# Fundamental limits of feedback: an information theoretic viewpoint

## Basic results and definitions of information theory

*Nuno C. Martins*

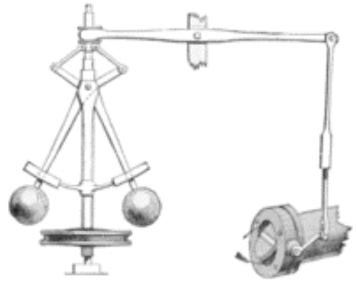
*nmartins@isr.umd.edu*

*Department of Electrical and Computer Engineering*
*Institute for Systems Research*
*University of Maryland, College Park*

**Connections II, CALTECH** August 14th, 2006

# Tomorrow’s talk will focus on

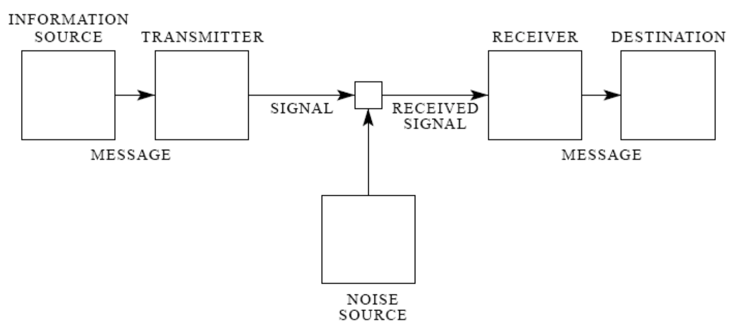
H. W. Bode



*Linear Time-Invariant Feedback Systems*

Sensitivity to external excitation

C. Shannon




(Extracted from: A Mathematical Theory of Communication)

Maximum rate of reliable information transfer

## Entropy

Consider a random variable **z** with alphabet $\mathcal{Z} = \{1, \ldots, M\}$

There is a representation where we assign $\lceil -\log p_z(z) \rceil$ bits to each z and the expected size becomes:

$$E\left[ \lceil -\log p_z(\mathbf{z}) \rceil \right]$$

## Entropy

Consider a random variable **z** with alphabet $\mathcal{Z} = \{1, \ldots, M\}$

There is a representation where we assign $\lceil -\log p_z(z) \rceil$ bits to each z and the expected size becomes:

$$E\left[\,\lceil -\log p_z(\mathbf{z}) \rceil\,\right]$$

Entropy is the expected size of the most bit-economic representation:

$$H(\mathbf{z}) = E\left[\, -\log p_z(\mathbf{z})\,\right]$$

(valid on average for a large collection of variables)

## Entropy

Consider a random variable **z** with alphabet $\boldsymbol{Z} = \{1,\ldots,M\}$

There is a representation where we assign $\lceil -\log p_z(z) \rceil$ bits to each z and the expected size becomes:

$$E\left[\,\lceil -\log p_z(\mathbf{z}) \rceil\,\right]$$

Entropy is the expected size of the most bit-economic representation:

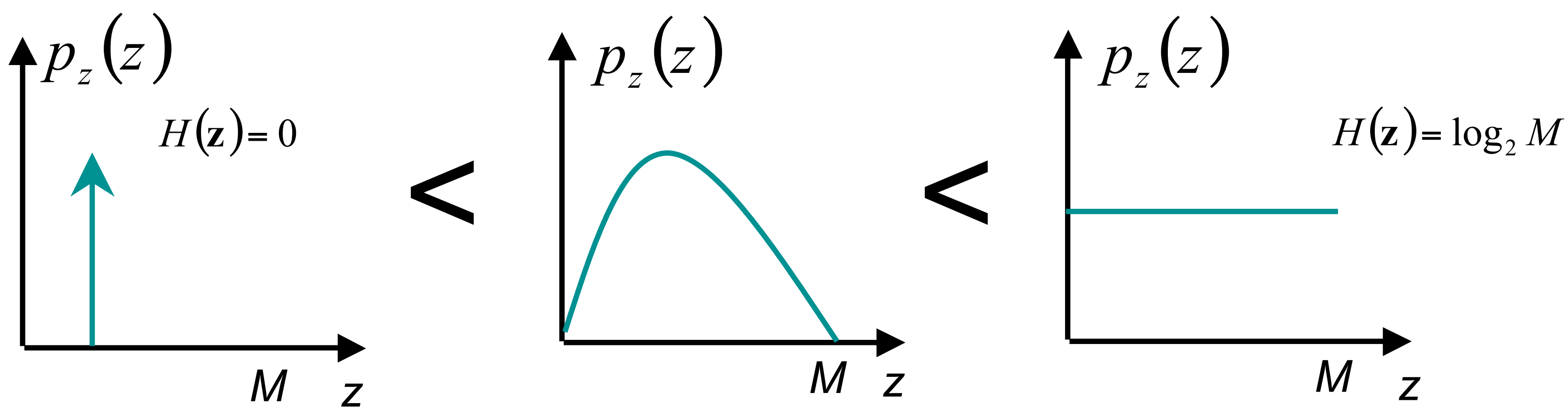
$$H(\mathbf{z}) = -\sum_{z\in\mathbf{Z}} p_z(z)\log p_z(z)$$

## Properties of Entropy:

$$H(\mathbf{z}) \geq 0$$

**z** can be represented as $2^{H(\mathbf{z})}$ uniformly distributed binary random variables

Measure of randomness

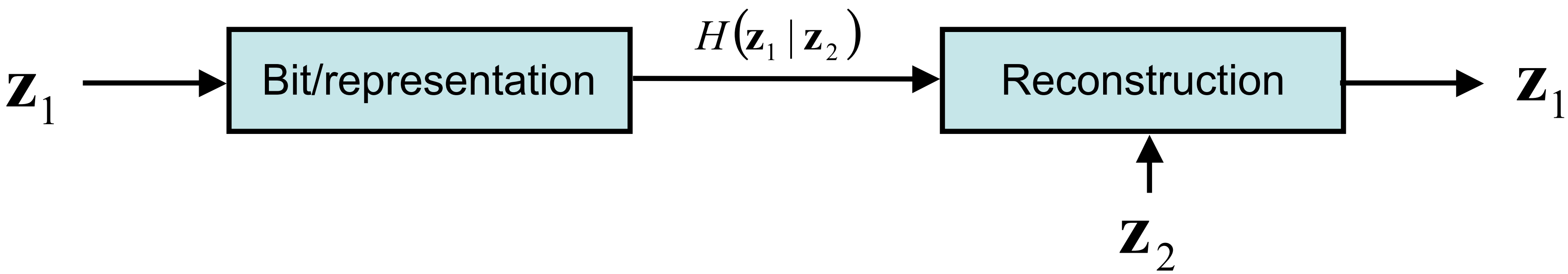


# Information Theory Basic Quantities

## Properties of Entropy:

### Conditional Entropy

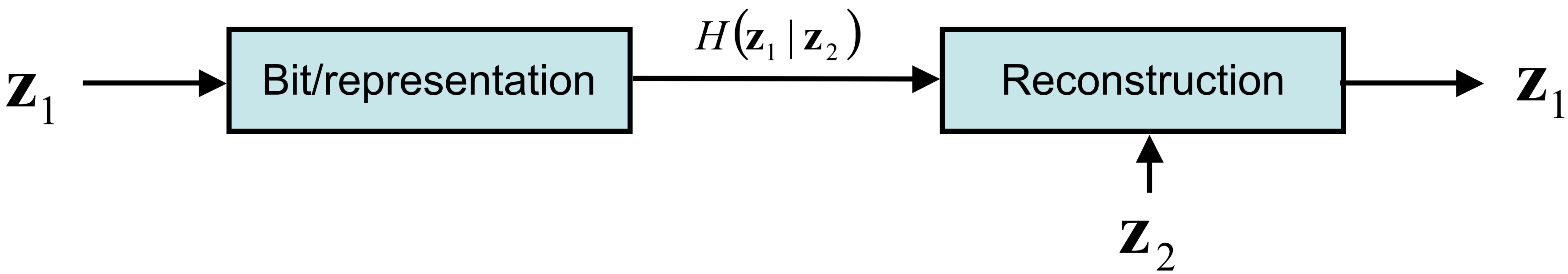
$$H(\mathbf{z}_1 \mid \mathbf{z}_2) = H(\mathbf{z}_1, \mathbf{z}_2) - H(\mathbf{z}_2) = E\left[-\log p_{Z_1|Z_2}(\mathbf{z}_1 \mid \mathbf{z}_2)\right]$$



## Properties of Entropy:

### Conditional Entropy

$$H(\mathbf{z}_1 \mid \mathbf{z}_2) = H(\mathbf{z}_1, \mathbf{z}_2) - H(\mathbf{z}_2) = E\left[-\log p_{Z_1|Z_2}(\mathbf{z}_1 \mid \mathbf{z}_2)\right]$$




How much information does $\mathbf{z}_2$ carry about $\mathbf{z}_1$ ?

$$I(\mathbf{z}_1, \mathbf{z}_2) = H(\mathbf{z}_1) - H(\mathbf{z}_1 \mid \mathbf{z}_2)$$

## “Continuous” random variables

Assume that **w** is a random variable with alphabet $\mathrm{W} = \Re^n$

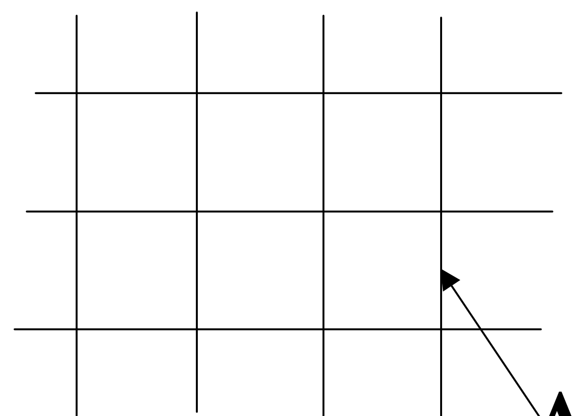
Differential entropy or “entropy density”:

$$2^{h(\mathbf{w})} = \lim_{\Delta \to 0}\left[\Delta 2^{H(f_\Delta(\mathbf{w}))}\right]$$

unif. dist. volume

This many elements unif. dist.

$f_\Delta$ is a quantizer



$\Delta$ sensitivity of the quantizer

“Continuous” random variables

$$I(\mathbf{w}_1, \mathbf{w}_2) \overset{def}{=} \lim_{\Delta \to 0} I\left(f_\Delta(\mathbf{w}_1), f_\Delta(\mathbf{w}_2)\right) = h(\mathbf{w}_1) - h(\mathbf{w}_1 \mid \mathbf{w}_2)$$

The function of the quantizer is to extract as “much” information as possible from the continuous random variables.

"Continuous" random variables

$$I(\mathbf{w}_1, \mathbf{w}_2) \overset{def}{=} \lim_{\Delta \to 0} I\left(f_\Delta(\mathbf{w}_1), f_\Delta(\mathbf{w}_2)\right) = h(\mathbf{w}_1) - h(\mathbf{w}_1 \mid \mathbf{w}_2)$$

The function of the quantizer is to extract as "much" information as possible from the continuous random variables.

"Continuous" and discrete random variables

$$I(\mathbf{z}, \mathbf{w}) = \lim_{\Delta \to 0} I\left(f_\Delta(\mathbf{z}), \mathbf{w}\right) = h(\mathbf{z}) - h(\mathbf{z} \mid \mathbf{w})$$

Properties:

I (Positivity)

$$I(\mathbf{w},\mathbf{z}) = I(\mathbf{z},\mathbf{w}) \geq 0$$

Properties:

I (Positivity)

$$I(\mathbf{w},\mathbf{z}) = I(\mathbf{z},\mathbf{w}) \geq 0$$

II (Kolmogorov's Formula)

Given **v**, how much more information about **z** can I get from **w**?

$$I(\mathbf{w},\mathbf{z} \mid \mathbf{v}) = I((\mathbf{w},\mathbf{v}),\mathbf{z}) - I(\mathbf{v},\mathbf{z}) = h(\mathbf{z} \mid \mathbf{v}) - h(\mathbf{z} \mid \mathbf{w},\mathbf{v})$$

Properties:

I (Positivity)

$$I(\mathbf{w},\mathbf{z}) = I(\mathbf{z},\mathbf{w}) \geq 0$$

II (Kolmogorov's Formula)

Given **v**, how much more information about **z** can I get from **w**?

$$I(\mathbf{w},\mathbf{z}\,|\,\mathbf{v}) = I((\mathbf{w},\mathbf{v}),\mathbf{z}) - I(\mathbf{v},\mathbf{z}) = h(\mathbf{z}\,|\,\mathbf{v}) - h(\mathbf{z}\,|\,\mathbf{w},\mathbf{v})$$

III (maximum entropy bounds)

$$H(\mathbf{a}) \leq \log(\#\mathbf{A})$$

$$h(\mathbf{z}^k) \leq \frac{1}{2}\log\left((2\pi e)^k \left|\Sigma_{z^k}\right|\right) \leq \frac{1}{2}\sum_{i=1}^{k}\log\left(2\pi e\sigma_{z(i)}^2\right)$$

Equality is achieved if $\mathbf{z}^k$ is Gaussian

Equality is achieved if $\mathbf{z}(i)$ are uncorrelated

Important Aspects about (differential) entropy and Mutual Information

Equality if g is injective $\longrightarrow$

$$h(a \mid b) \leq h\left(a - f \circ g(b) \mid g(b)\right)$$

Cannot Reduce Uncertainty without more information

Equality under Invertible f and g $\longrightarrow$

$$I(a,b) \geq I\left(g(a), f(b)\right)$$

data processing inequality

Limit on the ability To send Information

$$h(a - f \circ g(b) \mid g(b)) \geq h(a) - I(a,b)$$

Limit on the ability to reduce uncertainty

# Information Theory Basic Quantities

Definition in terms of rates:

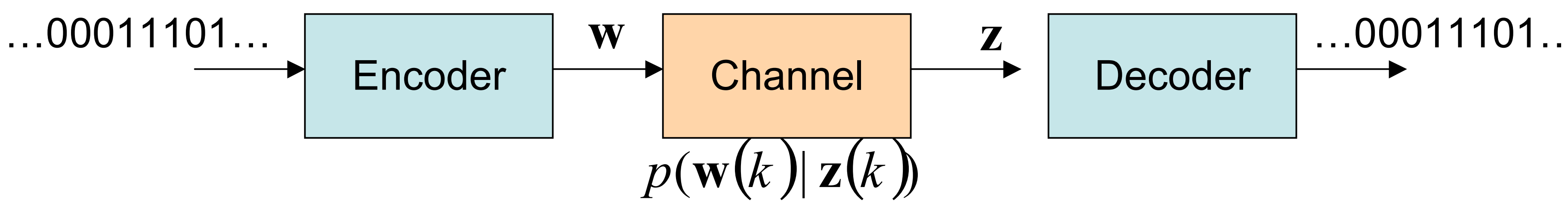
Consider two “continuous” or discrete stochastic processes:

$$\mathbf{w}^k = \left(\mathbf{w}(0),\ldots,\mathbf{w}(k)\right)$$

$$\mathbf{z}^k = \left(\mathbf{z}(0),\ldots,\mathbf{z}(k)\right)$$

(Information rate) $$I_\infty(\mathbf{w},\mathbf{z}) = \lim_{k\to\infty} \frac{I\left(\mathbf{z}^k,\mathbf{w}^k\right)}{k}$$

Maximum reliable bit-rate.

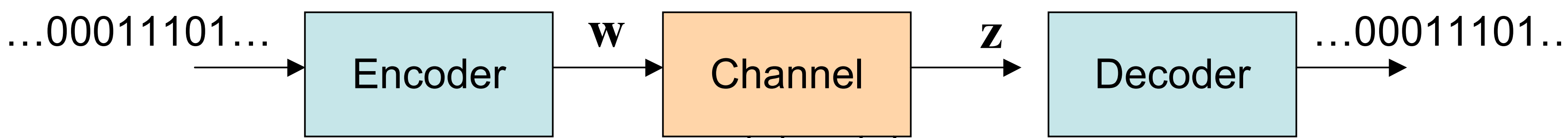



(Channel coding Theorem)

# Channel Coding Theorem

## Channel Capacity

(Information rate) $$I_\infty(\mathbf{w},\mathbf{z}) = \lim_{k\to\infty} \frac{I(\mathbf{z}^k,\mathbf{w}^k)}{k}$$ Rate of reliable transmission




$$p(\mathbf{w}(k) \mid \mathbf{z}(k))$$

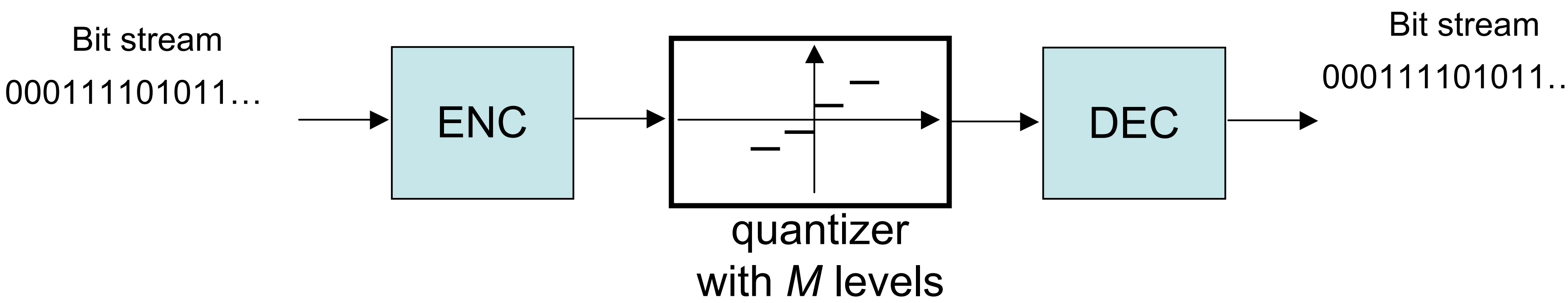
(Channel coding Theorem)

Given a (constrained) set of stochastic processes $S_{\mathbf{w}}$, channel capacity is given by:

$$C = \sup_{\mathbf{w}\in S_{\mathbf{w}}} \lim_{k\to\infty} \frac{I(\mathbf{z}^k,\mathbf{w}^k)}{k} \underset{memoryless}{=} \sup_{p_{\mathbf{w}(k)}} I(\mathbf{z}(k),\mathbf{w}(k))$$

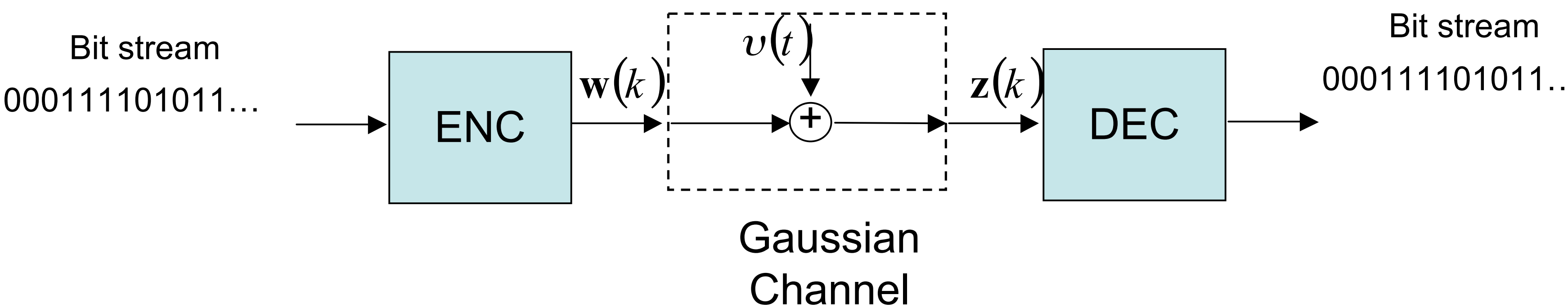
# Channel Coding Theorem

## Examples: Quantizer with M levels




$$C = \log_2 M$$

## Examples: Gaussian channel

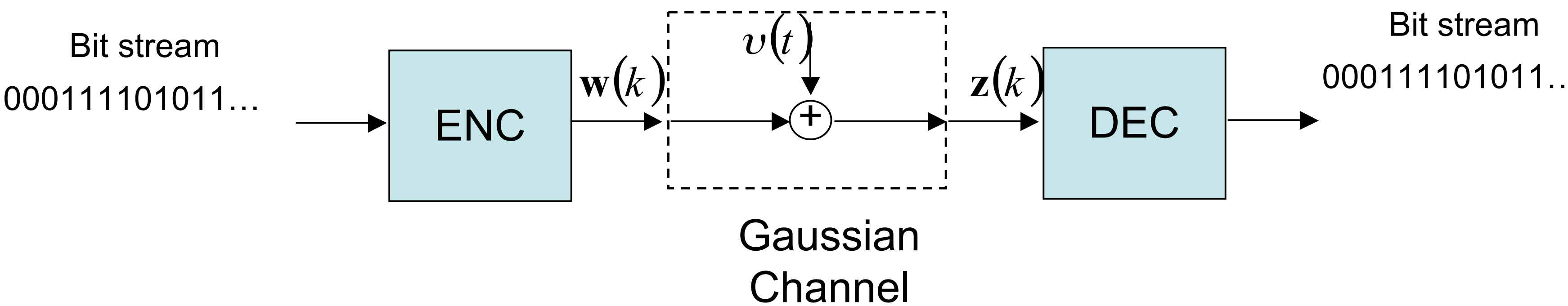



$$C = \frac{1}{2}\log_2\left(1 + \frac{\sigma_w^2}{\sigma_v^2}\right) \longleftarrow \text{Shannon Capacity}$$

# Channel Coding Theorem

## Examples: Gaussian channel in more detail

Information capacity is the supremum of the bit-rate for which information can be transmitted through a medium:




If $N(k)$ represents the total number of bits transmitted up to time *k* then we know that

$$\begin{cases} \sup_k \dfrac{N(k)}{k} \leq \dfrac{1}{2}\log_2\left(1+\dfrac{\sigma_w^2}{\sigma_\upsilon^2}\right) \longleftarrow \text{Shannon Capacity} \\ P[Error(k)] \rightarrow 0 \end{cases}$$

## Rates and bounds

Definition in terms of rates:

Consider two “continuous” or discrete stochastic processes:

$$\mathbf{w}^k = \left(\mathbf{w}(0),\ldots,\mathbf{w}(k)\right)$$

$$\mathbf{z}^k = \left(\mathbf{z}(0),\ldots,\mathbf{z}(k)\right)$$

(Information rate) $$I_\infty(\mathbf{w},\mathbf{z}) = \lim_{k\to\infty}\frac{I\left(\mathbf{z}^k,\mathbf{w}^k\right)}{k}$$

Maximum reliable bit-rate: used to define capacity.

(Entropy rate) $$h_\infty(\mathbf{z}) = \lim_{k\to\infty}\frac{h\left(\mathbf{z}^k\right)}{k}$$

$$h_\infty(\mathbf{z}) \le \frac{1}{4\pi}\int_{-\pi}^{\pi}\log\left(2\pi e F_z(\omega)\right)d\omega$$

Equality is achieved if $\mathbf{z}$ is Gaussian

Consider the following scalar linear and time-invariant system:

$$\mathbf{x}(k+1) = a\mathbf{x}(k) + \mathbf{u}(k) \qquad k \geq 0 \qquad |a| > 1$$

$\mathbf{x}(0)$ is a random variable uniformly distributed in the interval $[-1,1]$

Consider the following scalar linear and time-invariant system:

$$\mathbf{x}(k+1) = a\mathbf{x}(k) + \mathbf{u}(k) \qquad k \geq 0 \qquad |a| > 1$$

$\mathbf{x}(0)$ is a random variable uniformly distributed in the interval $[-1,1]$

We will prove that:

$$\sup_k E\left[\mathbf{x}^2(k)\right] < \beta \quad \Rightarrow \quad \lim_{k \to \infty} \frac{I\left(\mathbf{u}^k, \mathbf{x}(0)\right)}{k} \geq \log|a|$$

## Example of application

Proof:

$$\mathbf{x}(k+1) = a\mathbf{x}(k) + \mathbf{u}(k)$$

The state can be computed as:

$$\mathbf{x}(k) = a^k \mathbf{x}(0) + f\left(k, \mathbf{u}^k\right)$$

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right) \qquad \tilde{f}\left(k, \mathbf{u}^k\right) = -a^{-k} f\left(k, \mathbf{u}^k\right)$$

can be viewed as an estimate of $\mathbf{x}(0)$

## Example of application

Proof:

$$\mathbf{x}(k+1) = a\mathbf{x}(k) + \mathbf{u}(k)$$

The state can be computed as:

$$\mathbf{x}(k) = a^k\mathbf{x}(0) + f(k,\mathbf{u}^k)$$

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}(k,\mathbf{u}^k) \qquad \tilde{f}(k,\mathbf{u}^k) = -a^{-k} f(k,\mathbf{u}^k)$$

can be viewed as an estimate of $\mathbf{x}(0)$

Now, using the formula for mutual information:

$$I(\mathbf{x}(0), \tilde{f}(k,\mathbf{u}^k)) = h(\mathbf{x}(0)) - h(\mathbf{x}(0) \mid \tilde{f}(k,\mathbf{u}^k))$$

$$I(\mathbf{x}(0), \tilde{f}(k,\mathbf{u}^k)) = 2 - h(\mathbf{x}(0) - \tilde{f}(k,\mathbf{u}^k) \mid \tilde{f}(k,\mathbf{u}^k)) \geq 2 - h(\mathbf{x}(0) - \tilde{f}(k,\mathbf{u}^k))$$

## Example of application

Proof (Continuation):

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right) \qquad\qquad \sup_k E\left[\mathbf{x}^2(k)\right] < \beta$$

Now, using the formula for mutual information:

$$I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^k\right)\right) = h(\mathbf{x}(0)) - h\left(\mathbf{x}(0) \mid \tilde{f}\left(k, \mathbf{u}^k\right)\right)$$

$$I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^k\right)\right) = 2 - h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right) \mid \tilde{f}\left(k, \mathbf{u}^k\right)\right) \geq 2 - h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right)\right)$$

# Example of application

Proof (Continuation):

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right) \qquad\qquad \sup_k E\left[\mathbf{x}^2(k)\right] < \beta$$

Now, using the formula for mutual information:

$$I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^k\right)\right) = h(\mathbf{x}(0)) - h\left(\mathbf{x}(0) \mid \tilde{f}\left(k, \mathbf{u}^k\right)\right)$$

$$I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^k\right)\right) = 2 - h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right) \mid \tilde{f}\left(k, \mathbf{u}^k\right)\right) \geq 2 - h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right)\right)$$

Now notice that:

$$h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right)\right) \leq \frac{1}{2}\log\left(2\pi e E\left[\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^k\right)\right)^2\right]\right) \leq \frac{1}{2}\log\left(2\pi e a^{-2k}\beta\right)$$

Proof (Continuation):

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}(k, \mathbf{u}^k) \qquad \sup_k E\left[\mathbf{x}^2(k)\right] < \beta$$

Now, using the formula for mutual information:

$$I\left(\mathbf{x}(0), \tilde{f}(k, \mathbf{u}^k)\right) = h(\mathbf{x}(0)) - h\left(\mathbf{x}(0) \mid \tilde{f}(k, \mathbf{u}^k)\right)$$

Proof (Continuation):

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}(k, \mathbf{u}^k) \qquad \sup_k E\left[\mathbf{x}^2(k)\right] < \beta$$

Now, using the formula for mutual information:

$$I\left(\mathbf{x}(0), \tilde{f}(k, \mathbf{u}^k)\right) = h(\mathbf{x}(0)) - h\left(\mathbf{x}(0) \mid \tilde{f}(k, \mathbf{u}^k)\right)$$

$$h\left(\mathbf{x}(0) - \tilde{f}(k, \mathbf{u}^k)\right) \leq \frac{1}{2}\log\left(2\pi e a^{-2k}\beta\right)$$

## Example of application

Proof (Continuation):

$$a^{-k}\mathbf{x}(k) = \mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^{k}\right) \qquad \sup_{k} E\left[\mathbf{x}^{2}(k)\right] < \beta$$

Now, using the formula for mutual information:

$$I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^{k}\right)\right) = h(\mathbf{x}(0)) - h\left(\mathbf{x}(0) \mid \tilde{f}\left(k, \mathbf{u}^{k}\right)\right)$$

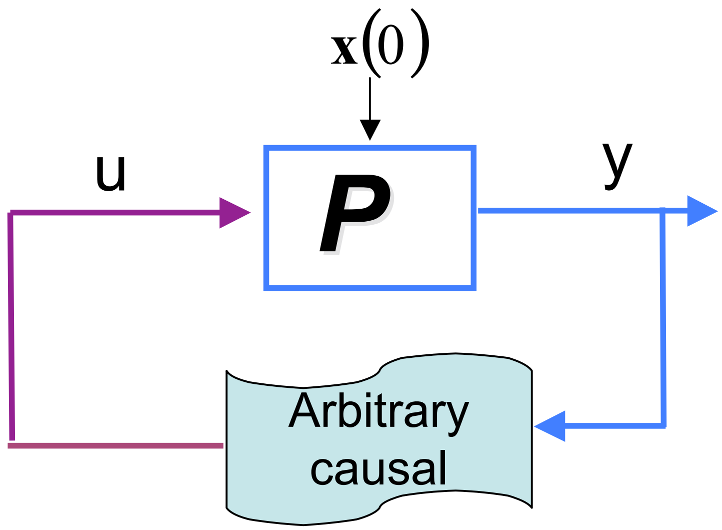
$$h\left(\mathbf{x}(0) - \tilde{f}\left(k, \mathbf{u}^{k}\right)\right) \leq \frac{1}{2}\log\left(2\pi e a^{-2k}\beta\right)$$

$$\lim_{k \to \infty} \frac{I\left(\mathbf{x}(0), \tilde{f}\left(k, \mathbf{u}^{k}\right)\right)}{k} \geq \log|a|$$

data processing inequality: $\Rightarrow$ $$\lim_{k \to \infty} \frac{I\left(\mathbf{x}(0), \mathbf{u}^{k}\right)}{k} \geq \log|a|$$

Example of application

The previous analysis leads to an alternative proof for the following result of Tatikona and Mitter (00):




If the feedback interconnection is second moment stable then:

$$\lim_{k \to \infty} \frac{I(\mathbf{x}(0), \mathbf{u}^k)}{k} \geq \sum_{unstable} \log |pole_i|$$