Nonstochastic Information for Worst-Case Networked Estimation and Control

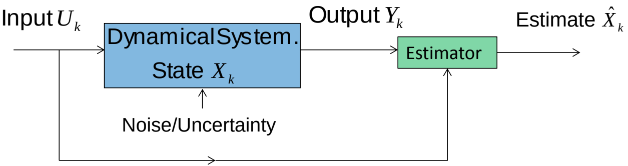
Girish Nair

Department of Electrical and Electronic Engineering
University of Melbourne

IEEE Information Theory Workshop
5 November, 2014
Hobart

State Estimation...

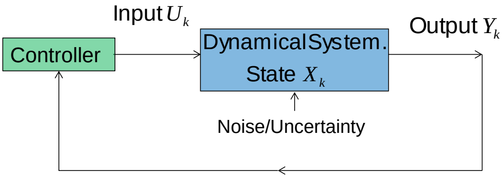
- Object of interest is a given dynamical system - a *plant* - with input U_k, output Y_k, and state X_k, all possibly vector-valued.
- Typically the plant is subject to noise, disturbances and/or model uncertainty.
- In *state estimation*, the inputs $U_0, \ldots, U_k$ and outputs $Y_0, \ldots, Y_k$ are used to estimate/predict the plant state in real-time.



Often assumed that $U_k = 0$.

...and Feedback Control

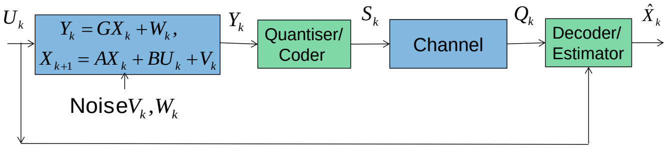
- In control, the outputs $Y_0, \ldots, Y_k$ are used to generate the input U_k, which is fed back into the plant.
  Aim is to regulate closed-loop system behaviour in some desired sense.

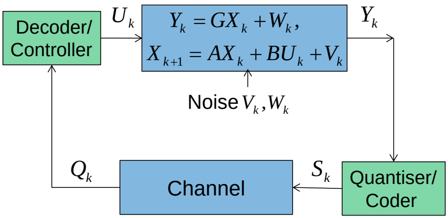


Networked State Estimation/Control

- 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
- Focus here on limited quantiser resolution and capacity, which are less understood than delay in control.

Estimation/Control over Communication Channels





Main Results in Area

'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.

- For errorless digital channels, FoM = data rate R [Baillieul'02, Tatikonda-Mitter TAC04, N.-Evans SIAM04]
- But if channel is noisy, then 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 $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$.

MELBOURNE SCHOOL OF ENGINEERING

Main Results in Area

'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.

- For errorless digital channels, FoM = data rate R [Baillieul'02, Tatikonda-Mitter TAC04, N.-Evans SIAM04]
- But if channel is noisy, then 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 $C \geq C_{\text{any}} \geq C_{0f} \geq C_0$.

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?

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.
- 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 physical laws.

Why Nonstochastic? (continued)

- 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.
- 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.

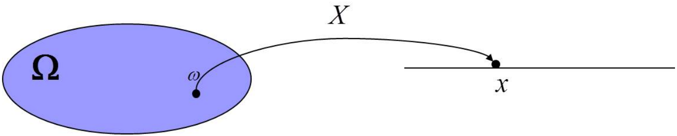
Probability in Practice

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

– Lawrence 'Yogi' Berra, former US baseball player (attributed).

Uncertain Variable Formalism

- Define an *uncertain variable (uv)* X to be a mapping from a sample space Ω to a (possibly continuous) 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

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

- X, Y called *unrelated* if

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

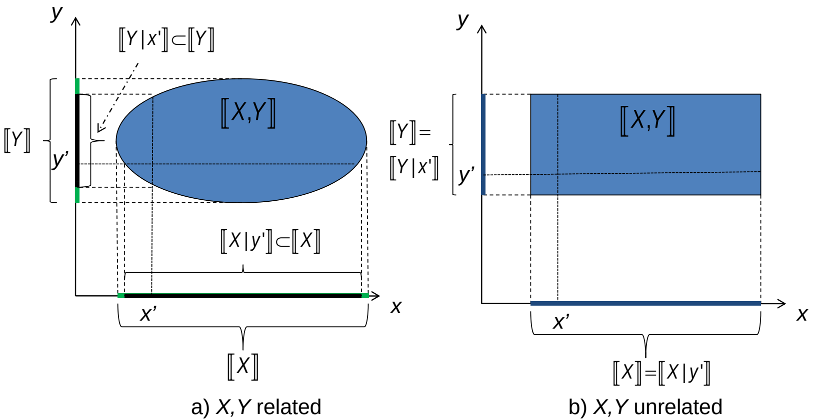
  or equivalently

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

  Else called *related*.

- 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

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

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

Equivalently,

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

i.e. X, Z are *conditionally unrelated given* Y.

Information without Probability

- Call two points $(x, y), (x', y') \in [\![X, Y]\!]$ *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.
- As $\leftrightsquigarrow$ is an equivalence relation, it induces a *taxicab partition* $\mathscr{T}[X; Y]$ of $[\![X, Y]\!]$.
- Define a nonstochastic information index

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

Information without Probability

- Call two points $(x,y),(x',y') \in [\![X,Y]\!]$ *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.
- As $\leftrightsquigarrow$ is an equivalence relation, it induces a *taxicab partition* $\mathscr{T}[X;Y]$ of $[\![X,Y]\!]$.
- Define a nonstochastic information index

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

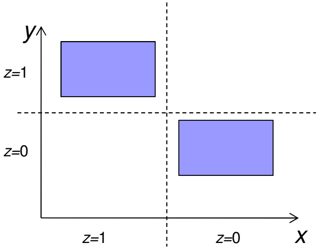
Common Random Variables

- $\mathscr{T}[X;Y]$ 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].

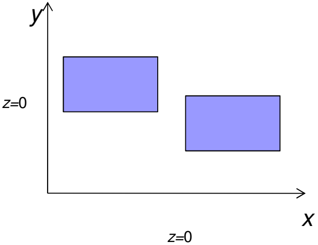
Common Random Variables

- $\mathscr{T}[X;Y]$ 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.

Similarities to Mutual Information I

- Nonnegativity $I_*[X;Y] \geq 0$.
- Symmetry: $I_*[X;Y] = I_*[Y;X]$
- Monotonicity: $I_*[X;Y] \leq I_*[X;Y,Z]$
- Data processing: For Markov **uncertainty** chains $X - Y - Z$,

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

Stationary Memoryless Uncertain Channels

- 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* consists 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 $\mathscr{G}$ of all uncertain input-output signal pairs (X, Y) s.t.

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

C.f. [Massey isit90].

Zero Error Capacity in terms of I_*

- Zero-error capacity C_0 defined *operationally*, as the highest block-code rate that yields exactly zero (probability of) errors.
- [N. TAC13]:

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

- In [Wolf-Wullschleger itw04], C_0 was characterised as the largest *Shannon* entropy rate of the maximal rv Z_n common to discrete $X_{0:n}, Y_{0:n}$.
- Similar proof here, but nonstochastic and applicable to continuous-valued X, Y.

Zero Error Capacity in terms of I_*

- Zero-error capacity C_0 defined *operationally*, as the highest block-code rate that yields exactly zero (probability of) errors.
- [N. TAC13]:

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

- In [Wolf-Wullschleger itw04], C_0 was characterised as the largest *Shannon* entropy rate of the maximal rv Z_n common to discrete $X_{0:n}, Y_{0:n}$.
- Similar proof here, but nonstochastic and applicable to continuous-valued X, Y.

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

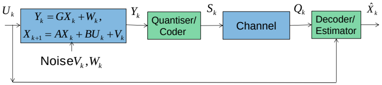
$$I_*[X_{0:n} \to Y_{0:n}] := \sum_{k=0}^{n} I_*[X_{0:k}; Y_k | Y_{0:k-1}]$$

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

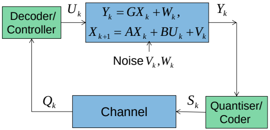
$$C_{0f} = \sup_{n \geq 0, (X,Y) \in \mathscr{G}} \frac{I_*[X_{0:n} \to Y_{0:n}]}{n+1}.$$

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?