} > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

where \( \lambda_1, \ldots, \lambda_n \) = eigenvalues of plant matrix \( A \).

- Unlike noiseless channel case, FoM depends critically on stability notion and noise model:
  - FoM = \( C \) - states/est. errors \( \rightarrow 0 \) almost surely (a.s.) [Matveev-Savkin SIAM07], or mean-square bounded (MSB) states over AWGN channel [Braslavsky et al. TAC07]
  - FoM = \( C_{\text{any}} \) - MSB states over random discrete memoryless channels [Sahai-Mitter TIT06]
  - FoM = \( C_{0f} \) for control or \( C_0 \) for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]

As \( C \geq C_{\text{any}} \geq C_{0f} \geq C_0 \), these criteria generally do not coincide.
Noisy Channels

- ‘Stable’ states/estimation errors possible iff a suitable channel figure-of-merit (FoM) satisfies

$$\text{FoM} > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|,$$

where $\lambda_1, \ldots, \lambda_n$ = eigenvalues of plant matrix $A$.

- Unlike noiseless channel case, FoM depends critically on stability notion and noise model:
  - FoM = $C$ - states/est. errors $\rightarrow 0$ almost surely (a.s.) [Matveev-Savkin SIAM07], or mean-square bounded (MSB) states over AWGN channel [Braslavsky et al. TAC07]
  - FoM = $C_{\text{any}}$ - MSB states over random discrete memoryless channels [Sahai-Mitter TIT06]
  - FoM = $C_{0f}$ for control or $C_0$ for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]

As $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$, these criteria generally do not coincide.
Noisy Channels

- ‘Stable’ states/estimation errors possible iff a suitable channel figure-of-merit (FoM) satisfies

\[ \text{FoM} > \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|, \]

where \(\lambda_1, \ldots, \lambda_n\) = eigenvalues of plant matrix \(A\).

- Unlike noiseless channel case, FoM depends critically on stability notion and noise model:
  - FoM = \(C\) - states/est. errors \(\rightarrow 0\) almost surely (a.s.) [Matveev-Savkin SIAM07], or mean-square bounded (MSB) states over AWGN channel [Braslavsky et al. TAC07]
  - FoM = \(C_{\text{any}}\) - MSB states over random discrete memoryless channels [Sahai-Mitter TIT06]
  - FoM = \(C_{0f}\) for control or \(C_0\) for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]

As \(C \geq C_{\text{any}} \geq C_{0f} \geq C_0\), these criteria generally do not coincide.
If the goal is MSB or a.s. convergence $\to 0$ of states/estimation errors, then information theory is crucial for finding lower bounds.

However, when the goal is a.s. bounded states/errors, classical information theory has played no role so far in networked estimation/control.

Yet information in some sense must be flowing across the channel, even without a probabilistic model/objective.
Questions

- Is there a meaningful theory of information for nonrandom variables?
- Can we construct an information-theoretic basis for networked estimation/control with nonrandom noise?
- Are there intrinsic, information-theoretic interpretations of $C_0$ and $C_{0f}$?
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electro-magnetic/optical.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electro-magnetic/optical.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electro-magnetic/optical.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electro-magnetic/optical.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electro-magnetic/optical.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Why Nonstochastic Anyway?

Long tradition in control of treating noise as nonrandom perturbation with bounded magnitude, energy or power:

- Control systems usually have mechanical/chemical components, as well as electrical.
  - Dominant disturbances may not be governed by known probability distributions.
  - E.g. in mechanical systems, main disturbance may be vibrations at resonant frequencies determined by machine dimensions and material properties.

- In contrast, communication systems are mainly electrical/electrical-optical/magnetic.
  - Dominant disturbances - thermal noise, shot noise, fading etc. - well-modelled by probability distributions derived from statistical/quantum physics.
Related to the previous points,

- In most digital comm. systems, bit periods $T_b \approx 2 \times 10^{-5}$ s or shorter.
  $\Rightarrow$ Thermal and shot noise ($\sigma \propto \sqrt{T_b}$) noticeable compared to detected signal amplitudes ($\propto T_b$).

- Control systems typically operate with longer sample or bit periods, $10^{-2}$ or $10^{-3}$ s.
  $\Rightarrow$ Thermal/shot noise negligible compared to signal amplitudes.
Why Nonstochastic Anyway? (cont.)

Related to the previous points,

- In most digital comm. systems, bit periods $T_b \approx 2 \times 10^{-5}$s or shorter.
  - ⇒ Thermal and shot noise ($\sigma \propto \sqrt{T_b}$) noticeable compared to detected signal amplitudes ($\propto T_b$).

- Control systems typically operate with longer sample or bit periods, $10^{-2}$ or $10^{-3}$s.
  - ⇒ Thermal/shot noise negligible compared to signal amplitudes.
For safety or mission-critical reasons, stability and performance guarantees often required *every time* a control system is used, if disturbances within rated bounds. Especially if plant is unstable or marginally stable. Or if we wish to interconnect several control systems and still be sure of performance.

In contrast, most consumer-oriented communications requires good performance only on average, or with high probability. Occasional violations of specifications permitted, and cannot be prevented within a probabilistic framework.
For safety or mission-critical reasons, stability and performance guarantees often required *every time* a control system is used, if disturbances within rated bounds. Especially if plant is unstable or marginally stable. Or if we wish to interconnect several control systems and still be sure of performance.

In contrast, most consumer-oriented communications requires good performance only on average, or with high probability. Occasional violations of specifications permitted, and cannot be prevented within a probabilistic framework.
For safety or mission-critical reasons, stability and performance guarantees often required *every time* a control system is used, if disturbances within rated bounds. Especially if plant is unstable or marginally stable. Or if we wish to interconnect several control systems and still be sure of performance.

In contrast, most consumer-oriented communications requires good performance only on average, or with high probability. Occasional violations of specifications permitted, and cannot be prevented within a probabilistic framework.
‘If there’s a fifty-fifty chance that something can go wrong, nine out of ten times, it will.’

(attrib. L. ‘Yogi’ Berra, former US baseball player)

(Photo from Wikipedia)
Uncertain Variable Formalism

- Define an *uncertain variable* \( (uv) X \) to be a mapping from an underlying sample space \( \Omega \) to a space \( X \).

- Each \( \omega \in \Omega \) may represent a specific combination of noise/input signals into a system, and \( X \) may represent a state/output variable.

- For a given \( \omega, x = X(\omega) \) is the *realisation* of \( X \).

- Unlike probability theory, no \( \sigma \)-algebra \( \in 2^\Omega \) or measure on \( \Omega \) is imposed.

- Assume \( \Omega \) is uncountable to accommodate continuous \( X \).
Uncertain Variable Formalism

- Define an *uncertain variable (uv)* $X$ to be a mapping from an underlying sample space $\Omega$ to a space $\mathbb{X}$.
- Each $\omega \in \Omega$ may represent a specific combination of noise/input signals into a system, and $X$ may represent a state/output variable.
- For a given $\omega$, $x = X(\omega)$ is the *realisation* of $X$.

Unlike probability theory, *no* $\sigma$-algebra $\subset 2^\Omega$ or measure on $\Omega$ is imposed.
- Assume $\Omega$ is uncountable to accommodate continuous $\mathbb{X}$.


**UV Formalism- Ranges and Conditioning**

- **Marginal range** $[X] := \{X(\omega) : \omega \in \Omega\} \subseteq X$.
- **Joint range** $[X, Y] := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq X \times Y$.
- **Conditional range** $[X|y] := \{X(\omega) : Y(\omega) = y, \omega \in \Omega\}$.

In the absence of statistical structure, the joint range fully characterises the relationship between $X$ and $Y$. Note

$$[X, Y] = \bigcup_{y \in [Y]} [X|y] \times \{y\},$$

i.e. joint range is given by the conditional and marginal, analogously to probability theory.
Independence Without Probability

Definition

The uv’s $X, Y$ are called (mutually) unrelated if

$$[X, Y] = [X] \times [Y],$$

(1)

denoted $X \perp Y$. Else called related.

- Equivalent characterisation:

Proposition

The uv’s $X, Y$ unrelated if

$$[X|y] = [X], \quad \forall y \in [Y].$$

(2)

- Unrelatedness is equivalent to $X$ and $Y$ inducing qualitatively independent [Rényi’70] partitions of $\Omega$ when $\Omega$ is finite.
Examples of Relatedness and Unrelatedness

a) \( X, Y \) related

\[
\begin{align*}
[Y|X'] &\subseteq [Y] \\
[X'|Y] &\subseteq [X] \\
[X] &\subseteq [Y]
\end{align*}
\]

b) \( X, Y \) unrelated

\[
\begin{align*}
[Y] &= [Y|X'] \\
[X] &= [X|Y']
\end{align*}
\]
Markovness without Probability

Definition

\( X, Y, Z \) said to form a Markov uncertainty chain \( X - Y - Z \) if

\[
[X|y, z] = [X|y], \quad \forall (y, z) \in [Y, Z].
\]  

(3)

- Equivalent to

\[
[X, Z|y] = [X|y] \times [Z|y], \quad \forall y \in [Y],
\]

i.e. \( X, Z \) conditionally unrelated given \( Y \), or in other words \( X \perp Z|Y \).

- \( X, Y, Z \) said to form a conditional Markov uncertainty chain given \( W \) if \( X - (Y, W) - Z \).

Also write as \( X - Y - Z|W \) or \( X \perp Z|(Y, W) \).
Markovness without Probability

**Definition**

$X, Y, Z$ said to form a Markov uncertainty chain $X \rightarrow Y \rightarrow Z$ if

$$[X|y, z] = [X|y], \ \forall (y, z) \in [Y, Z].$$

(3)

- Equivalent to

  $$[X, Z|y] = [X|y] \times [Z|y], \ \forall y \in [Y],$$

  i.e. $X, Z$ *conditionally unrelated given* $Y$, or in other words $X \perp Z|Y$.

- $X, Y, Z$ said to form a *conditional Markov uncertainty chain given* $W$ if $X \rightarrow (Y, W) \rightarrow Z$.
  Also write as $X \rightarrow Y \rightarrow Z|W$ or $X \perp Z|(Y, W)$.

Nair (Uni. Melbourne) Nonstochastic Information DISC School 2015 23 / 68
Markovness without Probability

Definition

\[ X, Y, Z \text{ said to form a Markov uncertainty chain } X - Y - Z \text{ if} \]

\[ [X|y, z] = [X|y], \forall (y, z) \in [Y, Z]. \quad (3) \]

- Equivalent to

\[ [X, Z|y] = [X|y] \times [Z|y], \forall y \in [Y], \]

i.e. \( X, Z \text{ conditionally unrelated given } Y \), or in other words \( X \perp Z|Y \).

- \( X, Y, Z \text{ said to form a conditional Markov uncertainty chain given } W \text{ if } X - (Y, W) - Z. \)
  Also write as \( X - Y - Z|W \) or \( X \perp Z|(Y, W) \).
Definition

Two points \((x, y), (x', y') \in [X, Y]\) are called taxicab connected \((x, y) \leftrightarrow (x', y')\) if \(\exists\) a sequence

\[(x, y) = (x_1, y_1), (x_2, y_2), \ldots, (x_{n-1}, y_{n-1}), (x_n, y_n) = (x', y')\]

of points in \([X, Y]\) such that each point differs in only one coordinate from its predecessor.

- Not hard to see that \(\leftrightarrow\) is an equivalence relation on \([X, Y]\).
- Call its equivalence classes a taxicab partition \(\mathcal{T}[X; Y]\) of \([X, Y]\).
- Define a nonstochastic information index

\[I_\ast[X; Y] := \log_2 |\mathcal{T}[X; Y]| \in [0, \infty]. \quad (4)\]
Information without Probability

Definition

Two points \((x, y), (x', y') \in [X, Y]\) are called taxicab connected \((x, y) \leftrightarrow (x', y')\) if \(\exists\) a sequence

\[
(x, y) = (x_1, y_1), (x_2, y_2), \ldots, (x_{n-1}, y_{n-1}), (x_n, y_n) = (x', y')
\]

of points in \([X, Y]\) such that each point differs in only one coordinate from its predecessor.

- Not hard to see that \(\leftrightarrow\) is an equivalence relation on \([X, Y]\).
- Call its equivalence classes a taxicab partition \(\mathcal{T}[X; Y]\) of \([X, Y]\).
- Define a nonstochastic information index

\[
I_*[X; Y] := \log_2 |\mathcal{T}[X; Y]| \in [0, \infty]. \tag{4}
\]
Connection to Common Random Variables

- $\mathcal{T}[X; Y]$ also called *ergodic decomposition* [Gács-Körner PCIT72].

- For discrete $X, Y$, the elements of $\mathcal{T}[X; Y]$ are the *connected components* of [Wolf-Wullschleger itw04], which were shown there to be the maximal *common rv* $Z_*$, i.e.
  - $Z_* = f_*(X) = g_*(Y)$ under suitable mappings $f_*, g_*$
    (since points in distinct sets in $\mathcal{T}[X; Y]$ are not taxicab-connected)
  - If another rv $Z \equiv f(X) \equiv g(Y)$, then $Z \equiv k(Z_*)$
    (since all points in the same set in $\mathcal{T}[X; Y]$ are taxicab-connected)

- Not hard to see that $Z_*$ also has the largest no. distinct values of any common rv $Z \equiv f(X) \equiv g(Y)$.

- $I_*[X; Y] = \text{Hartley entropy of } Z_*$.

- Maximal common rv’s first described in the brief paper ‘The lattice theory of information’ [Shannon TIT53].
Connection to Common Random Variables

- $\mathcal{T}[X; Y]$ also called *ergodic decomposition* [Gács-Körner PCIT72].
- For discrete $X, Y$, the elements of $\mathcal{T}[X; Y]$ are the *connected components* of [Wolf-Wullschleger itw04], which were shown there to be the maximal *common rv* $Z^*$, i.e.
  - $Z^* = f^*(X) = g^*(Y)$ under suitable mappings $f^*, g^*$
    (since points in distinct sets in $\mathcal{T}[X; Y]$ are not taxicab-connected)
  - If another rv $Z \equiv f(X) \equiv g(Y)$, then $Z \equiv k(Z^*)$
    (since all points in the same set in $\mathcal{T}[X; Y]$ are taxicab-connected)
- Not hard to see that $Z^*$ also has the largest no. distinct values of any common rv $Z \equiv f(X) \equiv g(Y)$.
- $I^*[X; Y] = \text{Hartley entropy of } Z^*$.
- Maximal common rv’s first described in the brief paper *‘The lattice theory of information’* [Shannon TIT53].
Examples

\[ |\mathcal{T}| = 2 = \max \text{. # distinct values} \]
that can always be agreed on
from separate observations of \(X \& Y\).

\[ |\mathcal{T}| = 1 = \max \text{. # distinct values} \]
that can always be agreed on
from separate observations of \(X \& Y\).
Equivalent View via Overlap Partitions

- As in probability, often easier to work with conditional rather than joint ranges.
- Let \([X|Y] := \{[X|y] : y \in [Y]\}\) be the conditional range family.

**Definition**

*Two points* \(x, x'\) *are called* \([X|Y]\)*-overlap-connected* *if* \(\exists\) *a sequence of sets* \(B_1, \ldots, B_n \in [X|Y]\) *s.t.*

- \(x \in B_1\) *and* \(x' \in B_n\)
- \(B_i \cap B_{i+1} \neq \emptyset, \forall i \in [1 : n - 1]\).

- Overlap connectedness is an equivalence relation on \([X]\), induced by \([X|Y]\).
- Let the *overlap partition* \([X|Y]*\ of \([X]\) denote the equivalence classes.
Equivalent View via Overlap Partitions

- As in probability, often easier to work with conditional rather than joint ranges.
- Let \([X|Y] := \{[X|y] : y \in [Y]\}\) be the conditional range family.

**Definition**

Two points \(x, x'\) are called \([X|Y]\)-overlap-connected if \(\exists\) a sequence of sets \(B_1, \ldots, B_n \in [X|Y]\) s.t.

- \(x \in B_1\) and \(x' \in B_n\)
- \(B_i \cap B_{i+1} \neq \emptyset, \forall i \in [1 : n - 1]\).

- Overlap connectedness is an equivalence relation on \([X]\), induced by \([X|Y]\).
- Let the overlap partition \([X|Y]^*\) of \([X]\) denote the equivalence classes.
Equivalent View via Overlap Partitions (cont.)

**Proposition**

For any uv’s $X, Y,$

$$I_*[X; Y] = \log_2 |[X|Y]_*|.$$  \hspace{1cm} (5)

**Proof Sketch:**

- For any two points $(x, y), (x', y') \in [X, Y], (x, y) \leftrightarrow (x', y') \iff x'$ and $x'$ are $[X|Y]$-overlap-connected.

- This allows us to set up a bijection between the partitions $\mathcal{T}[X; Y]$ and $[X|Y]_*$.

$\Rightarrow \mathcal{T}[X; Y]$ and $[X|Y]_*$ must have the same cardinality.
Proposition

For any uv’s $X, Y$,

$$I_*[X; Y] = \log_2 |[X|Y]_*|.$$  \hspace{1cm} (5)

Proof Sketch:

- For any two points $(x, y), (x', y') \in [X, Y]$, $(x, y) \leftrightarrow (x', y')$ iff $x'$ and $x'$ are $[X|Y]$-overlap-connected.

- This allows us to set up a bijection between the partitions $\mathcal{T}[X; Y]$ and $[X|Y]_*$.

$\Rightarrow$ $\mathcal{T}[X; Y]$ and $[X|Y]_*$ must have the same cardinality.
Equivalent View via Overlap Partitions (cont.)

Proposition

For any uv’s $X, Y,$

$$I_\ast[X; Y] = \log_2 |[X|Y]_\ast|.$$  \hspace{1cm} (5)

Proof Sketch:

- For any two points $(x, y), (x', y') \in [X, Y]$, $(x, y) \leftrightarrow (x', y')$ iff $x'$ and $x'$ are $[X|Y]$-overlap-connected.

- This allows us to set up a bijection between the partitions $\mathcal{T}[X; Y]$ and $[X|Y]_\ast$.

- $\Rightarrow \mathcal{T}[X; Y]$ and $[X|Y]_\ast$ must have the same cardinality.
Properties of $I_*$

- (Nonnegativity) $I_*[X; Y] \geq 0$ (obvious)
- (Symmetry) $I_*[X; Y] = I_*[Y; X]$. Follows from the fact that

\[(x, y) \leftrightarrow (x', y') \in [X, Y] \iff (y, x) \leftrightarrow (y', x') \in [Y, X]. \quad (6)\]

From this property and (5), knowing just one of the conditional range families $[X | Y]$ or $[Y | X]$ is enough to determine $I_*[X; Y]$.

Not like ordinary mutual information.
Properties of $I_\ast$ (cont.)

**Proposition (Monotonicity)**

*For any uv’s $X, Y, Z$,*

$$I_\ast[X; Y, Z] \geq I_\ast[X; Y].$$  \hspace{1cm} (7)

*Proof:* Idea is to find a surjection from $[X|Y, Z] \rightarrow [X|Y]$. This would automatically imply that the latter cannot have greater cardinality.

- Pick any set $B \in [X|Y, Z]$ and choose a $B' \in [X|Y]$ s.t. $B \cap B' \neq \emptyset$.
- At least one such $B'$ exists, since $[X|Y]$ covers $[X] \supset B$. 

Nair (Uni. Melbourne) Nonstochastic Information DISC School 2015 30 / 68
Properties of \( I_* \) (cont.)

### Proposition (Monotonicity)

For any uv’s \( X, Y, Z \),

\[
I_*[X; Y, Z] \geq I_*[X; Y].
\]  \( (7) \)

**Proof:** Idea is to find a surjection from \([X|Y, Z]_* \rightarrow [X|Y]_*\). This would automatically imply that the latter cannot have greater cardinality.

- Pick any set \( B \in [X|Y, Z]_* \) and choose a \( B' \in [X|Y]_* \) s.t. \( B \cap B' \neq \emptyset \).
- At least one such \( B' \) exists, since \([X|Y]_* \) covers \([X] \supseteq B\).
Proof of Monotonic Property (cont.)

Furthermore, exactly one such intersecting $B' \in [X|Y]*$ exists for each $B \in [X|Y,Z]*$, since $B \subseteq B'$:

- By definition, any $x \in B$ and $x' \in B \cap B'$ are connected by a sequence of successively overlapping sets in $[X|Y,Z]$.
- As $[X|y,z] \subseteq [X|y]$, $x,x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
- But $B' =$ all pts. that are $[X|Y]$-overlap connected with the representative pt. $x' \in B'$, so $x \in B'$.
- As $x$ was arbitrary, $B \subseteq B'$.

Thus $B \mapsto B'$ is a well-defined map from $[X|Y,Z]* \rightarrow [X|Y]*$.

It is also onto since, as noted before, every set $B' \in [X|Y]$ intersects some $B$ in $[X|Y,Z]*$, which covers $[X]$.

So $B \mapsto B'$ is the required surjection from $[X|Y,Z]* \rightarrow [X|Y]*$. □
Proof of Monotonic Property (cont.)

Furthermore, exactly one such intersecting $B' \in [X|Y]^*$ exists for each $B \in [X|Y, Z]^*$, since $B \subseteq B'$:

- By definition, any $x \in B$ and $x' \in B \cap B'$ are connected by a sequence of successively overlapping sets in $[X|Y, Z]$.
- As $[X|y, z] \subseteq [X|y]$, $x, x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
- But $B' = \text{all pts. that are } [X|Y]-\text{overlap connected with the representative pt. } x' \in B'$, so $x \in B'$.
- As $x$ was arbitrary, $B \subseteq B'$.

Thus $B \mapsto B'$ is a well-defined map from $[X|Y, Z]^* \rightarrow [X|Y]^*$.

It is also onto since, as noted before, every set $B' \in [X|Y]^*$ intersects some $B$ in $[X|Y, Z]^*$, which covers $[X]$.

So $B \mapsto B'$ is the required surjection from $[X|Y, Z]^* \rightarrow [X|Y]^*$. □
Proof of Monotonic Property (cont.)

Furthermore, exactly one such intersecting $B' \in [X|Y]_*$ exists for each $B \in [X|Y, Z]_*$, since $B \subseteq B'$:

- By definition, any $x \in B$ and $x' \in B \cap B'$ are connected by a sequence of successively overlapping sets in $[X|Y, Z]$.
- As $[X|y, z] \subseteq [X|y]$, $x, x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
- But $B' = \text{all pts. that are } [X|Y]\text{-overlap connected with the representative pt. } x' \in B'$, so $x \in B'$.
- As $x$ was arbitrary, $B \subseteq B'$.

Thus $B \mapsto B'$ is a well-defined map from $[X|Y, Z]_* \rightarrow [X|Y]_*$. It is also onto since, as noted before, every set $B' \in [X|Y]_*$ intersects some $B$ in $[X|Y, Z]_*$, which covers $[X]$. So $B \mapsto B'$ is the required surjection from $[X|Y, Z]_* \rightarrow [X|Y]_*$. □
Proof of Monotonic Property (cont.)

Furthermore, exactly one such intersecting $B' \in [X|Y]_*$ exists for each $B \in [X|Y, Z]_*$, since $B \subseteq B'$:

- By definition, any $x \in B$ and $x' \in B \cap B'$ are connected by a sequence of successively overlapping sets in $[X|Y, Z]$.
- As $[X|y, z] \subseteq [X|y]$, $x, x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
- But $B' = $ all pts. that are $[X|Y]$-overlap connected with the representative pt. $x' \in B'$, so $x \in B'$.
- As $x$ was arbitrary, $B \subseteq B'$.

Thus $B \mapsto B'$ is a well-defined map from $[X|Y, Z]_* \rightarrow [X|Y]_*$. It is also onto since, as noted before, every set $B' \in [X|Y]_*$ intersects some $B$ in $[X|Y, Z]_*$, which covers $[X]$. So $B \mapsto B'$ is the required surjection from $[X|Y, Z]_* \rightarrow [X|Y]_*$. □
Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $B' \in [X|Y]_*$ exists for each $B \in [X|Y, Z]_*$, since $B \subseteq B'$:
  - By definition, any $x \in B$ and $x' \in B \cap B'$ are connected by a sequence of successively overlapping sets in $[X|Y, Z]$.
  - As $[X|y, z] \subseteq [X|y]$, $x, x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
  - But $B' = \text{all pts. that are } [X|Y]-\text{overlap connected with the representative pt. } x' \in B'$, so $x \in B'$.
  - As $x$ was arbitrary, $B \subseteq B'$.

Thus $B \mapsto B'$ is a well-defined map from $[X|Y, Z]_* \rightarrow [X|Y]_*$. It is also onto since, as noted before, every set $B' \in [X|Y]_*$ intersects some $B$ in $[X|Y, Z]_*$, which covers $[X]$. So $B \mapsto B'$ is the required surjection from $[X|Y, Z]_* \rightarrow [X|Y]_*$. \qed
Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathcal{B}' \in [X|Y]_\ast$ exists for each $\mathcal{B} \in [X|Y, Z]_\ast$, since $\mathcal{B} \subseteq \mathcal{B}'$:
  - By definition, any $x \in \mathcal{B}$ and $x' \in \mathcal{B} \cap \mathcal{B}'$ are connected by a sequence of successively overlapping sets in $[X|Y, Z]$.
  - As $[X|y, z] \subseteq [X|y]$, $x, x'$ are also connected by a sequence of successively overlapping sets in $[X|Y]$.
  - But $\mathcal{B}' = \text{all pts. that are } [X|Y]-\text{overlap connected with the representative pt. } x' \in \mathcal{B}'$, so $x \in \mathcal{B}'$.
  - As $x$ was arbitrary, $\mathcal{B} \subseteq \mathcal{B}'$.

- Thus $\mathcal{B} \mapsto \mathcal{B}'$ is a well-defined map from $[X|Y, Z]_\ast \rightarrow [X|Y]_\ast$.
  - It is also onto since, as noted before, every set $\mathcal{B}' \in [X|Y]_\ast$ intersects some $\mathcal{B}$ in $[X|Y, Z]_\ast$, which covers $[X]$.
  - So $\mathcal{B} \mapsto \mathcal{B}'$ is the required surjection from $[X|Y, Z]_\ast \rightarrow [X|Y]_\ast$. \(\square\)
Proof of Monotonic Property (cont.)

Furthermore, exactly one such intersecting $\mathbb{B}' \in \mathbb{[X|Y]}_*$ exists for each $\mathbb{B} \in \mathbb{[X|Y, Z]}_*$, since $\mathbb{B} \subseteq \mathbb{B}'$:

- By definition, any $x \in \mathbb{B}$ and $x' \in \mathbb{B} \cap \mathbb{B}'$ are connected by a sequence of successively overlapping sets in $\mathbb{[X|Y, Z]}$.
- As $\mathbb{[X|y, z]} \subseteq \mathbb{[X|y]}$, $x, x'$ are also connected by a sequence of successively overlapping sets in $\mathbb{[X|Y]}$.
- But $\mathbb{B}' = \text{all pts. that are } \mathbb{[X|Y]}\text{-overlap connected with the representative pt. } x' \in \mathbb{B}'$, so $x \in \mathbb{B}'$.
- As $x$ was arbitrary, $\mathbb{B} \subseteq \mathbb{B}'$.

Thus $\mathbb{B} \mapsto \mathbb{B}'$ is a well-defined map from $\mathbb{[X|Y, Z]}_* \rightarrow \mathbb{[X|Y]}_*$.

It is also onto since, as noted before, every set $\mathbb{B}' \in \mathbb{[X|Y]}_*$ intersects some $\mathbb{B}$ in $\mathbb{[X|Y, Z]}_*$, which covers $\mathbb{[X]}$.

So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $\mathbb{[X|Y, Z]}_* \rightarrow \mathbb{[X|Y]}_*$. □
Properties of $I_*$ (cont.)

Proposition (Data Processing)

For Markov uncertainty chains $X - Y - Z$ (3),

$$I_*[X; Z] \leq I_*[X; Y].$$

Proof:

- By monotonicity and the overlap partition characterisation of $I_*$,

$$I_*[X; Z] \overset{(7)}{=} I_*[X; Y, Z] \overset{(5)}{=} \log |\mathcal{J}_{X|Y, Z}|.$$  

- By Markovness (3), $[X|y, z] = [X|y], \forall y \in [Y]$ and $z \in [Z|y]$.

- $\Rightarrow [X|Y, Z] = [X|Y]$.

- $\Rightarrow [X|Y, Z]^* = [X|Y]^*$.

- Substitute into the RHS of the equation above. $\square$
Properties of \( I_* \) (cont.)

### Proposition (Data Processing)

*For Markov uncertainty chains \( X \rightarrow Y \rightarrow Z \) (3),*

\[
I_*[X; Z] \leq I_*[X; Y].
\]

**Proof:**

- By monotonicity and the overlap partition characterisation of \( I_* \),

\[
I_*[X; Z] \overset{(7)}{\leq} I_*[X; Y, Z] \overset{(5)}{=} \log \left| [X|Y, Z]_* \right|.
\]

- By Markovness (3), \([X|y, z] = [X|y], \forall y \in [Y] \text{ and } z \in [Z|y].\)

\[\Rightarrow [X|Y, Z] = [X|Y].\]

\[\Rightarrow [X|Y, Z]_* = [X|Y]_* .\]

- Substitute into the RHS of the equation above. \( \Box \)
Properties of $I_*$ (cont.)

**Proposition (Data Processing)**

*For Markov uncertainty chains $X \rightarrow Y \rightarrow Z$ (3),*

$$I_*[X; Z] \leq I_*[X; Y].$$

**Proof:**

- By monotonicity and the overlap partition characterisation of $I_*$,

$$I_*[X; Z] \overset{(7)}{\leq} I_*[X; Y, Z] \overset{(5)}{=} \log |\mathbb{I}[X|Y, Z]^*|.$$ (8)

- By Markovness (3), $[X|y, z] = [X|y]$, $\forall y \in [Y]$ and $z \in [Z|y]$.

- $\Rightarrow [X|Y, Z] = [X|Y]$.

- $\Rightarrow [X|Y, Z]^* = [X|Y]^*$.

- Substitute into the RHS of the equation above. \(\square\)
Properties of $I_\star$ (cont.)

**Proposition (Data Processing)**

*For Markov uncertainty chains $X \rightarrow Y \rightarrow Z$ (3),*

$$I_\star[X;Z] \leq I_\star[X;Y].$$

**Proof:**

- By monotonicity and the overlap partition characterisation of $I_\star$,

$$I_\star[X;Z] \overset{(7)}{\leq} I_\star[X;Y,Z] \overset{(5)}{=} \log||[X|Y,Z]_*||. \quad (8)$$

- By Markovness (3), $[X|y,z] = [X|y]$, $\forall y \in [Y]$ and $z \in [Z|y]$.

- $\Rightarrow [X|Y,Z] = [X|Y]$.

- $\Rightarrow [X|Y,Z]_* = [X|Y]*$.

- Substitute into the RHS of the equation above. □
Properties of $I_*$ (cont.)

Proposition (Data Processing)

For Markov uncertainty chains $X \rightarrow Y \rightarrow Z$ (3),

$$I_*[X;Z] \leq I_*[X;Y].$$

Proof:

- By monotonicity and the overlap partition characterisation of $I_*$,

$$I_*[X;Z] \overset{(7)}{=} I_*[X;Y,Z] \overset{(5)}{=} \log|\mathbb{K}[X|Y,Z]|.$$  \hfill (8)

- By Markovness (3), $\mathbb{K}[X|y,z] = \mathbb{K}[X|y], \forall y \in \mathbb{K}[Y]$ and $z \in \mathbb{K}[Z|y]$.

- $\Rightarrow \mathbb{K}[X|Y,Z] = \mathbb{K}[X|Y]$.

- $\Rightarrow \mathbb{K}[X|Y,Z]_* = \mathbb{K}[X|Y]_*$.

- Substitute into the RHS of the equation above. $\square$
An *uncertain signal* $X$ is a mapping from $\Omega$ to the space $X^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $X$.

A stationary memoryless *uncertain* channel may be defined in terms of

- input and output spaces $X, Y$;
- a set-valued transition function $T : X \rightarrow 2^Y$;
- and the family of all uncertain input-output signal pairs $(X, Y)$ s.t.

$$\left[ Y_k | x_{0:k}, y_{0:k-1} \right] = \left[ Y_k | x_{k} \right] = T(x_k), \quad k \in \mathbb{Z}_{\geq 0}. \quad (9)$$

If channel ‘used without feedback’, then impose the extra constraint

$$\left[ X_k | x_{0:k-1}, y_{0:k-1} \right] = \left[ X_k | x_{0:k-1} \right], \quad k \in \mathbb{Z}_{\geq 0}, \quad (10)$$
on $(X, Y)$. 

Nair (Uni. Melbourne)  Nonstochastic Information  DISC School 2015  33 / 68
Stationary Memoryless Uncertain Channels - Take 1

- An *uncertain signal* $X$ is a mapping from $\Omega$ to the space $X^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $X$.

- A stationary memoryless *uncertain* channel may be defined in terms of
  
  - input and output spaces $X, Y$;
  - a set-valued transition function $T : X \rightarrow 2^Y$;
  - and the family of all uncertain input-output signal pairs $(X, Y)$ s.t.

$$\left[ Y_k \mid x_0:k, y_0:k-1 \right] = \left[ Y_k \mid x_k \right] = T(x_k), \quad k \in \mathbb{Z}_{\geq 0}. \quad (9)$$

- If channel ‘used without feedback’, then impose the extra constraint

$$\left[ X_k \mid x_0:k-1, y_0:k-1 \right] = \left[ X_k \mid x_0:k-1 \right], \quad k \in \mathbb{Z}_{\geq 0}, \quad (10)$$

on $(X, Y)$. 
An uncertain signal $X$ is a mapping from $\Omega$ to the space $X^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $X$.

A stationary memoryless uncertain channel may be defined in terms of

- input and output spaces $X, Y$;
- a set-valued transition function $T : X \rightarrow 2^Y$;
- and the family of all uncertain input-output signal pairs $(X, Y)$ s.t.

$$[[Y_k | x_{0:k}, y_{0:k-1}] = [[Y_k | x_k]] = T(x_k), \ k \in \mathbb{Z}_{\geq 0}. \ (9)$$

If channel ‘used without feedback’, then impose the extra constraint

$$[[X_k | x_{0:k-1}, y_{0:k-1}] = [[X_k | x_{0:k-1}], \ k \in \mathbb{Z}_{\geq 0}, \ (10)$$
on $(X, Y)$. 

Nair (Uni. Melbourne)
Channel Noise?

- Previous formulation parallels [Massey isit90] for stationary memoryless stochastic channels:

$$f_{Y_k|X_{0:k},Y_{0:k-1}} (y_k|x_{0:k},y_{0:k-1}) = f_{Y_k|X_k} (y_k|x_k) \equiv q(y_k,x_k).$$

- In many cases, it is enough to think in terms of these conditional ranges. Channel noise implicit.

- However, in many cases it is convenient to model channel noise explicitly. E.g.
  - when the transmitter has access to some function of past channel noise, not just past channel outputs,
  - or when the channel is part of a larger system, with other input and noise signals.

In this case, previous formulation would have to be changed to include the other terms in the conditioning arguments.
Channel Noise?

- Previous formulation parallels [Massey isit90] for stationary memoryless stochastic channels:

\[ f_{Y_k|X_{0:k}, Y_{0:k-1}}(y_k|x_{0:k}, y_{0:k-1}) = f_{Y_k|X_k}(y_k|x_k) \equiv q(y_k, x_k). \]

- In many cases, it is enough to think in terms of these conditional ranges. Channel noise implicit.

- However, in many cases it is convenient to model channel noise explicitly. E.g.
  - when the transmitter has access to some function of past channel noise, not just past channel outputs,
  - or when the channel is part of a larger system, with other input and noise signals.

  In this case, previous formulation would have to be changed to include the other terms in the conditioning arguments.
Definition

A stationary memoryless uncertain channel (SMUC) consists of

- an unrelated, identically spread (uis) noise signal $V = (V_k)_{k=0}^\infty$ taking values over a space $\mathbb{V}$, i.e.

$$[V_k | v_{0:k-1}] = [V_k] = \mathbb{V}, \quad \forall v_{0:k-1} \in \mathbb{V}^k, k \in \mathbb{Z}_{\geq 0};$$

- input and output spaces $X, Y$, and a transition function $\tau : X \times \mathbb{V} \rightarrow Y$;

- and the family $G$ of all uncertain input-output signal pairs $(X, Y)$ s.t. $\forall k \in \mathbb{Z}_{\geq 0}$,
  - $Y_k = \tau(X_k, V_k)$,
  - and $X_{0:k} \perp V_k$

If channel used w/o feedback, then tighten last condition so that $X \perp V$. Yields smaller family $G_{nf} \subset G$. 

Nair (Uni. Melbourne)  Nonstochastic Information  DISC School 2015  35 / 68
Zero Error Coding in UV Framework (No Feedback)

A zero-error code w/o feedback is defined by

- a block length $n + 1 \in \mathbb{N}$;
- a message cardinality $\mu \geq 1$;
- and an encoder mapping $\gamma : [1 : \mu] \rightarrow \mathbb{X}^{n+1}$, s.t. for any $M \perp V$ taking $\mu$ distinct values $m_1, \ldots, m_\mu$,
  
  - $X_{0:n} = \gamma(i) \text{ if } M = m_i$.
  - $\|M|Y_{0:n}\| = 1, \forall Y_{0:n} \in \|Y_{0:n}\|$.

Last condition equivalent to existence of a decoder that always maps $Y_{0:n} \leftrightarrow M$, despite channel noise.
A zero-error code w/o feedback is defined by

- a block length \( n + 1 \in \mathbb{N} \);
- a message cardinality \( \mu \geq 1 \);
- and an encoder mapping \( \gamma : [1 : \mu] \rightarrow \mathbb{X}^{n+1} \), s.t. for any \( M \perp V \) taking \( \mu \) distinct values \( m^1, \ldots, m^\mu \),
  - \( X_0:n = \gamma(i) \) if \( M = m^i \).
  - \( \| [{M|Y_0:n}] \| = 1, \forall y_0:n \in [{Y_0:n}] \).

Last condition equivalent to existence of a decoder that always maps \( Y_0:n \mapsto M \), despite channel noise.
Zero Error Capacity and $I_*$

Zero-error capacity $C_0$ defined *operationally*, as the highest block-coding rate that yields zero errors:

$$C_0 := \sup_{n, \mu \in \mathbb{N}, \gamma_1 : n} \frac{\log_2 \mu}{n + 1} = \lim_{n \to \infty} \sup_{\mu \in \mathbb{N}, \gamma_1 : n} \frac{\log_2 \mu}{n + 1}. \quad (12)$$

Theorem (after N. TAC13)

$$C_0 = \sup_{n \geq 0, (X, Y) \in \mathcal{G}_{nf}} \frac{I_*[X_0 : n ; Y_0 : n]}{n + 1} \left(= \lim_{n \to \infty} \sup_{(X, Y) \in \mathcal{G}_{nf}} \frac{I_*[X_0 : n ; Y_0 : n]}{n + 1} \right). \quad (13)$$

- In [Wolf-Wullschleger itw04], $C_0$ was characterised as the largest Shannon entropy rate of the maximal rv $Z_n$ common to discrete $X_0 : n$, $Y_0 : n$.
- Key idea is similar, but nonstochastic and applicable to continuous-valued $X$, $Y$. 
Zero Error Capacity and $I_*$

Zero-error capacity $C_0$ defined operationally, as the highest block-coding rate that yields zero errors:

$$C_0 := \operatorname{sup}_{n, \mu \in \mathbb{N}, Y_n} \frac{\log_2 \mu}{n + 1} = \lim_{n \to \infty} \operatorname{sup}_{\mu \in \mathbb{N}, Y_n} \frac{\log_2 \mu}{n + 1}. \quad (12)$$

Theorem (after N. TAC13)

$$C_0 = \operatorname{sup}_{n \geq 0, (X,Y) \in \mathcal{G}_{nf}} \frac{I_*[X_0:n; Y_0:n]}{n + 1} = \lim_{n \to \infty} \operatorname{sup}_{(X,Y) \in \mathcal{G}_{nf}} \frac{I_*[X_0:n; Y_0:n]}{n + 1}. \quad (13)$$

- In [Wolf-Wullschleger itw04], $C_0$ was characterised as the largest Shannon entropy rate of the maximal rv $Z_n$ common to discrete $X_0:n, Y_0:n$.
- Key idea is similar, but nonstochastic and applicable to continuous-valued $X, Y$. 

Nair (Uni. Melbourne)  Nonstochastic Information  DISC School 2015  37 / 68
Zero Error Capacity and $I_*$

Zero-error capacity $C_0$ defined operationally, as the highest block-coding rate that yields zero errors:

$$C_0 : = \sup_{n, \mu \in \mathbb{N}, \gamma: n} \frac{\log_2 \mu}{n + 1} = \lim_{n \to \infty} \sup_{\mu \in \mathbb{N}, \gamma: n} \frac{\log_2 \mu}{n + 1}. \quad (12)$$

**Theorem (after N. TAC13)**

$$C_0 = \sup_{n \geq 0, (X, Y) \in \mathcal{G}_{nf}} \frac{I_*[X_{0:n}; Y_{0:n}]}{n + 1} = \left( \lim_{n \to \infty} \sup_{(X, Y) \in \mathcal{G}_{nf}} \frac{I_*[X_{0:n}; Y_{0:n}]}{n + 1} \right). \quad (13)$$

- In [Wolf-Wullschleger itw04], $C_0$ was characterised as the largest Shannon entropy rate of the maximal rv $Z_n$ common to discrete $X_{0:n}, Y_{0:n}$.

- Key idea is similar, but nonstochastic and applicable to continuous-valued $X, Y$
Proof: \( \geq \) (Construct a Code)

- Pick any \((X, Y) \in \mathcal{G}_{nf}, n \in \mathbb{N}\). Let

\[
\mu = |\{X_{0:n}; Y_{0:n}\}_*| = |\{Y_{0:n}; X_{0:n}\}_*|,
\]

and index the overlap partition sets:

\[
\begin{align*}
\{X_{0:n}|Y_{0:n}\}_* & \equiv \{P_X(z) : z \in [1 : \mu]\}, \quad (14) \\
\{Y_{0:n}|X_{0:n}\}_* & \equiv \{P_Y(z) : z \in [1 : \mu]\}. \quad (15)
\end{align*}
\]

- Define \(uv Z\) as the unique index s.t. \(P_X(Z) \ni X_{0:n}\).
This is also the unique index s.t. \(P_Y(Z) \ni Y_{0:n}\).

- For each \(z \in [1 : \mu]\), pick an input sequence \(x(z) \in P_X(z) \subseteq \{X_{0:n}\}\)
and define the coder map

\[
\gamma(z) = x(z) \in \{X_{0:n}\}, \quad \forall z \in [1 : \mu].
\]
Proof: \( \geq \) (Construct a Code)

- Pick any \((X, Y) \in \mathcal{G}_{nf}, n \in \mathbb{N}\). Let
  \[
  \mu = |[X_{0:n}; Y_{0:n}]^*| \equiv |[Y_{0:n}; X_{0:n}]^*| ,
  \]
  and index the overlap partition sets:
  \[
  [X_{0:n} | Y_{0:n}]^* \equiv \{P_X(z) : z \in [1 : \mu]\} , \quad (14)
  \]
  \[
  [Y_{0:n} | X_{0:n}]^* \equiv \{P_Y(z) : z \in [1 : \mu]\} . \quad (15)
  \]

- Define \(uvZ\) as the unique index s.t. \(P_X(Z) \ni X_{0:n}\).
  This is also the unique index s.t. \(P_Y(Z) \ni Y_{0:n}\).
- For each \(z \in [1 : \mu]\), pick an input sequence \(x(z) \in P_X(z) \subseteq [X_{0:n}]\)
  and define the coder map
  \[
  \gamma(z) = x(z) \in [X_{0:n}] , \quad \forall z \in [1 : \mu].
  \]
Proof: \( \geq \) (cont.)

- Now, consider any message \( M \perp V \) that can take \( \mu \) distinct values \( m^1, \ldots, m^\mu \). Encode this message to give an input uv sequence

\[
X'_{0:n} = x(i) \text{ if } M = m^i.
\]

This yields an output sequence \( Y'_{0:n} \), where

\[
Y'_k = \tau(X'_k, V_k), \quad k \in [0 : n].
\]

- As \( M \) and \( X'_{0:n} \) each \( \perp V \), it follows that if \( M = m^i \) then

\[
\mathbb{P}[Y'_{0:n} | X'_{0:n} = x(i)] = \mathbb{P}[Y_{0:n} | X_{0:n} = x(i)] \subseteq P_Y(i).
\]

- Sets \( P_Y(1), \ldots, P_Y(\mu) \) are disjoint since they form a partition

\[\Rightarrow\] Message \( M \) can be recovered from \( Y'_{0:n} \) with this code.
Proof: \( \geq \) (cont.)

- Now, consider any message \( M \perp V \) that can take \( \mu \) distinct values \( m^1, \ldots, m^\mu \). Encode this message to give an input uv sequence

\[
X'_0:n = x(i) \quad \text{if} \quad M = m^i.
\]

This yields an output sequence \( Y'_0:n \), where

\[
Y'_k = \tau(X'_k, V_k), \quad k \in [0:n].
\]

- As \( M \) and \( X_0:n \) each \( \perp V \), it follows that if \( M = m^i \) then

\[
[ Y'_0:n | X'_0:n = x(i) ] = [ Y_0:n | X_0:n = x(i) ] \subseteq P_Y(i).
\]

- Sets \( P_Y(1), \ldots, P_Y(\mu) \) are disjoint since they form a partition.

\( \implies \) Message \( M \) can be recovered from \( Y'_0:n \) with this code.
Proof: (cont.)

- Now, consider any message $M \perp V$ that can take $\mu$ distinct values $m^1, \ldots, m^\mu$. Encode this message to give an input uv sequence

$$X'_0:n = x(i) \text{ if } M = m^i.$$ 

This yields an output sequence $Y'_0:n$, where

$$Y'_k = \tau(X'_k, V_k), \quad k \in [0 : n].$$

- As $M$ and $X_0:n$ each $\perp V$, it follows that if $M = m^i$ then

$$\Pr[Y'_0:n|X'_0:n = x(i)] = \Pr[Y_0:n|X_0:n = x(i)] \subseteq P_Y(i).$$

- Sets $P_Y(1), \ldots, P_Y(\mu)$ are disjoint since they form a partition

$\Rightarrow$ Message $M$ can be recovered from $Y'_0:n$ with this code.
Thus

\[ C_0 \geq \frac{\log_2 \mu}{n+1} = \frac{\log_2 \left[ I\left[ X_0:n \mid Y_0:n \right] \right]}{n+1} = \frac{l^*_\left[ X_0:n; Y_0:n \right]}{n+1}. \]

As \((X, Y) \in \mathcal{G}_{nf}\) and \(n \in \mathbb{Z}\) were arbitrary,

\[ C_0 \geq \sup_{n \geq 0, (X, Y) \in \mathcal{G}_{nf}} \frac{l^*_\left[ X_0:n; Y_0:n \right]}{n+1}. \]
Proof: ≥ (cont.)

Thus

\[ C_0 \geq \frac{\log_2 \mu}{n+1} = \frac{\log_2 [\mathbb{E}[X_0:n|Y_0:n]]^*}{n+1} = \frac{l_*[X_0:n; Y_0:n]}{n+1}. \]

As \((X, Y) \in \mathcal{G}_{nf}\) and \(n \in \mathbb{Z}\) were arbitrary,

\[ C_0 \geq \sup_{n \geq 0, (X, Y) \in \mathcal{G}_{nf}} \frac{l_*[X_0:n; Y_0:n]}{n+1}. \]
Proof: \( \leq (\text{Construct } (X, Y) \in \mathcal{G}_{nf}) \)

- Select an arbitrary zero-error code \((n, \mu, \gamma)\).
- Pick a message \(uv\) \(M \perp V\) taking distinct values \(m^1, \ldots, m^\mu\).
- Set

\[
X_0:n = \gamma(i) \text{if } M = m_i \\
X_k = X_n, \quad k > n. \\
Y_k = \tau(X_k, V_k), \quad k \in \mathbb{Z}_{\geq 0}.
\]

As \(X_0:n\) is a function of \(M \perp V\), it follows that \(X \perp V\).

Thus \((X, Y) \in \mathcal{G}_{nf}\).
Proof: \( \leq \) (cont.)

- By zero-error property, the sets \([Y_{0:n}|X_{0:n} = \gamma(i)]\), \(i = 1, \ldots, \mu\), are disjoint, therefore distinct.
- Thus each partition set in \([Y_{0:n}|X_{0:n}]^*\) contains exactly one of these sets:
  - It includes at least one set \([Y_{0:n}|X_{0:n}]\).
  - If it includes more than one such set then, by definition of the overlap partition they would have overlaps, which is impossible.
- \(\Rightarrow\) \(\mu = |[Y_{0:n}|X_{0:n}]^*|\).
Proof: $\leq$ (cont.)

Thus

$$\frac{\log_2 \mu}{n+1} = \frac{\log_2 |\{ Y: n | X: n \}^\ast |}{n+1} \leq \sup_{n \geq 0, (X, Y) \in G_{nf}} \frac{l^\ast [X: n; Y: n]}{n+1}.$$

As the zero-error code $(n, \mu, \gamma)$ was arbitrary, we can take a supremum in the LHS to get

$$C_0 \leq \sup_{n \geq 0, (X, Y) \in G_{nf}} \frac{l^\ast [X: n; Y: n]}{n+1}.$$

$\square$
Conditional Maximin Information

- Let $\mathcal{P}[X; Y|w] := \text{taxicab partition of the conditional joint range of } [X, Y|w]$, given $W = w$.
- Then define conditional nonstochastic information

$$I_*[X; Y|W] := \min_{w \in [W]} \log_2 |\mathcal{P}[X; Y|w]|.$$ 

= Log-cardinality of most refined variable common to $(X, W)$ and $(Y, W)$ but unrelated to $W$.

I.e. if two agents each observe $X, Y$ separately but also share $W$, then $I_*[X; Y|W]$ captures the most refined variable that is ‘new’ with respect to $W$ and on which they can both agree.
Conditional Maximin Information

- Let \( \mathcal{T}[X; Y|w] := \) taxicab partition of the conditional joint range \([X, Y|w]\), given \( W = w \).
- Then define *conditional nonstochastic information* 

\[
I^*_{\mathbf{w}}[X; Y|W] := \min_{w \in [W]} \log_2 |\mathcal{T}[X; Y|w]|.
\]

- = Log-cardinality of most refined variable common to \((X, W)\) and \((Y, W)\) but **unrelated to** \( W \).
- I.e. if two agents each observe \( X, Y \) separately but also share \( W \), then \( I^*_{\mathbf{w}}[X; Y|W] \) captures the most refined variable that is ‘new’ with respect to \( W \) and on which they can both agree.
A zero-error code with feedback is defined by
- a block length \( n + 1 \in \mathbb{N} \);
- a message cardinality \( \mu \geq 1 \);
- and a sequence \( y_{0:n} \) of encoder mappings s.t. for any message \( M \perp V \) taking values \( m^1, \ldots, m^\mu \),
  - \( X_k = r_k(i, Y_{0:k-1}) \) if \( M = m^i \),
  - \( \| M | y_{0:n} \| = 1, \forall y_{0:n} \in [Y_{0:n}] \).

Last condition equivalent to existence of a decoder that can reconstruct \( M \) from \( Y_{0:n} \) without error.
Zero-error feedback capacity $C_{0f}$ defined \textit{operationally}, as the highest feedback coding rate that yields zero errors:

$$
C_{0f} := \sup_{n, \mu \in \mathbb{N}, \gamma_1:n} \frac{\log_2 \mu}{n+1} = \lim_{n \to \infty} \sup_{\mu \in \mathbb{N}, \gamma_1:n} \frac{\log_2 \mu}{n+1}.
$$

(16)

Growth rate of maximum cardinality of sets of feedback coding functions that can be unambiguously determined from channel outputs.

Define \textit{directed nonstochastic information}

$$
l_\star[X_0:n \rightarrow Y_0:n] := \sum_{k=0}^{n} l_\star[X_0:k; Y_k | Y_0:k-1]
$$
Theorem (N. cdc12)

For a stationary memoryless uncertain channel,

\[ C_{0f} = \sup_{n \geq 0, (X,Y) \in \mathcal{G}} \frac{I^*_{X_0:n \rightarrow Y_0:n}}{n+1}. \]

Parallels characterisation in [Kim TIT08, Tatikonda-Mitter TIT09] for ordinary feedback capacity \( C_f \) of stochastic channels:

\[ C_f = \sup_{n \geq 0, p_{X_k|X_0:k-1,Y_0:k-1}, 0 \leq k \leq n} \frac{I_{X_0:n \rightarrow Y_0:n}}{n+1}, \]

where Marko-Massey directed information

\[ I_{X_0:n \rightarrow Y_0:n} := \sum_{k=0}^{n} I_{X_0:k \rightarrow Y_0:k-1}, \]

and conditional information \( I[X; Y|Z] := H[X|Z] - H[X|Y, Z] \).
LTI State Estimation over Noisy Channels

\[ Y_k = GX_k + W_k, \]
\[ X_{k+1} = AX_k + BU_k + V_k \]

\[ U_k \]

\[ Y_k \]

Quantiser/Coder

Channel

Decoder/Estimator

\[ S_k \]

\[ Q_k \]

\[ \hat{X}_k \]

Noise \[ V_k, W_k \]
LTI State Estimation - Disturbance-Free

**Plant**: LTI, noiseless, zero input:

\[ X_{k+1} = AX_k, \quad Y_k = GX_k, \quad X_0 \text{ a uv.} \]

**Coder**: \( Y_{0:k} \mapsto S_k \)

**Channel**: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

**Estimator**: \( Q_{0:k} \mapsto \hat{X}_{k+1} \).

**Objective**: Uniform \( \rho \)-exponential convergence from an \( l \)-ball. I.e. given \( \rho, l > 0 \), construct a coder-estimator s.t. for any uv \( X_0 \) with \( \| X_0 \| \subseteq B_l(0) \),

\[ \lim_{k \to \infty} \sup_{\omega \in \Omega} \rho^{-k} \| X_k - \hat{X}_k \| = 0. \]
LTI State Estimation - Disturbance-Free

**Plant:** LTI, noiseless, zero input:

\[ X_{k+1} = AX_k, \quad Y_k = GX_k, \quad X_0 \text{ a uv.} \]

**Coder:** \[ Y_{0:k} \mapsto S_k \]

**Channel:** Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \text{channel noise.} \)

**Estimator:** \[ Q_{0:k} \mapsto \hat{X}_{k+1} \]

**Objective:** Uniform \( \rho \)-exponential convergence from an \( l \)-ball. I.e. given \( \rho, l > 0 \), construct a coder-estimator s.t. for any uv \( X_0 \) with \( \|X_0\| \subseteq B_l(0) \),

\[
\limsup_{k \to \infty} \sup_{\omega \in \Omega} \rho^{-k} \|X_k - \hat{X}_k\| = 0.
\]
Disturbance-Free State Estimation and $C_0$

Assumptions:

**DF1:** $A$ has one or more eigenvalues with magnitude $> \rho$.

**DF2:** $(G, A_\rho)$ is observable, where $A_\rho := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq \rho$.

**DF3:** $X_0 \perp Z$

---

**Theorem (N. TAC13)**

*If uniform $\rho$-exponential convergence is achieved from some $l$-ball, then*

$$C_0 \geq \sum_{|\lambda_i| \geq \rho} \log_2 \left( \frac{|\lambda_i|}{\rho} \right).$$

(17)

*Conversely, if (26) holds strictly, then for any $l > 0$, a coder-estimator that achieves uniform $\rho$-exponential convergence from $B_l(0)$ can be constructed.*
Disturbance-Free State Estimation and $C_0$

Assumptions:

**DF1:** $A$ has one or more eigenvalues with magnitude $> \rho$.

**DF2:** $(G, A_\rho)$ is observable, where $A_\rho := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq \rho$.

**DF3:** $X_0 \perp Z$

---

**Theorem (N. TAC13)**

*If uniform $\rho$-exponential convergence is achieved from some $l$-ball, then*

$$C_0 \geq \sum_{|\lambda_i| \geq \rho} \log_2 \left( \frac{|\lambda_i|}{\rho} \right). \quad (17)$$

*Conversely, if (26) holds strictly, then for any $l > 0$, a coder-estimator that achieves uniform $\rho$-exponential convergence from $B_l(0)$ can be constructed.*
Necessity Argument - Scalar Case

- Pick arbitrarily large \( t \in \mathbb{N} \) and small \( \varepsilon \in \left( 0, 1 - \frac{\rho}{|\lambda|} \right) \).
- Divide \([-l, l]\) into
  
  \[
  \kappa := \left\lfloor \frac{(1 - \varepsilon)\lambda}{\rho} \right\rfloor^t \geq 1
  \]

  equal intervals of length \( 2l/\kappa \).
- Inside each interval construct a centred subinterval \( I(s) \) of shorter length \( l/\kappa \). Define the subinterval family
  
  \[
  \mathcal{H} := \{ I(s) : s = 1, \ldots, \kappa \}, \quad (18)
  \]

  noting that subintervals \( \in \mathcal{H} \) are separated by a gap \( \geq l/\kappa \).
- Set the initial state range \( \lfloor X_0 \rfloor = \bigcup_{\mathcal{H} \in \mathcal{H}} \mathcal{H} \subset [-l, l] \).
Let $E_k := X_k - \hat{X}_k$. By hypothesis, $\exists \phi > 0$ s.t.

\[
\phi \rho^k \geq \sup \| E_k \|
\]

\[
\geq 0.5 \text{diam}[E_k]
\]

\[
\geq 0.5 \text{diam}[E_k | q_{0:k-1}]
\]

\[
= 0.5 \text{diam} \left[ \lambda^k X_0 - \eta_k (q_{0:k-1}) | q_{0:k-1} \right]
\]

\[
= 0.5 \text{diam}[\lambda^k X_0 | q_{0:k-1}]
\]

\[
= 0.5 |\lambda|^k \text{diam}[X_0 | q_{0:k-1}]
\]
Let $E_k := X_k - \hat{X}_k$. By hypothesis, $\exists \phi > 0$ s.t.

$$\phi \rho^k \geq \sup \| E_k \|$$

$$\geq 0.5 \mathrm{diam} \| E_k \|$$

$$\geq 0.5 \mathrm{diam} \| E_k \|_{q_0:k-1}$$

$$= 0.5 \mathrm{diam} \left( \lambda^k X_0 - \eta_k (q_{0:k-1}) | q_{0:k-1} \right)$$

$$= 0.5 \mathrm{diam} \| \lambda^k X_0 | q_{0:k-1} \|$$

$$= 0.5 |\lambda|^k \mathrm{diam} \| X_0 | q_{0:k-1} \|$$
Necessity Argument - Scalar Case (cont.)

Let $E_k := X_k - \hat{X}_k$. By hypothesis, $\exists \phi > 0$ s.t.

$$\phi \rho^k \geq \sup \| E_k \|$$

$$\geq 0.5 \text{diam} [ E_k ]$$

$$\geq 0.5 \text{diam} [ E_k | q_{0:k-1} ]$$

$$= 0.5 \text{diam} [ \lambda^k X_0 - \eta_k (q_{0:k-1}) | q_{0:k-1} ]$$

$$= 0.5 \text{diam} [ \lambda^k X_0 | q_{0:k-1} ]$$

$$= 0.5 |\lambda|^k \text{diam} [ X_0 | q_{0:k-1} ]$$

(19) (20) (21) (22)
Let $E_k := X_k - \hat{X}_k$. By hypothesis, $\exists \phi > 0$ s.t.

\[
\phi \rho^k \
\geq \sup \| E_k \| \\
\geq 0.5 \text{diam} [E_k] \\
\geq 0.5 \text{diam} [E_k | q_{0:k-1}] \\
= 0.5 \text{diam} \left[ \lambda^k X_0 - \eta_k (q_{0:k-1}) | q_{0:k-1} \right] \\
= 0.5 \text{diam} [\lambda^k X_0 | q_{0:k-1}] \\
= 0.5 |\lambda|^k \text{diam} [X_0 | q_{0:k-1}] 
\]
Let $E_k := X_k - \hat{X}_k$. By hypothesis, $\exists \phi > 0$ s.t.

\[
\phi \rho^k \geq \sup \|E_k\| \\
\geq 0.5 \text{diam}[E_k] \\
\geq 0.5 \text{diam}[E_k | q_{0:k-1}] \\
= 0.5 \text{diam} \left[ \lambda^k X_0 - \eta_k (q_{0:k-1}) | q_{0:k-1} \right] \\
= 0.5 \text{diam}[\lambda^k X_0 | q_{0:k-1}] \\
= 0.5 |\lambda|^k \text{diam}[X_0 | q_{0:k-1}] 
\]
Next show that for large $t$, no two sets in $\mathcal{H}$ (18) can be $[X_0|Q_{0:t-1}]$-overlap-connected:

Suppose in contradiction that $\exists \mathcal{H} \in \mathcal{H}$ that is $[X_0|Q_{0:t-1}]$-overlap-connected with another set in $\mathcal{H}$.

$\Rightarrow \exists [X_0|q_{0:t-1}]$ containing both a point $u \in \mathcal{H}$ and a point $v$ in some $\mathcal{H}' \in \mathcal{H} \setminus \{\mathcal{H}\}$

$\Rightarrow |u - v| \leq \text{diam}[X_0|q_{0:t-1}] \overset{(22)}{\leq} \frac{2\phi \rho^t}{|\lambda|^t}. \quad (23)$
Next show that for large $t$, no two sets in $\mathcal{H}$ (18) can be $[X_0|Q_{0:t-1}]$-overlap-connected:

Suppose in contradiction that $\exists H \in \mathcal{H}$ that is $[X_0|Q_{0:t-1}]$-overlap-connected with another set in $\mathcal{H}$.

$\Rightarrow \exists [X_0|q_{0:t-1}]$ containing both a point $u \in H$ and a point $v$ in some $H' \in \mathcal{H} \setminus \{H\}$

$\Rightarrow$

$$|u - v| \leq \text{diam}[X_0|q_{0:t-1}] \overset{(22)}{\leq} \frac{2\phi \rho^t}{|\lambda|^t}. \quad (23)$$
Next show that for large $t$, no two sets in $H$ (18) can be $[X_0|Q_{0:t-1}]$-overlap-connected:

Suppose in contradiction that $\exists H \in H$ that is $[X_0|Q_{0:t-1}]$-overlap-connected with another set in $H$.

$\Rightarrow \exists [X_0|q_{0:t-1}]$ containing both a point $u \in H$ and a point $v$ in some $H' \in H \setminus \{H\}$

$\Rightarrow$

$$|u - v| \leq \text{diam}[X_0|q_{0:t-1}] \overset{(22)}{\leq} \frac{2\phi \rho^t}{|\lambda| t}. \tag{23}$$
However, any two sets $\in \mathcal{H}$ are separated by a distance of at least $l/\kappa$. So

$$|u - v| \geq \frac{l}{\kappa} = \frac{l}{((1 - \varepsilon)\lambda|/\rho)^t}$$

$$\geq \frac{l}{((1 - \varepsilon)\lambda|/\rho)^t} = \frac{l\rho^t}{((1 - \varepsilon)\lambda)^t}.$$ 

The RHS of this would exceed the RHS of (23) when $t$ is sufficiently large that $(\frac{1}{1-\varepsilon})^t > 2\phi/l$, yielding a contradiction.
Necessity Argument - Scalar Case (cont.)

- So for large enough $t$, no two sets of $\mathcal{H}$ are $\llbracket X_0\mid Q_{0:t-1} \rrbracket$-overlap-connected.

- So

$$2^t_\ast[X_0;Q_{0:t-1}] \equiv |\llbracket X_0\mid Q_{0:t-1} \rrbracket| \geq |\mathcal{H}|$$

$$= \left| \frac{(1-\varepsilon)\lambda}{\rho} \right|^t$$

$$\geq 0.5 \left| \frac{(1-\varepsilon)\lambda}{\rho} \right|^t, \quad (24)$$

since $|x| > x/2$, for every $x \geq 1$. 
So for large enough $t$, no two sets of $\mathcal{H}$ are $[X_0|Q_0:t-1]$-overlap-connected.

So

$$2^{l_*[X_0;Q_0:t-1]} \equiv \|\{X_0|Q_0:t-1\}\| \geq |\mathcal{H}|$$

$$= \left| \frac{(1-\varepsilon)\lambda}{\rho} \right|^t$$

$$\geq 0.5 \left| \frac{(1-\varepsilon)\lambda}{\rho} \right|^t,$$

(24)

since $\lfloor x \rfloor > x/2$, for every $x \geq 1$. 
Necessity Argument - Scalar Case (cont.)

- But $X_0 - S_{0:t-1} - Q_{0:t-1}$ is a Markov uncertainty chain, so

$$I^*_X[X_0; Q_{0:t-1}] \leq I^*_S[S_{0:t-1}; Q_{0:t-1}] \leq tC_0.$$

- Substitute into the LHS of (24), take logarithms and divide by $t$ to get

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right| - \frac{1}{t}.$$

- Letting $t \to \infty$ yields

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right|.$$

As $\varepsilon$ can be made arbitrarily small, we are done. $\square$
Necessity Argument - Scalar Case (cont.)

- But $X_0 - S_{0:t-1} - Q_{0:t-1}$ is a Markov uncertainty chain, so

$$I^*[X_0; Q_{0:t-1}] \leq I^*[S_{0:t-1}; Q_{0:t-1}] \leq tC_0.$$ 

- Substitute into the LHS of (24), take logarithms and divide by $t$ to get

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right| - \frac{1}{t}.$$ 

- Letting $t \rightarrow \infty$ yields

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right|.$$ 

As $\varepsilon$ can be made arbitrarily small, we are done. $\square$
Necessity Argument - Scalar Case (cont.)

- But $X_0 - S_{0:t-1} - Q_{0:t-1}$ is a Markov uncertainty chain, so

$$I^* [X_0; Q_{0:t-1}] \leq I^* [S_{0:t-1}; Q_{0:t-1}] \leq tC_0.$$  

- Substitute into the LHS of (24), take logarithms and divide by $t$ to get

$$C_0 \geq \log_2 (1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right| - \frac{1}{t}.$$  

- Letting $t \to \infty$ yields

$$C_0 \geq \log_2 (1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right|.$$  

As $\varepsilon$ can be made arbitrarily small, we are done. □
Necessity Argument - Scalar Case (cont.)

- But $X_0 - S_{0:t-1} - Q_{0:t-1}$ is a Markov uncertainty chain, so

$$I^*[X_0; Q_{0:t-1}] \leq I^*[S_{0:t-1}; Q_{0:t-1}] \leq tC_0.$$  

- Substitute into the LHS of (24), take logarithms and divide by $t$ to get

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right| - \frac{1}{t}.$$  

- Letting $t \to \infty$ yields

$$C_0 \geq \log_2(1 - \varepsilon) + \log_2 \left| \frac{\lambda}{\rho} \right|.$$  

As $\varepsilon$ can be made arbitrarily small, we are done. $\square$
State Estimation with Plant Disturbances

Plant: LTI

\[ X_{k+1} = AX_k + V_k, \quad Y_k = GX_k + W_k, \]

Coder: \( Y_{0:k} \mapsto S_k \)

Channel: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

Estimator: \( Q_{0:k} \mapsto \hat{X}_{k+1} \).

Objective: Uniformly bounded estimation errors beginning from an \( l \)-ball. I.e. given \( l > 0 \), construct a coder-estimator s.t. for any initial state \( X_0 \) with \( \|X_0\| \subseteq B_l(0) \),

\[ \sup_{k \in \mathbb{Z}_{\geq 0}, \omega \in \Omega} \|X_k - \hat{X}_k\| < \infty. \]
State Estimation with Plant Disturbances

Plant: LTI

\[ X_{k+1} = AX_k + V_k, \quad Y_k = GX_k + W_k, \]

Coder: \( Y_{0:k} \mapsto S_k \)

Channel: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

Estimator: \( Q_{0:k} \mapsto \hat{X}_{k+1} \).

Objective: Uniformly bounded estimation errors beginning from an \( l \)-ball. I.e. given \( l > 0 \), construct a coder-estimator s.t. for any initial state \( X_0 \) with \( \|X_0\| \subseteq B_l(0) \),

\[ \sup_{k \in \mathbb{Z}_{\geq 0}, \omega \in \Omega} \|X_k - \hat{X}_k\| < \infty. \]
Estimation with Disturbances and $C_0$

Assumptions:

D1: $A$ has one or more eigenvalues with magnitude $\geq 1$.

D2: $(G, A_1)$ is observable, where $A_1 := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq 1$.

D3: $\|V_k\|$ and $\|W_k\|$ are uniformly bounded over $k$.

D4: $X_0, V, W$ and $Z$ are mutually unrelated.

D5: The zero-noise sequence pair $(v, w) = (0, 0)$ is valid, i.e. $(0, 0) \in \|V, W\|$.

Theorem (N. TAC13)

If uniformly bounded estimation errors are achieved from some $l$-ball, then

$$C_0 \geq \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|.$$  \hspace{1cm} \text{(25)}

Conversely, if (25) holds strictly, then for any $l > 0$, a coder-estimator that achieves uniformly bounded estimation errors from $B_l(0)$ can be constructed.
Estimation with Disturbances and $C_0$

Assumptions:

D1: $A$ has one or more eigenvalues with magnitude $\geq 1$.

D2: $(G, A_1)$ is observable, where $A_1 := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq 1$.

D3: $[V_k]$ and $[W_k]$ are uniformly bounded over $k$.

D4: $X_0, V, W$ and $Z$ are mutually unrelated.

D5: The zero-noise sequence pair $(v, w) = (0, 0)$ is valid, i.e. $(0, 0) \in [V, W]$.

Theorem (N. TAC13)

If uniformly bounded estimation errors are achieved from some $l$-ball, then

$$C_0 \geq \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|.$$  (25)

Conversely, if (25) holds strictly, then for any $l > 0$, a coder-estimator that achieves uniformly bounded estimation errors from $B_l(0)$ can be constructed.
Control over Noisy Channels

\[ Y_k = GX_k + W_k , \]
\[ X_{k+1} = AX_k + BU_k + V_k \]

Noise \( V_k, W_k \)

Decoder/Controller

Quantiser/Coder

Channel

\[ U_k \]

\[ Q_k \]

\[ S_k \]
LTI Control - Disturbance-Free

Plant: LTI, noiseless, zero input:

\[ X_{k+1} = AX_k + BU_k, \quad Y_k = GX_k, \quad X_0 \text{ a uv.} \]

Coder: \( Y_{0:k} \mapsto S_k \)

Channel: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

Controller: \( Q_{0:k} \mapsto U_k \).

Objective: Uniform \( \rho \)-exponential stability on an \( l \)-ball. I.e. given \( \rho, l > 0 \), construct a coder-controller s.t. for any uv \( X_0 \) with \( \|X_0\| \leq B_l(0) \),

\[
\limsup_{k \to \infty} \rho^{-k} \|X_k\| = 0.
\]
LTI Control - Disturbance-Free

**Plant:** LTI, noiseless, zero input:

\[ X_{k+1} = AX_k + BU_k, \quad Y_k = GX_k, \quad X_0 \text{ a uv.} \]

**Coder:** \( Y_{0:k} \mapsto S_k \)

**Channel:** Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \text{channel noise} \).

**Controller:** \( Q_{0:k} \mapsto U_k \).

**Objective:** Uniform \( \rho \)-exponential stability on an \( l \)-ball. I.e. given \( \rho, l > 0 \), construct a coder-controller s.t. for any uv \( X_0 \) with \( \| X_0 \| \subseteq B_l(0) \),

\[
\limsup_{k \to \infty} \omega \in \Omega \rho^{-k} \| X_k \| = 0.
\]
Disturbance-Free Control and $C_{0f}$

Assumptions:

DF1: $A$ has one or more eigenvalues with magnitude $> \rho$.

DF2: $(G, A_\rho)$ is observable and $(A_\rho, B)$ is controllable, where $A_\rho := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq \rho$.

DF3: $X_0 \perp Z$

Proposition

If uniform $\rho$-exponential stability is achieved on some $l$-ball, then

$$C_{0f} \geq \sum_{|\lambda_i| \geq \rho} \log_2 \left( \frac{|\lambda_i|}{\rho} \right).$$  \hspace{1cm} (26)

Conversely, if (26) holds strictly, then for any $l > 0$, a coder-controller that achieves uniform $\rho$-exponential stability on $B_l(0)$ can be constructed.
Disturbance-Free Control and $C_{0f}$

Assumptions:

DF1: $A$ has one or more eigenvalues with magnitude $> \rho$.

DF2: $(G, A_{\rho})$ is observable and $(A_{\rho}, B)$ is controllable, where $A_{\rho} := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq \rho$.

DF3: $X_0 \perp Z$

**Proposition**

If uniform $\rho$-exponential stability is achieved on some $l$-ball, then

$$C_{0f} \geq \sum_{|\lambda_i| \geq \rho} \log_2 \left( \frac{|\lambda_i|}{\rho} \right).$$  \hspace{1cm} (26)

Conversely, if (26) holds strictly, then for any $l > 0$, a coder-controller that achieves uniform $\rho$-exponential stability on $B_l(0)$ can be constructed.
Control with Plant Disturbances

Plant: LTI

\[ X_{k+1} = AX_k + BU_k + V_k, \quad Y_k = GX_k + W_k, \]

Coder: \( Y_{0:k} \mapsto S_k \)

Channel: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

Controller: \( Q_{0:k} \mapsto U_k \).

Objective: Uniformly bounded states beginning from an \( l \)-ball. I.e. given \( l > 0 \), construct a coder-controller s.t. for any initial state \( X_0 \) with \( \|X_0\| \subseteq B_l(0) \),

\[ \sup_{k \in \mathbb{Z}_{\geq 0}, \omega \in \Omega} \|X_k\| < \infty. \]
Control with Plant Disturbances

Plant: LTI

\[ X_{k+1} = AX_k + BU_k + V_k, \quad Y_k = GX_k + W_k, \]

Coder: \( Y_{0:k} \mapsto S_k \)

Channel: Stationary and memoryless, \( Q_k = \tau(S_k, Z_k) \), where \( Z = \) channel noise.

Controller: \( Q_{0:k} \mapsto U_k \).

Objective: Uniformly bounded states beginning from an \( l \)-ball. I.e. given \( l > 0 \), construct a coder-controller s.t. for any initial state \( X_0 \) with \( \|X_0\| \subseteq B_l(0) \),

\[
\sup_{k \in \mathbb{Z}_{\geq 0}, \omega \in \Omega} \|X_k\| < \infty.
\]
Control with Disturbances and $C_{0f}$

Assumptions:

D1: $A$ has one or more eigenvalues with magnitude $\geq 1$.

D2: $(G, A_1)$ is observable and $(A_1, B)$ is controllable, where $A_1 := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq 1$.

D3: $[V_k]$ and $[W_k]$ are uniformly bounded over $k$.

D4: $X_0, V, W$ and $Z$ are mutually unrelated.

D5: The zero-noise sequence pair $(v, w) = (0, 0)$ is valid, i.e. $(0, 0) \in [V, W]$.

Theorem (N. cdc12)

If uniformly bounded estimation errors are achieved from some $l$-ball, then

$$C_{0f} \geq \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|.$$  \hspace{1cm} (27)

Conversely, if (27) holds strictly, then for any $l > 0$, a coder-controller that achieves uniformly bounded states from $B_l(0)$ can be constructed.
Control with Disturbances and $C_{0f}$

Assumptions:

D1: $A$ has one or more eigenvalues with magnitude $\geq 1$.

D2: $(G, A_1)$ is observable and $(A_1, B)$ is controllable, where $A_1 := A$ restricted to eigenspace governed by eigenvalues of magnitude $\geq 1$.

D3: $\|V_k\|$ and $\|W_k\|$ are uniformly bounded over $k$.

D4: $X_0$, $V$, $W$ and $Z$ are mutually unrelated.

D5: The zero-noise sequence pair $(v, w) = (0, 0)$ is valid, i.e. $(0, 0) \in [V, W]$.

Theorem (N. cdc12)

If uniformly bounded estimation errors are achieved from some $l$-ball, then

$$C_{0f} \geq \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|.$$  \hspace{1cm} (27)

Conversely, if (27) holds strictly, then for any $l > 0$, a coder-controller that achieves uniformly bounded states from $B_l(0)$ can be constructed.
Matveev and Savkin considered similar estimation and control problems. Mixed formulation - plant noise was a nonstochastic, bounded disturbance, while initial state and channel were stochastic and independent.

Aim was to achieve a.s. boundedness for any plant noise.

Proof of necessity there used the randomness of the initial state and channel to apply a law of large numbers. No information theory.

Here, necessity is proved using data processing on Markov uncertainty chains, and analysing $I_*$ and directed $I_*$. No statistical assumptions.
Matveev and Savkin considered similar estimation and control problems. Mixed formulation - plant noise was a nonstochastic, bounded disturbance, while initial state and channel were stochastic and independent.

Aim was to achieve a.s. boundedness for any plant noise.

Proof of necessity there used the randomness of the initial state and channel to apply a law of large numbers. No information theory.

Here, necessity is proved using data processing on Markov uncertainty chains, and analysing $I_*$ and directed $I_*$. No statistical assumptions.
Matveev and Savkin considered similar estimation and control problems. Mixed formulation - plant noise was a nonstochastic, bounded disturbance, while initial state and channel were stochastic and independent.

Aim was to achieve a.s. boundedness for any plant noise.

Proof of necessity there used the randomness of the initial state and channel to apply a law of large numbers. No information theory.

Here, necessity is proved using data processing on Markov uncertainty chains, and analysing $I_*$ and directed $I_*$. No statistical assumptions.
This talk described:

- A nonstochastic theory of uncertainty and information, without assuming a probability space

- Intrinsic characterisations of the operational zero-error capacity and zero-error feedback capacity for stationary memoryless channels.

- An information-theoretic basis for analysing worst-case networked estimation/control with bounded noise.
Outlook

Theory is still far from mature!

- Tractable algorithms to estimate $C_0$ (perhaps Monte Carlo)?
- Disturbances with bounded energy or time-averages?
- $C_{0f}$ for channels with memory?
- Zero-error feedback capacity with imperfect channel feedback?
- Multi-agent systems...?


P. Gacs and J. Korner, “Common information is far less than mutual information”, *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–62, 1973


