From thermodynamics to information theory

Susanne Still
University of Hawaii at Manoa

Dissipation in the process of data representation and utilization

- Dissipation is lower bound by information trade-off between total memory and relevant information:

$$Q_{\text{out}} - Q_{\text{in}} \geq kT_1 I[X,Y] - kT_2 I[Y,Z]$$

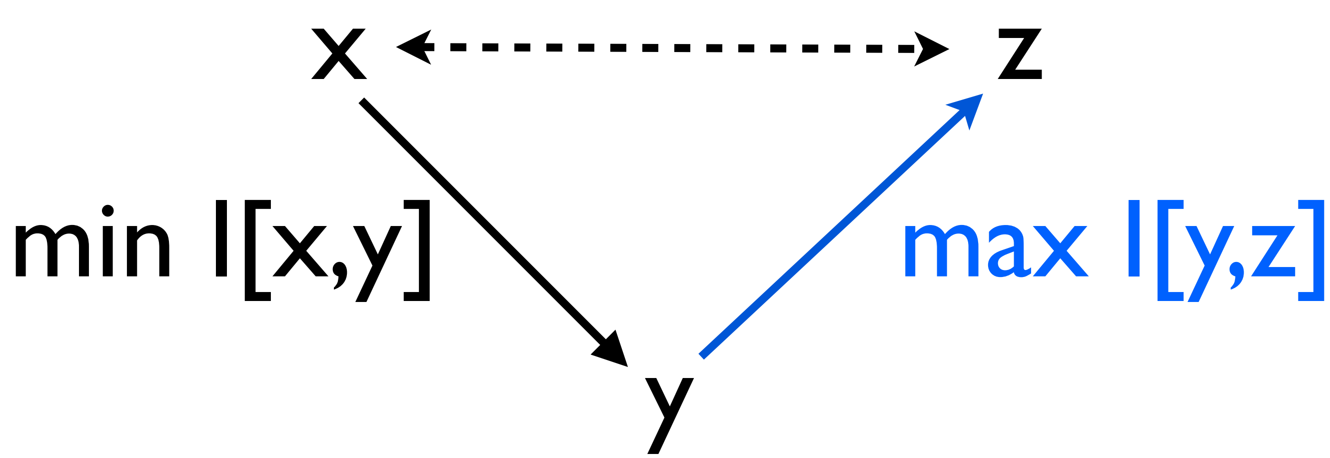
Measurement instrument (T_1) — Work medium (T_2)

- Minimize the bound over all data representations:

$$\min_{p(y|x)} \left(I[X,Y] - \frac{T_2}{T_1} I[Y,Z] \right)$$

This is the optimization performed by the *"Information Bottleneck" method (introduced by Tishby, Pereira and Bialek in 1999)*

Information Bottleneck

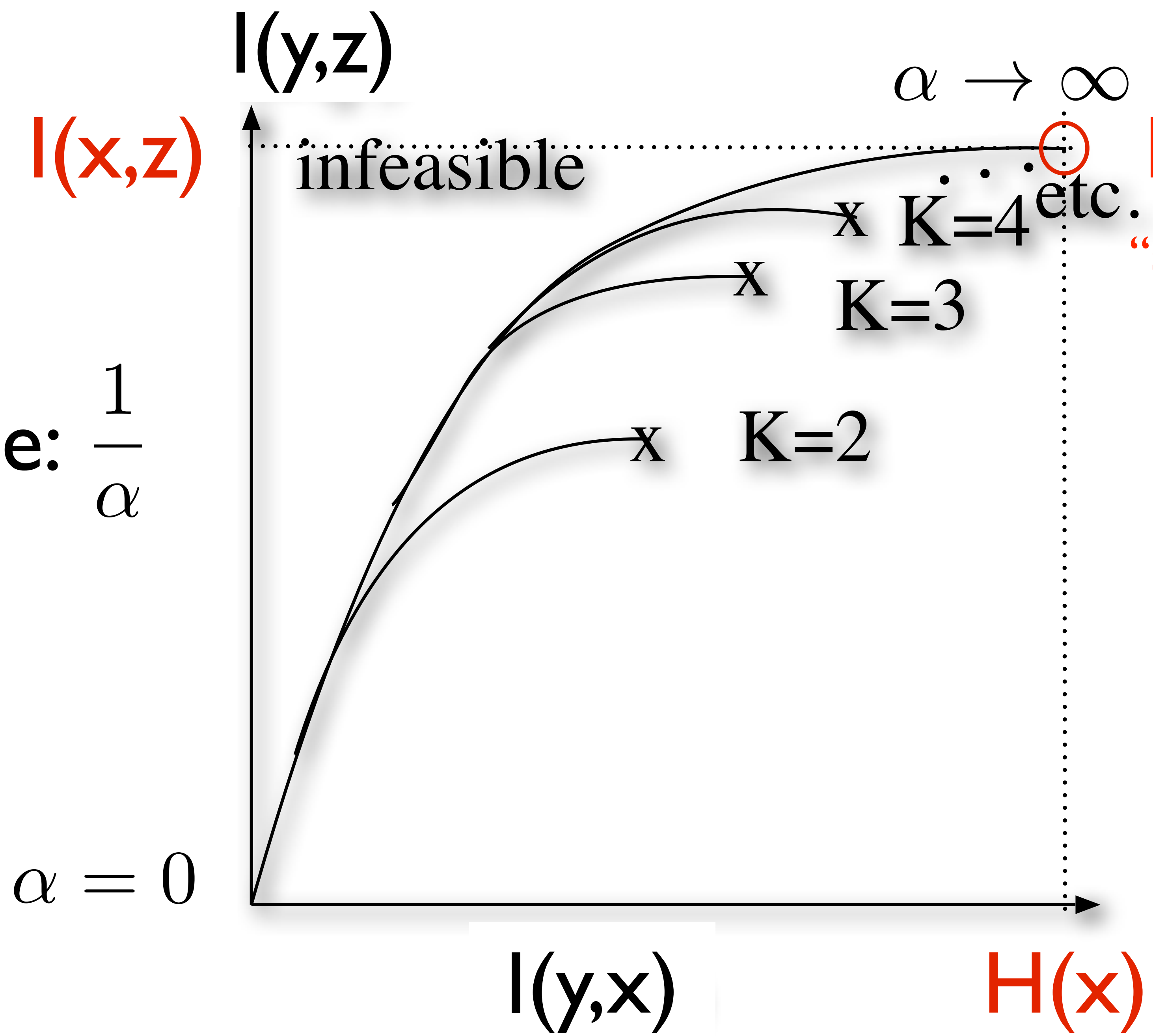


- Optimization: $\min_{p(y|x)} \left(I[X,Y] - \alpha I[Y,Z] \right)$
- Optimal data representation has to satisfy:

$$p(y|x) = \frac{p(y)}{Z(x,\alpha)} e^{-\alpha D_{KL}[p(z|x) \| p(z|y)]}$$

Information plane

Slope: $\frac{1}{\alpha}$



$T_1 \to 0$

$\alpha = \frac{T_2}{T_1}$

$T_1 \to \infty$

Limits

$$p(y|x) = \frac{p(y)}{Z(x,\alpha)} e^{-\alpha D_{KL}[p(z|x) \| p(z|y)]}$$

- inf. noisy data representation: $T_1 \to \infty; \alpha \to 0$

  $p(y|x) = p(y)$ uncorrelated, uniform

- noise free data representation: $T_1 \to 0; \alpha \to \infty$

see Blackboard

Shannon's rate-distortion theory

- Information source is described by p(x) and has information (rate) H[p(x)]
- A **continuous signal** has infinite information rate.
- But infinite resolution is irrelevant for most applications, some level of distortion is tolerable.

Shannon's rate-distortion theory

- *Therefore:* the achievable rate of a continuous information source, if transmitted to finite resolution, i.e. for fixed average distortion is:

$$R(D) := \min_{p(y|x)} I[y, x]$$

$$\text{s.t.} \langle d(y, x) \rangle_{p(y,x)} = D$$

- Represent original signal, x, by encoded signal, s. Given: distortion function d(s,x); information source p(x)
- (units: convert between information in bits per symbol, and rate in bits per second: multiply by a constant - symbols per second)

Shannon's rate-distortion theory

$$R(D) := \min_{p(y|x)} I[x, y]$$
$$\text{s.t.} \langle d(x, y) \rangle_{p(x,y)} = D$$

- Understand this as the work required for the ***least effort data representation***, given the desired level of fidelity.
- Formal connection to *Information Bottleneck*:

$$d(x, y) = D_{\text{KL}}[p(z|x) \| p(z|y)]$$

Shannon's channel capacity

- Channel capacity is the maximally transmittable (rate of) information, given the channel, *maximized over all information sources*

$$\max_{p(x)} I[x, y]$$

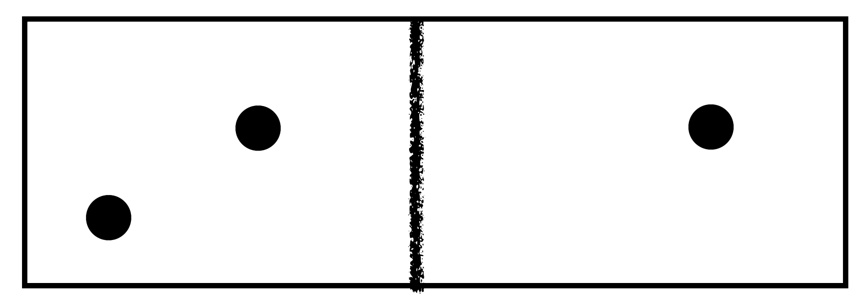
- Can understand this as the maximally extractable work, given the work extraction machinery, *maximized over all physical systems that can be used as work media*

- Alternative to fixing the system and maximizing over all possible work extraction mechanisms

Play with Szillard's box

- where to put partition?

- how many particles?



- how many partitions?