Control and estimation under communication constraints

Alexander Pogromsky

Lecture 2

Technische Universiteit **Eindhoven** University of Technology

October 7, 2021

4 Topological end restoration entropy

4.1 Topological entropy

Given system $\quad x(t+1) = \varphi(x(t)), \qquad \varphi \in C^0, \; x(0) \in K, \quad K$ is compact

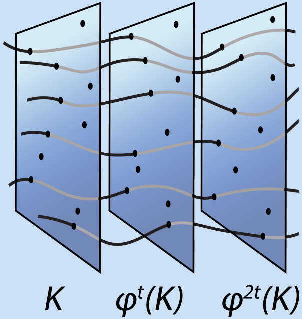
Given the time horizon k, and a finite set of points $\xi_i \in K$, $i = 1, \ldots N$. Consider the sequences

$$\Xi_i^k := \{\xi_i, \varphi(\xi_i), \varphi^2(\xi_i), \ldots, \varphi^k(\xi_i)\} = \{\xi_i, \varphi(\xi_i), \varphi(\varphi(\xi_i)), \ldots, \varphi((\ldots(\varphi(\xi_i))\ldots)\}$$

One aims at approximating an arbitrary sequence of iterations $\zeta, \varphi(\zeta), \ldots, \varphi^k(\zeta)$ by an element from Ξ_i^k.

4.1 Topological entropy

Spanning set



Spanning set

Def. (k,ε) **- spanning set** $\quad \mathcal{P}(\varepsilon,k) \subset K, \quad N := \#\mathcal{P}:$

$$\forall \zeta \in K \;\; \exists \xi \in \mathcal{P}(\varepsilon,k) \;:\; \max_{j=1,\dots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

$$N(\varepsilon,k)$$

4.1 Topological entropy

Spanning set

Def. (k, ε) **- spanning set** $\mathcal{P}(\varepsilon, k) \subset K, \quad N := \#\mathcal{P}:$

$$\forall \zeta \in K \;\; \exists \xi \in \mathcal{P}(\varepsilon, k) \;:\; \max_{j=1,\ldots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

$$\log_2 N(\varepsilon, k)$$

Spanning set

Def. (k, ε) **- spanning set** $\mathcal{P}(\varepsilon, k) \subset K, \quad N := \#\mathcal{P}$:

$$\forall \zeta \in K \;\; \exists \xi \in \mathcal{P}(\varepsilon, k) \;\; : \;\; \max_{j=1,\ldots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

$$\min_{\mathcal{P}} \log_2 N(\varepsilon, k)$$

Spanning set

Def. (k, ε) - spanning set $\quad \mathcal{P}(\varepsilon, k) \subset K, \quad N := \#\mathcal{P}$:

$$\forall \zeta \in K \;\; \exists \xi \in \mathcal{P}(\varepsilon, k) \;:\; \max_{j=1,\dots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

$$\frac{1}{k} \min_{\mathcal{P}} \log_2 N(\varepsilon, k)$$

Spanning set

Def. (k, ε) **- spanning set** $\quad \mathcal{P}(\varepsilon, k) \subset K, \quad N := \#\mathcal{P}$:

$$\forall \zeta \in K \;\; \exists \xi \in \mathcal{P}(\varepsilon, k) \;:\; \max_{j=1,\ldots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

$$\limsup_{k\to\infty} \frac{1}{k} \min_{\mathcal{P}} \log_2 N(\varepsilon, k)$$

Spanning set

Def. (k, ε) **- spanning set** $\mathcal{P}(\varepsilon, k) \subset K, \quad N := \#\mathcal{P}$:

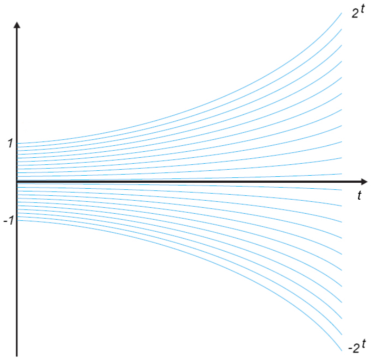
$$\forall \zeta \in K \ \exists \xi \in \mathcal{P}(\varepsilon, k) \ : \ \max_{j=1,\ldots,k} ||\varphi^j(\xi) - \varphi^j(\zeta)|| < \varepsilon$$

Topological entropy:

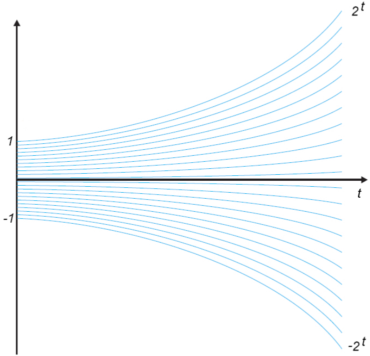
$$H(\varphi, K) = \lim_{\varepsilon \to 0} \limsup_{k \to \infty} \frac{1}{k} \min_{\mathcal{P}} \log_2 N(\varepsilon, k)$$

Example.

$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$



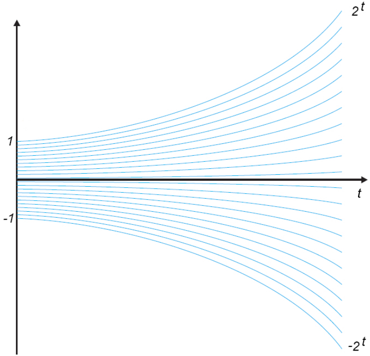
Example.



$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$

$$x(t) \in \varphi^t(K) = [-2^t, 2^t]$$

Example.

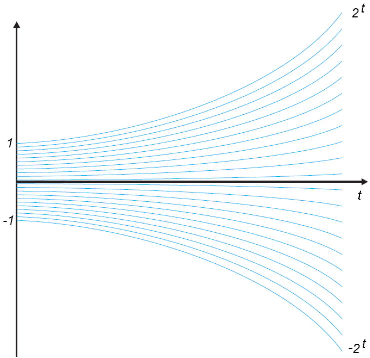


$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$

$$x(t) \in \varphi^t(K) = [-2^t, 2^t]$$

$$N = \frac{2 \cdot 2^t}{\varepsilon}$$

Example.



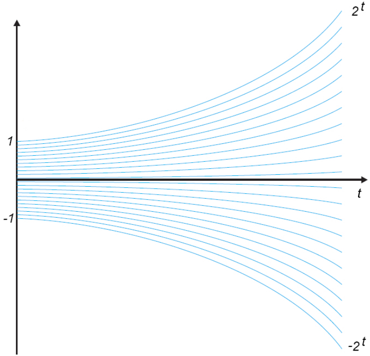
$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$

$$x(t) \in \varphi^t(K) = [-2^t, 2^t]$$

$$N = \frac{2 \cdot 2^t}{\varepsilon}$$

$$\log_2 N = \log_2 \frac{2}{\varepsilon} + t$$

Example.



$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$

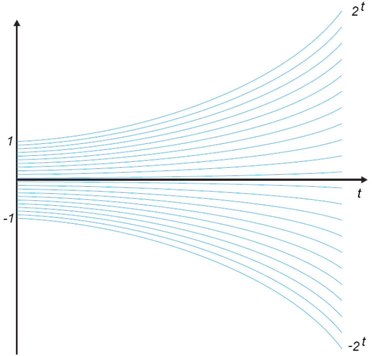
$$x(t) \in \varphi^t(K) = [-2^t, 2^t]$$

$$N = \frac{2 \cdot 2^t}{\varepsilon}$$

$$\log_2 N = \log_2 \frac{2}{\varepsilon} + t$$

$$\frac{1}{t} \log_2 N = \frac{1}{t} \log_2 \frac{2}{\varepsilon} + 1 \quad \rightarrow \quad = 1, \ \text{as } t \rightarrow \infty$$

Example.



$$x(t+1) = 2x(t), \quad \varphi(x) = 2x, \quad K = [-1, 1]$$

$$x(t) \in \varphi^t(K) = [-2^t, 2^t]$$

$$N = \frac{2 \cdot 2^t}{\varepsilon}$$

$$\log_2 N = \log_2 \frac{2}{\varepsilon} + t$$

$$\frac{1}{t} \log_2 N = \frac{1}{t} \log_2 \frac{2}{\varepsilon} + 1 \quad \rightarrow \quad = 1, \text{ as } t \rightarrow \infty$$

$$\boxed{H(\varphi, K) = 1}$$

Quiz time: find the topological entropy

$x(t+1) = 0.5x(t)$

Answers: $-\infty, \quad -1, \quad 0, \quad 1,$ Google

$x(t+1) = ax(t), \quad a > 1$

Answers: $a/2, \quad \log_2 a,$ Google

$x(t+1) = -ax(t), \quad a > 1$

Answers: $-\log_2 a, \quad \log_2 a,$ Google

$x(t+1) = 2x(t), \quad y(t+1) = 2y(t)$

Answers: $1, \quad 2, \quad 4$

4.1 Topological entropy

Linear systems

$$x(t+1) = Ax(t)$$

- H is invariant under linear coordinate change
- $H = \sum_{j,|\lambda_j|>1} \log_2 |\lambda_j|$

Alternative characterization

For any $\varepsilon > 0$ there is a positive definite matrix $P = P^\top > 0$ so that

$$\frac{1}{2} \sum_{j,\alpha_j > 1} \log_2 \alpha_j - \varepsilon \leq H \leq \frac{1}{2} \sum_{j,\alpha_j > 1} \log_2 \alpha_j$$

where α_js, $j = 1, \ldots, n$ are the solutions of

$$\det(A^\top PA - \alpha P) = 0$$

Sketch of the proof

- A is, with no loss in generality, in the (real) Jordan block.
- The matrices

$$J_1 = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix} \quad J_\varepsilon = \begin{bmatrix} \lambda & \varepsilon & 0 & \cdots & 0 \\ 0 & \lambda & \varepsilon & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda & \varepsilon \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

  are similar: $\forall \varepsilon > 0 \ \exists M(\varepsilon), \quad \det M \neq 0, \quad MJ_1M^{-1} = J_\varepsilon$. For J_ε one can take $P = I_n$.
- $\det(M^{-\top}J_1^\top M^\top M J_1 M^{-1} - \alpha I_n) = \det M^{-\top}(J_1^\top M^\top M J_1 - \alpha M^\top M)M = \det(J_1^\top P J_1 - \alpha P)$.
- $P = M^\top M$ gives an upper estimate, which is ε_*-close to the true value.

4.1 Topological Entropy. Linear systems.

Continuous time.

$$\dot{x} = Ax, \quad x \in \mathbb{R}^n, \; 0 \in K$$

Similar to discrete time

$$H = \frac{1}{\ln 2} \sum_{j, \mathrm{Re}\lambda_j > 0} \mathrm{Re}\lambda_j$$

Alternative characterization

For any $\varepsilon > 0$ there is a positive definite matrix $P = P^\top > 0$ so that

$$\frac{1}{2\ln 2}\sum_{j,\alpha_j>0}\alpha_j - \varepsilon \leq H \leq \frac{1}{2\ln 2}\sum_{j,\alpha_j>0}\alpha_j$$

where α_js, $j = 1, \ldots, n$ are the solutions of

$$\det(A^\top P + PA - \alpha P) = 0$$

4.2 Topological entropy and observability via constrained channels

Observability:

$$\forall\varepsilon > 0 \exists\delta > 0 \qquad ||x(0) - \widehat{x}(0)|| \leq \delta \implies ||x(t) - \widehat{x}(t)|| \leq \varepsilon, \qquad \forall t \geq 0$$

Capacity threshold: $\mathcal{R}_o : c > \mathcal{R}_o \implies$ observability.

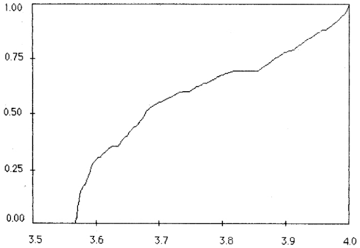
Data rate theorem

$$\mathcal{R}_o(\varphi, K) = H(\varphi, K)$$

The main idea: H estimates the number of bits required to represent the spanning set. So, the coders/decoders have to calculate the spanning set. A very tricky problem.

Example. Logistic map

$$x(t+1) = \mu x(t)(1 - x(t)), \;\; \mu \in \mathbb{R} \;\; - \text{parameter}, \;\; K = [0,1]$$



H vs μ.

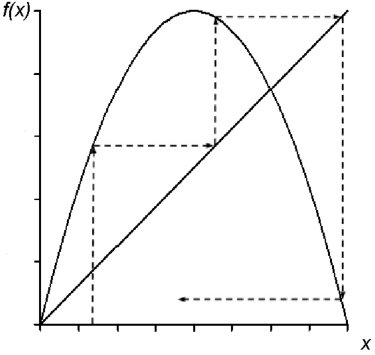
From: L. Block, J. Keesliing, S. Li, K. Petersen, An improved algorithm for computing the topological entropy, Journal of Statistical Physics, 55(5/6), 929-939, 1989

Example. Logistic map

Fixed points:

$$x = \mu x(1-x) \quad \Longrightarrow \quad x_{1e} = 0, \quad x_{2e} = \frac{\mu - 1}{\mu}$$

$$Df(x_{1e}) = \mu, \;\; Df(x_{2e}) = 2 - \mu$$



Instability of the equilibria does not necessarily imply that $H > 0$ (the solutions can eventually settle on one of the periodic orbits).

From the definition of observability: if ε is taken small, then δ has to be even much smaller.

Restoration entropy

Regular observability

$$\exists \delta_*, G > 0 \ \forall \delta \leq \delta_* \qquad ||x(0) - \widehat{x}(0)|| \leq \delta \quad \Longrightarrow \quad ||x(t) - \widehat{x}(t)|| \leq G\delta, \qquad \forall t \geq 0$$

Capacity threshold: $\mathcal{R}_r : c > \mathcal{R}_r \quad \Longrightarrow \quad$ regular observability.

Two issues:

- Uniformity in δ.
- $\varepsilon = G\delta$, so one can conjecture that if there are unstable fixed points, $\mathcal{R}_r > 0$. In other words, $\mathcal{R}_r > H$.

4.3 Restoration entropy

Notations

$\varphi(K) \subset K$, B_a^δ - a δ-ball centered at a. For CT $\varphi^t(\cdot)$ is the flow.
$N(T, a, \delta)$ - a minimal number of δ-balls to cover $\varphi^T(B_a^\delta \bigcap K)$. Recall communication between Alice and Bob.

Definition of the restoration entropy

$$\begin{aligned} H_{\text{res}}(\varphi, K) &:= \overline{\lim_{T\to\infty}} \frac{1}{T} \overline{\lim_{\delta\to 0}} \sup_{a\in K} \log_2 N(T, a, \delta) \\ &= \lim_{T\to\infty} \frac{1}{T} \overline{\lim_{\delta\to 0}} \sup_{a\in K} \log_2 N(T, a, \delta) \end{aligned}$$

Properties of the restoration entropy

Inequality, involving topological and restoration entropy

$$H \leq H_{\text{res}} \qquad \text{strict inequalities are ''more often'' than} \quad =$$

1. C. Kawan, On the Relation between Topological Entropy and Restoration Entropy, *Entropy*, 21(1), 2019.
2. A. Pogromsky, A. Matveev, Data rate limitations for observability of nonlinear systems, *IFAC-PapersOnLine*, 49(14), 2016.

Data rate theorem(s)

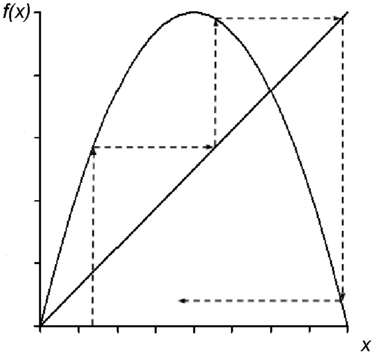
H is a threshold of the channel capacity for observability ($H = \mathcal{R}_o$).
H_{res} is a threshold of the channel capacity for regular(fine) observability ($H_{\text{res}} = \mathcal{R}_r = \mathcal{R}_{fine}$).

1. A. Savkin, Analysis and synthesis of networked control systems: topological entropy, observability, robustness, and optimal control, *Automatica*, 42, 2006.
2. A. Matveev, A. Pogromsky, Observation of nonlinear systems via finite capacity channels, Part II: Restoration entropy and its estimates, *Automatica*, 103, 2019.

Fixed points:

$$x = \mu x(1 - x) \implies x_{1e} = 0, \quad x_{2e} = \frac{\mu - 1}{\mu}$$

$$Df(x_{1e}) = \mu, \;\; Df(x_{2e}) = 2 - \mu$$

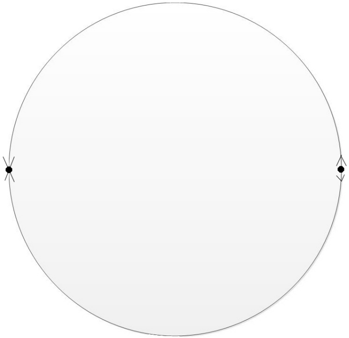


$$\mu > 1 \implies H_{\text{res}}(K) = \log_2 \mu.$$

$$\mu = 4 \implies H_{\text{res}}(K) = 2H(K).$$

More detail to come later.

Quiz time



$$\dot{x} = f(x), \;\; f : S^1 \to S^1, \;\; K = S^1, \;\; \mathrm{D}f(0) = 1, \;\; \mathrm{D}f(\pi) = -1, \;\; H = ?, H_{\text{res}} = ?$$

5. Estimating the entropy

Motivation

Data rate theorem(s) for observability and regular observability:

$$\mathcal{R}_o = H, \quad \mathcal{R}_\mathrm{r} = H_\mathrm{res}$$

H and H_res are substantially different quantities: H is a topological invariant, while H_res is invariant under bi-Lipschitz transformations of φ.

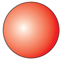
There are characterizations of H (e.g. Pesin formula, outside the scope of this course). However, their practical utilization is cumbersome.

Can we find analytically verifiable conditions to estimate H_res? - any such an upper estimate is also an upper estimate for H.

5. Estimating the entropy

$x(t+1) = \varphi(x(t))$

$x(t+1) = \varphi(x(t))$



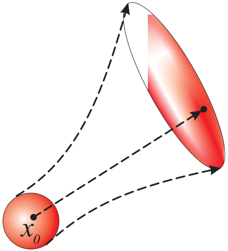
5. Estimating the entropy

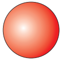
$x(t+1) = \varphi(x(t))$

Alice

$x(t+1) = \varphi(x(t))$

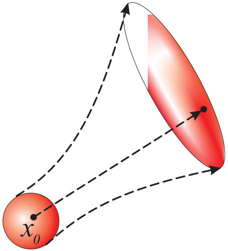
Bob



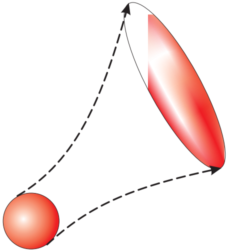


5. Estimating the entropy

$$x(t+1) = \varphi(x(t))$$



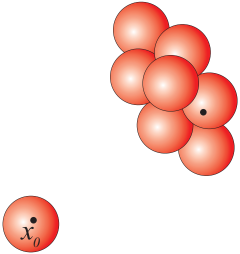
$$x(t+1) = \varphi(x(t))$$



5. Estimating the entropy

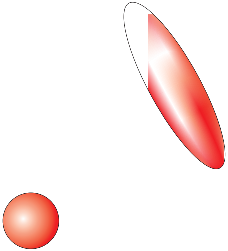
Alice

$x(t+1) = \varphi(x(t))$



$x(t+1) = \varphi(x(t))$

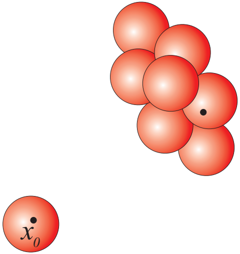
Bob



5. Estimating the entropy

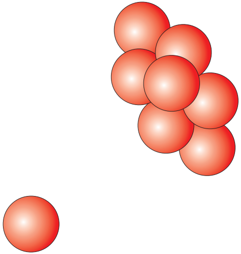
$$x(t+1) = \varphi(x(t))$$

Alice



$$x(t+1) = \varphi(x(t))$$

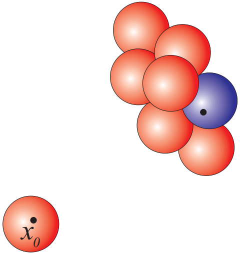
Bob



5. Estimating the entropy

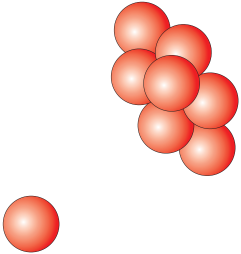
$x(t+1) = \varphi(x(t))$

Alice



$x(t+1) = \varphi(x(t))$

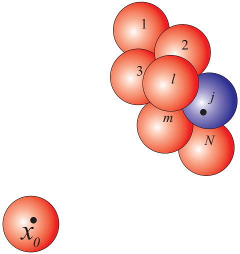
Bob



5. Estimating the entropy

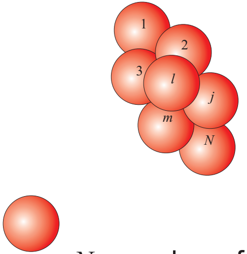
$x(t+1) = \varphi(x(t))$

Alice



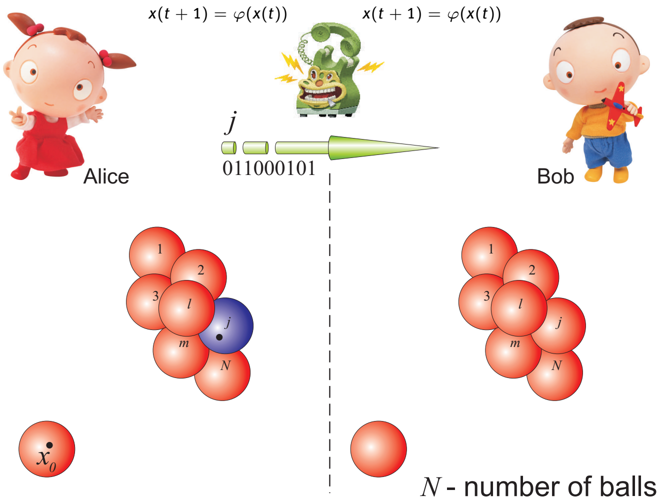
$x(t+1) = \varphi(x(t))$

Bob

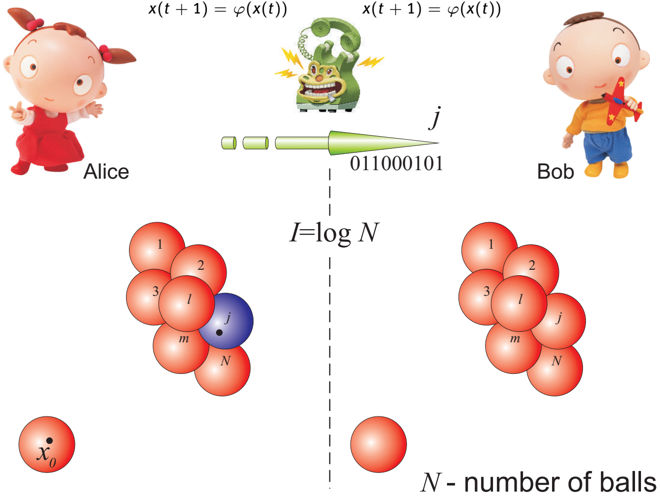


N - number of balls

5. Estimating the entropy



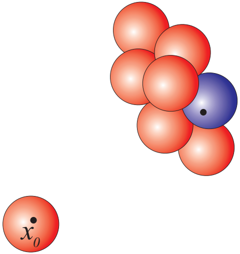
5. Estimating the entropy



5. Estimating the entropy

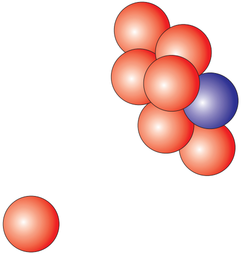
Alice

$$x(t+1) = \varphi(x(t))$$



$$x(t+1) = \varphi(x(t))$$

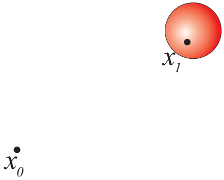
Bob



5. Estimating the entropy

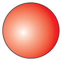
$x(t+1) = \varphi(x(t))$

Alice

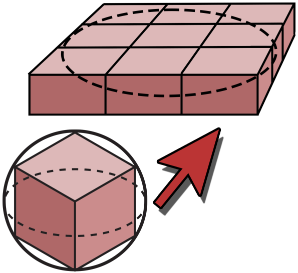


$x(t+1) = \varphi(x(t))$

Bob

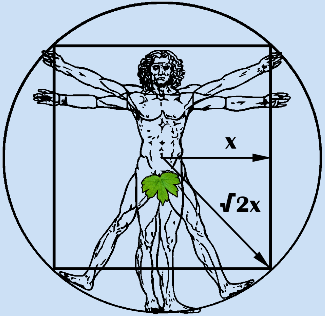


How to (over)-approximate N?



- Exploit C^1-smoothness of φ:
  $\varphi(\xi) - \varphi(\zeta) = \mathrm{D}\varphi(\xi)(\xi - \zeta) + \text{h.o.t.}$
- A linear mapping $\varphi(x) = Ax$ maps a ball into an ellipsoid with semi-axes being the singular values of A.
- Inscribe the ball into a box, so that the image of the ball is contained in the chocolate bar.

Approximation errors



- from $||\cdot||_2$ to $||\cdot||_\infty$ and back
- ceiling/floor effects

5.1 Discrete time

The main trick

- Given T (large enough). The image of a δ-ball centered at ξ can be approximated by an ellipsoid with semi-axes equal to the singular values of $\mathrm{D}\varphi^T(\xi)$.
- Chain rule ($\varphi^T(x) = \varphi(\varphi(\varphi(\ldots(x))\ldots))$):

$$\mathrm{D}\varphi^T(\xi) = \underbrace{\mathrm{D}\varphi(\varphi^{T-1}(\xi))}_{A} \cdot \underbrace{\mathrm{D}\varphi(\varphi^{T-2}(\xi))}_{B} \cdots \underbrace{\mathrm{D}\varphi(\varphi(\xi))}_{Y}\,\underbrace{\mathrm{D}\varphi(\xi)}_{Z}$$

- How to estimate the singular values of $ABC \cdot XYZ$? Answer: the Horn inequality. Let ω_d stand for the product of d first largest singular values. Then

$$\omega_d(AB) \leq \omega_d(A)\omega_d(B)$$

The main trick

- Suppose K is positively invariant: $\varphi(K) \subset K$. Then $\forall \xi \in K$

$$\omega_d(\mathrm{D}\varphi^T(\xi))) \leq \max_{a \in K} [\omega_d(\mathrm{D}\varphi(a))]^T$$

- Recall the definition

$$H_{\text{res}}(\varphi, K) = \lim_{T \to \infty} \frac{1}{T} \overline{\lim_{\delta \to 0}} \sup_{a \in K} \log_2 N(T, a, \delta)$$

N is ($\leq$) proportional to the product of largest first (and > 1) singular values of $\omega_d(\mathrm{D}\varphi^T(\xi)))$. Then lim's will kill the approximation errors due to log in the definition.

The main trick

- Finally,

$$H_{\text{res}}(\varphi, K) \leq \max_{x \in K} \max_{d} \omega_d(\mathrm{D}\varphi(x))$$

- Pro: constructive
- Con: coordinate-dependent. The right-hand side will change if one applies a linear change $x = Mz$

How to make the estimate coordinate independent?

- Recall the alternative characterization of the entropy of linear systems
- We can apply the same trick: suppose there is a $P = P^\top > 0$

$$A(x) := \mathrm{D}\varphi(x), \quad \alpha_i(x) = \sqrt{\lambda_i(x)}, \; \det(A^\top P A - \lambda P) = 0$$

$$H_{\mathrm{res}} \leq \max_{x \in K} \sum_{i=1}^{n} \max\{0, \log_2 \alpha_i(x)\}$$

- A linear change will change P. What to do to make the result independent with respect to a nonlinear (bi-Lipschitz) transformation? - Allow P to depend on x: $P = P(x)$.

How to make the estimate coordinate independent?

Suppose there is a $P(x) = P(x)^\top > 0,\ P(\cdot) \in C^1$

$$A(x) := \mathrm{D}\varphi(x), \quad \alpha_i(x) = \sqrt{\lambda_i(x)}, \ \det(A^\top P(\varphi(x))A - \lambda P(x)) = 0$$

$$H_{\text{res}} \leq \max_{x \in K} \sum_{i=1}^{n} \max\{0, \log_2 \alpha_i(x)\}$$

5.1 Discrete time

Converse statement

Assume that $K = \mathrm{cl}(\mathrm{int}K)$, $\mathrm{D}\varphi(x)$ is invertible $\forall x \in K$.

$$\forall \varepsilon > 0 \;\; \exists P(x) \in C^0, \;\; P(x) = P(x)^\top > 0, \;\; H_{\mathrm{res}} \geq \max_{x \in K} \sum_{i=1}^{n} \max\{0, \log_2 \alpha_i(x)\} - \varepsilon$$

In other words,

$$H_{\mathrm{res}}(\varphi, K) = \inf_{P \in C^0} \max_{x \in K} \sum_{i=1}^{n} \max\{0, \log_2 \alpha_i(x)\}.$$

C. Kawan, A. Matveev, A. Pogromsky, “Remote state estimation problem: Towards the data-rate limit along the avenue of the second Lyapunov method,” *Automatica*, 125, 2021

A lower estimate

Suppose K has a non-void interior and there is a fixed point $x_e \in \mathrm{int}(K)$.

Let $H_{\mathrm{loc}}(x_e)$ be the entropy of the linear system with $A = \mathrm{D}\varphi(x_e)$.

Then $H_{\mathrm{res}}(K) \geq H_{\mathrm{loc}}(x_e)$.

Logistic map

$$\varphi_\mu \ : \ \{x \mapsto \mu x(1-x)\}, \quad K = [0,1]$$

$$\mathrm{D}\varphi_\mu(x) = \mu(1-2x), \quad P = 1 \implies \mu^2(1-2x)^2 = \lambda(x)$$

$$\max_{x\in[0,1]} \lambda(x) = \mu^2$$

$$H_{res}(K) \le \log_2 \mu$$

The result from the previous slide does not allow to claim $=$ (the maximum is attained at the equilibrium $x_e = 0$, but this equilibrium is not in the interior of $[0,1]$).
The assumption $x_e \in \mathrm{int}(K)$ can be relaxed to prove the equality in this situation.