A Nonstochastic Theory of Information

Girish Nair

Department of Electrical and Electronic Engineering
University of Melbourne

Australian School of Information Theory
University of South Australia, Adelaide
12 November, 2014

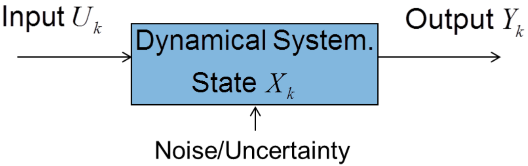
Outline

Outline

Outline

State Estimation and Control

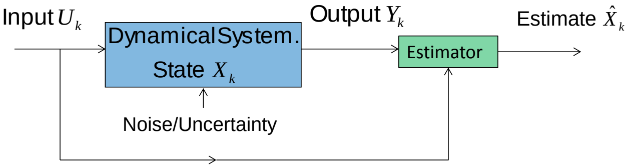
The object of interest is a given dynamical system - a *plant* - with input U_k, output Y_k, and state X_k, all possibly vector-valued.



Typically the plant is subject to noise, disturbances and/or model uncertainty.

State Estimation

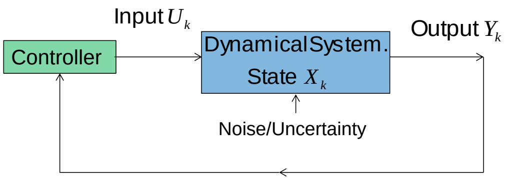
In *state estimation*, the inputs $U_0, \ldots, U_k$ and outputs $Y_0, \ldots, Y_k$ are used to estimate/predict the plant state in real-time.



Often assumed that $U_k = 0$.

Feedback Control

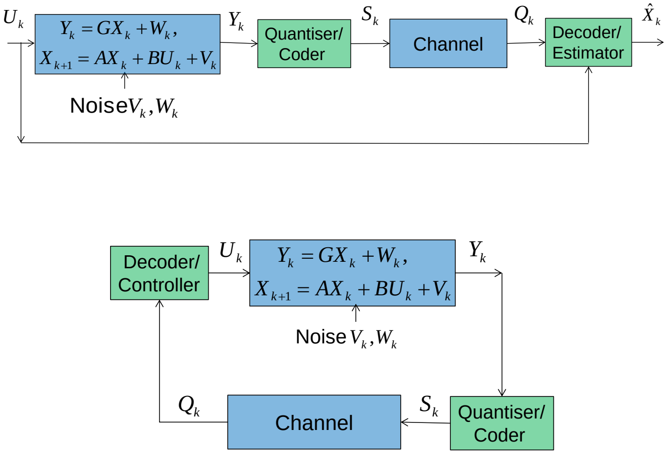
- In control, the outputs $Y_0, \ldots, Y_k$ are used to generate the input U_k, which is fed back into the plant.
- Aim is to regulate closed-loop system behaviour in some desired sense - e.g. 'small' X_k and U_k - despite noise and model uncertainty.



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 large *control area networks* in factories.
  - Quantisation errors in *sampled-data/digital* control,
  - Finite communication capacity (per-sensor) in long-range radar surveillance networks
- Limited quantiser resolution and capacity are less understood than delay in control.

Estimation/Control over Communication Channels



Additive Noise Model

- Early work considered static quantisation and errorless channels. Quantiser errors modelled as additive, uncorrelated measurement noise [e.g. Curry 1970] with variance $\propto 2^{-2R}$ ($R =$ errorless bit rate).
- Good 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 qualitatively wrong:
  1. If plant is noiseless and unstable, then states/estimation errors cannot converge to 0 .
  2. If plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.

Additive Noise Model

- Early work considered static quantisation and errorless channels. Quantiser errors modelled as additive, uncorrelated measurement noise [e.g. Curry 1970] with variance $\propto 2^{-2R}$ ($R =$ errorless bit rate).
- Good 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 qualitatively wrong:
  1. If plant is noiseless and unstable, then states/estimation errors cannot converge to 0 .
  2. If plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.

Additive Noise Model

- Early work considered static quantisation and errorless channels. Quantiser errors modelled as additive, uncorrelated measurement noise [e.g. Curry 1970] with variance $\propto 2^{-2R}$ ($R =$ errorless bit rate).
- Good 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 qualitatively wrong:
  1. If plant is noiseless and unstable, then states/estimation errors cannot converge to 0 .
  2. If plant is unstable, then mean-square-bounded states/estimation errors can always be achieved.

Errorless Channels

- In fact, 'reliable' state estimation or control is possible iff

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

where $\lambda_1, \ldots, \lambda_n =$ eigenvalues of plant matrix A. The RHS coincides with the *topological entropy (TE)* of the plant.

- Holds under various assumptions and reliability notions [Baillieu; Tatikonda-Mitter; N.-Evans]
  - Random initial state, noiseless plant; mean *r*th power convergence to 0.
  - Bounded initial state, noiseless plant; uniform convergence to 0
  - Random plant noise; mean-square boundedness.
  - Bounded plant noise; uniform boundedness
- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Errorless Channels

- In fact, 'reliable' state estimation or control is possible iff

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

  where $\lambda_1, \ldots, \lambda_n =$ eigenvalues of plant matrix A. The RHS coincides with the *topological entropy (TE)* of the plant.
- Holds under various assumptions and reliability notions [Baillieu; Tatikonda-Mitter; N.-Evans]
  - Random initial state, noiseless plant; mean *r*th power convergence to 0.
  - Bounded initial state, noiseless plant; uniform convergence to 0
  - Random plant noise; mean-square boundedness.
  - Bounded plant noise; uniform boundedness
- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Errorless Channels

- In fact, 'reliable' state estimation or control is possible iff

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

  where $\lambda_1, \ldots, \lambda_n =$ eigenvalues of plant matrix A. The RHS coincides with the *topological entropy (TE)* of the plant.
- Holds under various assumptions and reliability notions [Baillieu; Tatikonda-Mitter; N.-Evans]
  - Random initial state, noiseless plant; mean rth power convergence to 0.
  - Bounded initial state, noiseless plant; uniform convergence to 0
  - Random plant noise; mean-square boundedness.
  - Bounded plant noise; uniform boundedness
- Additive uncorrelated noise models of quantisation fail to capture the existence of such a threshold.

Noisy Channel

'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.

- FoM depends 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_{any} - MSB states over DMC [Sahai-Mitter TIT06]
  - FoM = C_{0f} for control or C_0 for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]
- Note $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$.

Noisy Channel

'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.

- FoM depends 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_{any} - MSB states over DMC [Sahai-Mitter TIT06]
  - FoM = C_{0f} for control or C_0 for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]
- Note $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$.

Noisy Channel

'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.

- FoM depends 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_{any} - MSB states over DMC [Sahai-Mitter TIT06]
  - FoM = C_{0f} for control or C_0 for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]
- Note $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$.

Noisy Channel

'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.

- FoM depends 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_{any} - MSB states over DMC [Sahai-Mitter TIT06]
  - FoM = C_{0f} for control or C_0 for state estimation, with a.s. bounded states/est. errors [Matveev-Savkin IJC07]
- Note $C \geq C_{\mathrm{any}} \geq C_{0f} \geq C_0$.

Missing Information

- If the goal is MSB or a.s. convergence $\to 0$ of states/estimation errors, then differential entropy, entropy power, mutual information, and the data processing inequality are crucial for proving 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? (cont.)

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.
  $\Rightarrow$ 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.)

- 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.

Probability in Practice

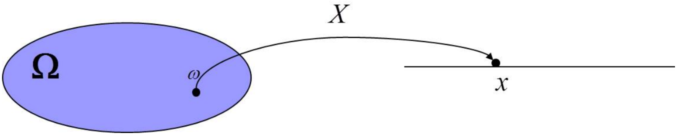
Proposition (attrib. L. 'Yogi' Berra, former US baseball player)

'If there's a fifty-fifty chance that something can go wrong, nine out of ten times, it will.'

(Photo from Wikipedia)

Uncertain Variable Formalism

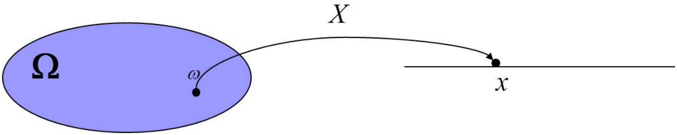
- Define an *uncertain variable (uv)* X to be a mapping from an underlying sample space Ω to a space $\mathbb{X}$.
- Each $\omega \in \Omega$ may represent a specific combination of noise/input signals into a system, and X may represent a state/output variable.
- For a given ω, $x = X(\omega)$ is the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



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

Uncertain Variable Formalism

- Define an *uncertain variable (uv)* X to be a mapping from an underlying sample space Ω 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 ω, $x = X(\omega)$ is the *realisation* of X.



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

UV Formalism- Ranges and Conditioning

- *Marginal* range $[\![X]\!] := \{X(\omega) : \omega \in \Omega\} \subseteq \mathbb{X}$.
- *Joint* range $[\![X, Y]\!] := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq \mathbb{X} \times \mathbb{Y}$.
- *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, similar to probability.

Independence Without Probability

Definition

The uv's X, Y are called (mutually) unrelated *if*

$$[\![X, Y]\!] = [\![X]\!] \times [\![Y]\!], \tag{1}$$

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

- Equivalent characterisation:

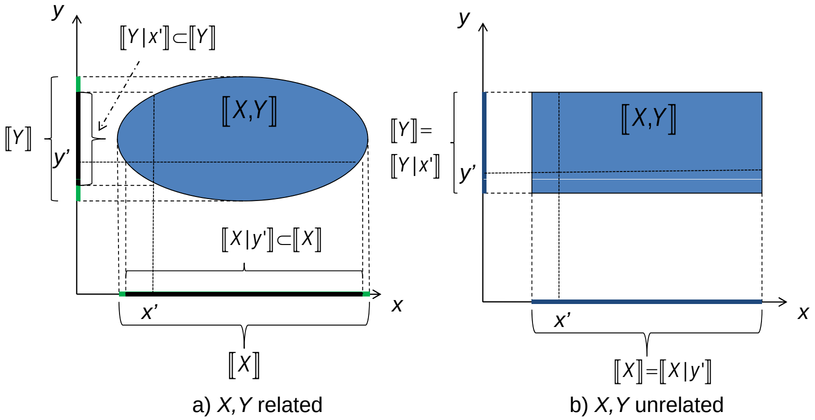
Proposition

The uv's X, Y unrelated if

$$[\![X|y]\!] = [\![X]\!], \quad \forall y \in [\![Y]\!]. \tag{2}$$

- Unrelatedness is equivalent to X and Y inducing *qualitatively independent* [Rényi'70] partitions of Ω when Ω is finite.

Examples of Relatedness and Unrelatedness



Markovness without Probability

Definition

X, Y, Z said to form a Markov uncertainty chain $X - Y - Z$ *if*

$$[\![X|y,z]\!] = [\![X|y]\!], \ \forall (y,z) \in [\![Y,Z]\!]. \tag{3}$$

- Equivalent to

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

  i.e. *X, Z conditionally unrelated given Y*, $X \perp Z|Y$.

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

Information without Probability

Definition

Two points $(x, y), (x', y') \in [\![X, Y]\!]$ *are called* taxicab connected $(x, y) \leftrightsquigarrow (x'y')$ *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 $\leftrightsquigarrow$ is an equivalence relation on $[\![X, Y]\!]$.
- Call its equivalence classes a *taxicab partition* $\mathscr{T}[X; Y]$ of $[\![X, Y]\!]$.
- Define a nonstochastic information index

$$I_*[X; Y] := \log_2 |\mathscr{T}[X; Y]| \in [0, \infty]. \quad (4)$$

Information without Probability

Definition

Two points $(x, y), (x', y') \in [\![X, Y]\!]$ *are called* taxicab connected $(x, y) \leftrightsquigarrow (x'y')$ *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 $\leftrightsquigarrow$ is an equivalence relation on $[\![X, Y]\!]$.
- Call its equivalence classes a *taxicab partition* $\mathscr{T}[X; Y]$ of $[\![X, Y]\!]$.
- Define a nonstochastic information index

$$I_*[X; Y] := \log_2 |\mathscr{T}[X; Y]| \in [0, \infty]. \tag{4}$$

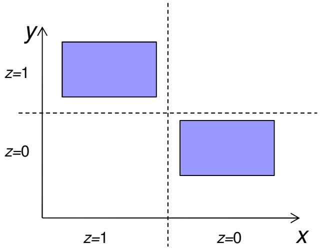
Connection to Common Random Variables

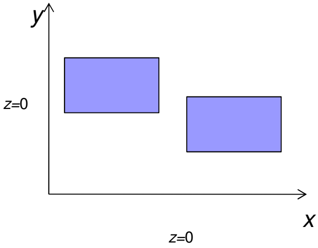
- $\mathscr{T}[X;Y]$ also called *ergodic decomposition* [Gács-Körner PCIT72].
- For discrete X, Y, equivalent to *connected components* of [Wolf-Wullschleger itw04], which were shown there to be the maximal *common rv* Z_*, i.e.
  - $Z_* = f_*(X) = g_*(Y)$ under suitable mappings f_*, g_* (since points in distinct sets in $\mathscr{T}[X;Y]$ 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 $\mathscr{T}[X;Y]$ are taxicab-connected)
- Not hard to see that Z_* also has the largest no. distinct values of any common rv $Z \equiv f(X) \equiv g(Y)$.
- $I_*[X;Y] =$ *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

- $\mathscr{T}[X;Y]$ also called *ergodic decomposition* [Gács-Körner PCIT72].
- For discrete X, Y, equivalent to *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 $\mathscr{T}[X;Y]$ 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 $\mathscr{T}[X;Y]$ are taxicab-connected)
- Not hard to see that Z_* also has the largest no. distinct values of any common rv $Z \equiv f(X) \equiv g(Y)$.
- $I_*[X;Y] =$ *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.# distinct values
that can always be agreed on
from separate observations of X & Y.

$|\mathcal{T}| = 1 =$ max.# 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* $\mathbb{B}_1, \ldots, \mathbb{B}_n \in [\![X|Y]\!]$ *s.t.*

- $x \in \mathbb{B}_1$ *and* $x' \in \mathbb{B}_n$
- $\mathbb{B}_i \cap \mathbb{B}_{i+1} \neq \emptyset$, $\forall i \in [1 : n-1]$.

- 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* $\mathbb{B}_1, \ldots, \mathbb{B}_n \in [\![X|Y]\!]$ *s.t.*

- $x \in \mathbb{B}_1$ *and* $x' \in \mathbb{B}_n$
- $\mathbb{B}_i \cap \mathbb{B}_{i+1} \neq \emptyset$, $\forall i \in [1 : n-1]$.

- 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,

$$I_*[X;Y] = \log_2 |[\![X|Y]\!]_*| . \quad (5)$$

Proof Sketch:

- For any two points $(x,y),(x',y') \in [\![X,Y]\!]$, $(x,y) \leftrightsquigarrow (x',y')$ iff x' and x' are $[\![X|Y]\!]$-overlap-connected.
- This allows us to set up a bijection between the partitions $\mathscr{I}[X;Y]$ and $[\![X|Y]\!]_*$.
- $\Rightarrow \mathscr{I}[X;Y]$ and $[\![X|Y]\!]_*$ must have the same cardinality.

Equivalent View via Overlap Partitions (cont.)

Proposition

For any uv's X, Y,

$$I_*[X;Y] = \log_2 |[\![X|Y]\!]_*| . \quad (5)$$

Proof Sketch:

- For any two points $(x,y),(x',y') \in [\![X,Y]\!]$, $(x,y) \leftrightsquigarrow (x',y')$ iff x' and x' are $[\![X|Y]\!]$-overlap-connected.
- This allows us to set up a bijection between the partitions $\mathscr{T}[X;Y]$ and $[\![X|Y]\!]_*$.
- $\Rightarrow \mathscr{T}[X;Y]$ and $[\![X|Y]\!]_*$ must have the same cardinality.

Equivalent View via Overlap Partitions (cont.)

Proposition

For any uv's X, Y,

$$I_*[X; Y] = \log_2 |[\![X|Y]\!]_*|. \quad (5)$$

Proof Sketch:

- For any two points $(x, y), (x', y') \in [\![X, Y]\!]$, $(x, y) \leftrightsquigarrow (x', y')$ iff x' and x' are $[\![X|Y]\!]$-overlap-connected.
- This allows us to set up a bijection between the partitions $\mathscr{T}[X; Y]$ and $[\![X|Y]\!]_*$.
- $\Rightarrow \mathscr{T}[X; Y]$ and $[\![X|Y]\!]_*$ 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

$$(x,y) \leftrightsquigarrow (x',y') \in [\![X,Y]\!] \iff (y,x) \leftrightsquigarrow (y',x') \in [\![Y,X]\!]. \qquad (6)$$

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_* (cont.)

Proposition (Monotonicity)

For any uv's X, Y *and* Z*,*

$$I_*[X;Y] \leq I_*[X;Y,Z]. \tag{7}$$

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

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

Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathbb{B}' \in [\![X|Y,Z]\!]_*$ exists for each $\mathbb{B} \in [\![X|Y,Z]\!]_*$, 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 $[\![X|Y,Z]\!]$.
  - As $[\![X|y,z]\!] \subseteq [\![X|y]\!]$, x, x' are also connected by a sequence of successively overlapping sets in $[\![X|Y]\!]$.
  - But $\mathbb{B}' =$ all pts. that are $[\![X|Y]\!]$-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 $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$.
- Furthermore it is onto, since every set $\mathbb{B}' \in [\![X|Y]\!]_*$ intersects some $\mathbb{B}$ in $[\![X|Y,Z]\!]_*$, which covers $[\![X]\!]$.
- So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$. $\square$

Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathbb{B}' \in [\![X|Y,Z]\!]_*$ exists for each $\mathbb{B} \in [\![X|Y,Z]\!]_*$, 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 $[\![X|Y,Z]\!]$.
  - As $[\![X|y,z]\!] \subseteq [\![X|y]\!]$, x, x' are also connected by a sequence of successively overlapping sets in $[\![X|Y]\!]$.
  - But $\mathbb{B}' =$ all pts. that are $[\![X|Y]\!]$-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 $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$.
- Furthermore it is onto, since every set $\mathbb{B}' \in [\![X|Y]\!]_*$ intersects some $\mathbb{B}$ in $[\![X|Y,Z]\!]_*$, which covers $[\![X]\!]$.
- So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$. $\square$

Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathbb{B}' \in [\![X|Y,Z]\!]_*$ exists for each $\mathbb{B} \in [\![X|Y,Z]\!]_*$, 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 $[\![X|Y,Z]\!]$.
  - As $[\![X|y,z]\!] \subseteq [\![X|y]\!]$, x, x' are also connected by a sequence of successively overlapping sets in $[\![X|Y]\!]$.
  - But $\mathbb{B}' =$ all pts. that are $[\![X|Y]\!]$-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 $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$.
- Furthermore it is onto, since every set $\mathbb{B}' \in [\![X|Y]\!]_*$ intersects some $\mathbb{B}$ in $[\![X|Y,Z]\!]_*$, which covers $[\![X]\!]$.
- So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$. $\square$

Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathbb{B}' \in [\![X|Y,Z]\!]_*$ exists for each $\mathbb{B} \in [\![X|Y,Z]\!]_*$, 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 $[\![X|Y,Z]\!]$.
  - As $[\![X|y,z]\!] \subseteq [\![X|y]\!]$, x, x' are also connected by a sequence of successively overlapping sets in $[\![X|Y]\!]$.
  - But $\mathbb{B}' =$ all pts. that are $[\![X|Y]\!]$-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 $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$.
- Furthermore it is onto, since every set $\mathbb{B}' \in [\![X|Y]\!]_*$ intersects some $\mathbb{B}$ in $[\![X|Y,Z]\!]_*$, which covers $[\![X]\!]$.
- So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$. □

Proof of Monotonic Property (cont.)

- Furthermore, exactly one such intersecting $\mathbb{B}' \in [\![X|Y,Z]\!]_*$ exists for each $\mathbb{B} \in [\![X|Y,Z]\!]_*$, 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 $[\![X|Y,Z]\!]$.
  - As $[\![X|y,z]\!] \subseteq [\![X|y]\!]$, x, x' are also connected by a sequence of successively overlapping sets in $[\![X|Y]\!]$.
  - But $\mathbb{B}' =$ all pts. that are $[\![X|Y]\!]$-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 $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$.
- Furthermore it is onto, since every set $\mathbb{B}' \in [\![X|Y]\!]_*$ intersects some $\mathbb{B}$ in $[\![X|Y,Z]\!]_*$, which covers $[\![X]\!]$.
- So $\mathbb{B} \mapsto \mathbb{B}'$ is the required surjection from $[\![X|Y,Z]\!]_* \to [\![X|Y]\!]_*$. $\square$

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_*,

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

- 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 $X - Y - Z$ (3),*

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

Proof:

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

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

- 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 $X - Y - Z$ (3),*

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

Proof:

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

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

- 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 $X - Y - Z$ (3),*

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

Proof:

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

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

- 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. □

Stationary Memoryless Uncertain Channels - Take 1

- An *uncertain signal* X is a mapping from Ω to the space $\mathbb{X}^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $\mathbb{X}$.
- A stationary memoryless *uncertain* channel may be defined in terms of
  - input and output spaces $\mathbb{X}, \mathbb{Y}$;
  - a set-valued transition function $\mathbf{T} : \mathbb{X} \to 2^{\mathbb{Y}}$;
  - and the family of all uncertain input-output signal pairs (X, Y) s.t.

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

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

$$[\![X_k | x_{0:k-1}, y_{0:k-1}]\!] = [\![X_k | x_{0:k-1}]\!], \; k \in \mathbb{Z}_{\geq 0}, \qquad (10)$$

on (X, Y).

Stationary Memoryless Uncertain Channels - Take 1

- An *uncertain signal* X is a mapping from Ω to the space $\mathbb{X}^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $\mathbb{X}$.
- A stationary memoryless *uncertain* channel may be defined in terms of
  - input and output spaces $\mathbb{X}, \mathbb{Y}$;
  - a set-valued transition function $\mathbf{T} : \mathbb{X} \to 2^{\mathbb{Y}}$;
  - and the family of all uncertain input-output signal pairs (X, Y) s.t.

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

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

$$[\![X_k | x_{0:k-1}, y_{0:k-1}]\!] = [\![X_k | x_{0:k-1}]\!], \; k \in \mathbb{Z}_{\geq 0}, \tag{10}$$

on (X, Y).

Stationary Memoryless Uncertain Channels - Take 1

- An *uncertain signal* X is a mapping from Ω to the space $\mathbb{X}^\infty$ of discrete-time sequences $x = (x_i)_{i=0}^\infty$ in $\mathbb{X}$.
- A stationary memoryless *uncertain* channel may be defined in terms of
  - input and output spaces $\mathbb{X}, \mathbb{Y}$;
  - a set-valued transition function $\mathbf{T} : \mathbb{X} \to 2^{\mathbb{Y}}$;
  - and the family of all uncertain input-output signal pairs (X, Y) s.t.

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

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

$$[\![X_k | x_{0:k-1}, y_{0:k-1}]\!] = [\![X_k | x_{0:k-1}]\!], \; k \in \mathbb{Z}_{\geq 0}, \tag{10}$$

on (X, Y).

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.

Channel as Noisy Function

Definition

A stationary memoryless uncertain channel (SMUC) *consists of*

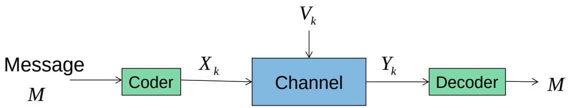
- *an* unrelated, identically spread (uis) *noise signal* $V = (V_k)_{k=0}^{\infty}$ *taking values over a space* $\mathbb{V}$*, i.e.*

$$[\![V_k|v_{0:k-1}]\!] = [\![V_k]\!] = \mathbb{V}, \;\; \forall v_{0:k-1} \in \mathbb{V}^k, k \in \mathbb{Z}_{\geq 0}; \qquad (11)$$

- *input and output spaces* $\mathbb{X}, \mathbb{Y}$*, and a* transition function $\tau : \mathbb{X} \times \mathbb{V} \to \mathbb{Y}$*;*
- *and the family* $\mathscr{G}$ *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* $X_{0:k} \perp V_k$

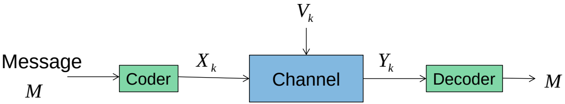
If channel used w/o feedback, then tighten last condition so that $X \perp V$. Yields smaller family $\mathscr{G}_{nf} \subset \mathscr{G}$.

Zero Error Coding in UV Framework (No Feedback)



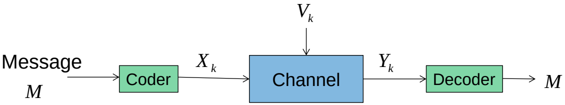
- Let $\mathscr{M}$:= set of all uv's $\perp$ V.
- A zero-error code w/o feedback is defined by
  - a block length $n+1 \in \mathbb{N}$;
  - a message cardinality $\mu \geq 1$;
  - and an encoder mapping $\gamma : [1:\mu] \to \mathbb{X}^{n+1}$, s.t. for any $M \in \mathscr{M}$ taking μ distinct values $m^1, \ldots, m^\mu$,
    - $X_{0:n} = \gamma(i)$ 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 Coding in UV Framework (No Feedback)



- Let $\mathscr{M} :=$ set of all uv's $\perp V$.
- 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 $\gamma : [1:\mu] \to \mathbb{X}^{n+1}$, s.t. for any $M \in \mathscr{M}$ taking μ 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 Coding in UV Framework (No Feedback)



- Let $\mathscr{M}$:= set of all uv's $\perp V$.
- 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 $\gamma : [1 : \mu] \to \mathbb{X}^{n+1}$, s.t. for any $M \in \mathscr{M}$ taking μ 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)

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

- 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:

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

Theorem (after N. TAC13)

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

- 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:

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

Theorem (after N. TAC13)

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

- 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 $(X, Y) \in \mathscr{G}_{nf}, n \in \mathbb{N}$. 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 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 $(X, Y) \in \mathscr{G}_{nf}, n \in \mathbb{N}$. 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 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

$$\gamma(z) = x(z) \in [\![X_{0:n}]\!], \ \ \forall z \in [1:\mu].$$

Proof: $\geq$ (cont.)

- Now, consider any message $M \in \mathscr{M}$ that can take μ 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

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

- Sets $P_Y(1), \ldots P_Y(\mu)$ are disjoint since they form a partition
- $\Rightarrow$ Message M can be recovered from $Y'_{0:n}$ with this code.

Proof: $\geq$ (cont.)

- Now, consider any message $M \in \mathscr{M}$ that can take μ 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

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

- Sets $P_Y(1), \ldots P_Y(\mu)$ are disjoint since they form a partition
- $\Rightarrow$ Message M can be recovered from $Y'_{0:n}$ with this code.

Proof: $\geq$ (cont.)

- Now, consider any message $M \in \mathscr{M}$ that can take μ 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

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

- Sets $P_Y(1), \ldots P_Y(\mu)$ are disjoint since they form a partition
- $\Rightarrow$ Message M can be recovered from $Y'_{0:n}$ with this code.

Proof: $\geq$ (cont.)

- Thus

$$C_0 \geq \frac{\log_2 \mu}{n+1} = \frac{\log_2 |[\![X_{0:n}|Y_{0:n}]\!]_*|}{n+1} = \frac{I_*[X_{0:n};Y_{0:n}]}{n+1}.$$

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

$$C_0 \geq \sup_{n \geq 0, (X,Y) \in \mathscr{G}_{nf}} \frac{I_*[X_{0:n};Y_{0:n}]}{n+1}.$$

Proof: $\geq$ (cont.)

- Thus

$$C_0 \geq \frac{\log_2 \mu}{n+1} = \frac{\log_2 |[\![X_{0:n}|Y_{0:n}]\!]_*|}{n+1} = \frac{I_*[X_{0:n}; Y_{0:n}]}{n+1}.$$

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

$$C_0 \geq \sup_{n \geq 0, (X,Y) \in \mathscr{G}_{nf}} \frac{I_*[X_{0:n}; Y_{0:n}]}{n+1}.$$

Proof: $\leq$ (Construct $(X, Y) \in \mathscr{G}_{nf}$)

- Select an arbitrary zero-error code (n, μ, γ).
- Pick a message uv $M \in \mathscr{M}$ taking distinct values $m^1, \ldots, m^\mu$.
- Set

$$\begin{aligned} X_{0:n} &= \gamma(i) \text{if } M = m_i \\ X_k &= X_n, \quad k > n. \\ Y_k &= \tau(X_k, V_k), \quad k \in \mathbb{Z}_{\geq 0}. \end{aligned}$$

- As $X_{0:n}$ is a function of $M \perp V$, it follows that $X \perp V$
  Thus $(X, Y) \in \mathscr{G}_{nf}$.

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

$$\frac{\log_2 \mu}{n+1} = \frac{\log_2 |[\![Y_{0:n} | X_{0:n}]\!]_*|}{n+1} \leq \sup_{n\geq 0,(X,Y)\in \mathscr{G}_{nf}} \frac{I_*[X_{0:n}; Y_{0:n}]}{n+1}.$$

- As the zero-error code (n, μ, γ) was arbitrary, we can take a supremum in the LHS to get

$$C_0 \leq \sup_{n\geq 0,(X,Y)\in \mathscr{G}_{nf}} \frac{I_*[X_{0:n}; Y_{0:n}]}{n+1}.$$

□

Conditional Maximin Information

An information-theoretic characterisation of C_{0f}, in terms of *directed* nonstochastic information:

- First, let $\mathscr{T}[X;Y|w] :=$ taxicab partition of the conditional joint range $[\![X,Y|w]\!]$, given $W=w$.
- Then define *conditional nonstochastic information*

$$I_*[X;Y|W] := \min_{w\in[\![W]\!]} \log_2 |\mathscr{T}[X;Y|w]|\,.$$

- $=$ 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 $I_*[X;Y|W]$ captures the most refined variable that is 'new' with respect to W and on which they can both agree.

Conditional Maximin Information

An information-theoretic characterisation of C_{0f}, in terms of *directed* nonstochastic information:

- First, let $\mathscr{T}[X;Y|w] :=$ taxicab partition of the conditional joint range $[\![X,Y|w]\!]$, given $W = w$.
- Then define *conditional nonstochastic information*

$$I_*[X;Y|W] := \min_{w \in [\![W]\!]} \log_2 |\mathscr{T}[X;Y|w]| \,.$$

- = 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 $I_*[X;Y|W]$ captures the most refined variable that is 'new' with respect to W and on which they can both agree.

C_{0f} in terms of I_*

- Zero-error feedback capacity C_{0f} is defined operationally (in terms of the largest log-cardinality of sets of feedback coding functions that can be unambiguously determined from channel outputs).
- Define directed nonstochastic information

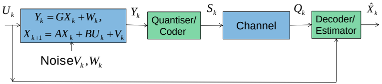
$$I_*[X_{0:n} \to Y_{0:n}] := \sum_{k=0}^{n} I_*[X_{0:k}; Y_k | Y_{0:k-1}]$$

- [N. cdc12]: For a stationary memoryless uncertain channel,

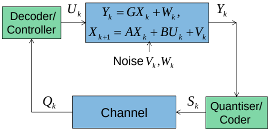
$$C_{0f} = \sup_{n \geq 0, (X,Y) \in \mathscr{G}} \frac{I_*[X_{0:n} \to Y_{0:n}]}{n+1}.$$

Parallels characterisation in [Kim TIT08, Tatikonda-Mitter TIT09] for C_f of stochastic channels (with memory) in terms of Marko-Massey directed information.

Networked State Estimation/Control Revisited



[N. TAC13]: It is possible to achieve uniformly bounded estimation errors iff $C_0 > H_A := \sum_{|\lambda_i| \geq 1} \log_2 |\lambda_i|$.



[N. cdc12]: It is possible to achieve uniformly bounded states iff $C_{0f} > H_A$.

Summary

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
  - New bounds or algorithms for C_0?
  - C_{0f} for channels with memory?
  - Zero-error capacity with partial/imperfect feedback?
  - Multiple users?

References

- G.N. Nair, "A nonstochastic information theory for communication and state estimation", *IEEE Trans. Automatic Control*, USA, vol. 58, no. 6, pp. 1497–510, 2013.
- G.N. Nair, "A nonstochastic information theory for feedback", *Proc. 51st IEEE Conf. Decision and Control*, Maui, USA, pp. 1343–8, 2012
- J. Baillieul, "Feedback designs in information-based control", *Stochastic Theory and Control. Proceedings of a Workshop held in Lawrence, Kansas*, pp. 35–57, Springer, 2002
- S. Tatikonda and S. Mitter, "Control under communication constraints", *IEEE Trans. Automatic Control*, USA, vol. 49, no. 7, pp. 1056–68, July 2004.
- G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates", *SIAM Jour. Control and Optimization*, vol. 43, no. 2, pp. 413–36, July 2004.
- A. S. Matveev and A. V. Savkin, "An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels", *SIAM Jour. Control and Optimization*, vol. 46, no. 4, pp. 1323–67, 2007.
- J. H. Braslavsky, R. H. Middleton and J. S. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels", *IEEE Trans. Automatic Control*, USA, vol. 52, no. 8, pp. 1391–403, 2007
- A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link part 1: scalar systems", *IEEE Trans. Info. Theory*, pp. 3369–95, vol. 52, no. 8, 2006.
- A.S. Matveev and A.V. Savkin, "Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels", *Int. Jour. Control*, vol. 80, pp. 241–55, 2007.

References (cont.)

- S. Wolf and J. Wullschleger, “Zero-error information and applications in cryptography”, in *Proc. Info. Theory Workshop*, San Antonio, USA, 2004, pp. 1–6.
- 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
- C.E. Shannon, “The lattice theory of information”, *Trans. IRE Professional Group on Info. Theory*, vol. 1, iss. 1, Feb. 1953., pp. 105–8.
- J. Massey, “Causality, feedback and directed information”, in *Proc. Int. Symp. Info. Theory App.*, Nov. 1990, pp. 1–6
- Y.H. Kim, “A coding theorem for a class of stationary channels with feedback”, *IEEE Trans. Info. Theory*, 1488–99, 2008.
- S. Tatikonda and S. Mitter, “The capacity of channels with feedback”, *IEEE Trans. Info. Theory*, pp. 323–49, 2009.
- C.E. Shannon, “The zero-error capacity of a noisy channel”, *Proc. IRE Trans. Info. Theory*, vol. 2, pp. 8–19, 1956.
- G. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*, Wiley, 2006, ch. 2.
- H. Shingin and Y. Ohta, “Disturbance rejection with information constraints: performance limitations of a scalar system for bounded and Gaussian disturbances”, *Automatica*, vol. 48, no. 6, pp. 1111–6, 2012.
- W. S. Wong and R. W. Brockett, “Systems with finite communication bandwidth constraints I”, *IEEE Trans. Automatic Control*, USA, vol. 42, pp. 1294–9, 1997.
- W. S. Wong and R. W. Brockett, “Systems with finite communication bandwidth constraints II: stabilization with limited information feedback”, *IEEE Trans. Automatic Control*, USA, vol. 44, pp. 1049–53, 1999.