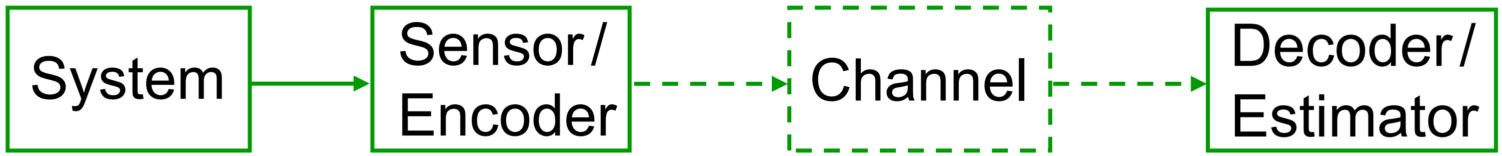
ENTROPY NOTIONS for STATE ESTIMATION and MODEL DETECTION with FINITE-DATA-RATE MEASUREMENTS

Daniel Liberzon and Sayan Mitra

Coordinated Science Laboratory and
Dept. of Electrical & Computer Eng.,
Univ. of Illinois at Urbana-Champaign

CDC, Las Vegas, Dec 2016

MOTIVATION and PROBLEM FORMULATION



State estimation:

How much data rate is needed to estimate system state with error converging to 0 at desired exponential rate?

Model detection:

How much data rate is needed to distinguish between two possible system models?

(Will be able to apply state estimation scheme if trajectories of two models are different enough)

Desired data rate will be described by estimation entropy

(variant of previous entropy notions for control and estimation [Nair, Evans et al.; Colonius, Kawan; Leonov, Matveev, Savkin et al.])

ENTROPY DEFINITION

$\dot{x} = f(x), \quad x \in \mathbb{R}^n, \; x(0) \in K \subset \mathbb{R}^n$ (known compact set)

$\xi(x,t)$ – solution from initial state x after time t

$\alpha > 0$ – desired exponential convergence rate (fixed)

Pick time horizon $T > 0$ and resolution $\varepsilon > 0$
(eventually $T \to \infty$ & $\varepsilon \to 0$)

Call a set of functions $\hat{x}_1, \ldots, \hat{x}_N : [0,T] \to \mathbb{R}^n$
(T,ε)-approximating if $\forall x \in K \; \exists \, \hat{x}_i(\cdot)$:

$$|\xi(x,t) - \hat{x}_i(t)| < \varepsilon e^{-\alpha t} \quad \forall t \in [0,T]$$

$s_{\text{est}}(T,\varepsilon) :=$ size N of smallest (T,ε)-approximating set

Estimation entropy:

$$h_{\text{est}} := \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T,\varepsilon)$$

Intuition: average # of bits needed to identify approx. functions

ALTERNATIVE ENTROPY DEFINITION

$\dot{x} = f(x), \quad x \in \mathbb{R}^n, \ x(0) \in K \subset \mathbb{R}^n$ (known compact set)

$\xi(x,t)$ – solution from initial state x after time t

$\alpha > 0$ – desired exponential convergence rate (fixed)

Can use system trajectories as approximating functions

Call a set of initial states $x_1, \ldots, x_N \in K$
(T, ε)-spanning if $\forall x \in K \ \exists \, x_i$:

$$|\xi(x,t) - \xi(x_i,t)| < \varepsilon e^{-\alpha t} \quad \forall t \in [0,T]$$

$s^*_{\text{est}}(T,\varepsilon) :=$ size N of smallest (T,ε)-spanning set

Alternative entropy notion:

$$h^*_{\text{est}} := \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log s^*_{\text{est}}(T,\varepsilon)$$

Intuition: $x_1, \ldots, x_N$ are quantization points, $h^*_{\text{est}} =$ bit rate

COMPARING TWO ENTROPY NOTIONS

A (T, ε)-spanning set of points $x_1, \ldots, x_N$ can be used to define a (T, ε)-approximating set of functions $\widehat{x}_i(t) := \xi(x_i, t)$

$\Rightarrow$ size of smallest (T, ε)-approximating set

$\leq$ size of smallest (T, ε)-spanning set

$\Rightarrow \ h_{\text{est}} \leq h^*_{\text{est}}$

Question: Can it be that $h_{\text{est}} < h^*_{\text{est}}$?

Answer: No!

Working with arbitrary (even discontinuous) approximating functions gives the same entropy as working only with system trajectories

COMPARING TWO ENTROPY NOTIONS

A (T, ε)-spanning set of points $x_1, \ldots, x_N$ can be used to define a (T, ε)-approximating set of functions $\widehat{x}_i(t) := \xi(x_i, t)$

$\Rightarrow$ size of smallest (T, ε)-approximating set

$\leq$ size of smallest (T, ε)-spanning set

$\Rightarrow \; h_{\text{est}} \leq h^*_{\text{est}}$

Theorem: $\boxed{h_{\text{est}} = h^*_{\text{est}}}$

Proof relies on (T, ε)-separated sets

A set of points $x_1, \ldots, x_M \in K$ is (T, ε)-separated if $\forall x_1, x_2$:

$$|\xi(x_1, t) - \xi(x_2, t)| \geq \varepsilon e^{-\alpha t} \quad \text{for some } \; t \in [0, T]$$

$n^*_{\text{est}}(T, \varepsilon) :=$ size M of largest (T, ε)-separated set

Can show (using standard techniques):

$$n^*_{\text{est}}(T, 2\varepsilon) \leq s_{\text{est}}(T, \varepsilon) \leq s^*_{\text{est}}(T, \varepsilon) \leq n^*_{\text{est}}(T, \varepsilon)$$

Divide by T, take $\limsup_{T \to \infty}$, then $\lim_{\varepsilon \to 0}$ – all become equal ◻

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L :=$ Lipschitz constant of f

Note: for $f \in C^1$ can replace L by upper bound on matrix measure of Jacobian $\partial f/\partial x$ to refine this bound

Theorem: $\boxed{h_{\text{est}} \le (L+\alpha)n}$

Claim: centers of balls of radius $\varepsilon e^{-(L+\alpha)T}$ that cover K form a (T,ε)-spanning set

Lemma: $|\xi(x_1,t) - \xi(x_2,t)| \le |x_1 - x_2|e^{Lt} \quad \forall t \ge 0,\ \forall x_1, x_2$

If $|x_1 - x_2| \le \varepsilon e^{-(L+\alpha)T}$ then $|\xi(x_1,t) - \xi(x_2,t)|$
$\le \varepsilon e^{-(L+\alpha)T} e^{Lt} \le \varepsilon e^{-\alpha t} \quad \forall t \in [0,T]$ which proves the Claim

How many balls of radius $\varepsilon e^{-(L+\alpha)T}$ are needed to cover K?

For K a hypercube of size ℓ, need $\left(\dfrac{\ell}{\varepsilon e^{-(L+\alpha)T}}\right)^n = \dfrac{\ell^n e^{(L+\alpha)Tn}}{\varepsilon^n}$

Taking $\frac{1}{T}\log$ gives $(L+\alpha)n$ $\blacksquare$

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L :=$ Lipschitz constant of f

Note: for $f \in C^1$ can replace L by upper bound on matrix measure of Jacobian $\partial f / \partial x$ to refine this bound

Theorem: $\boxed{h_{\text{est}} \leq (L + \alpha)n}$

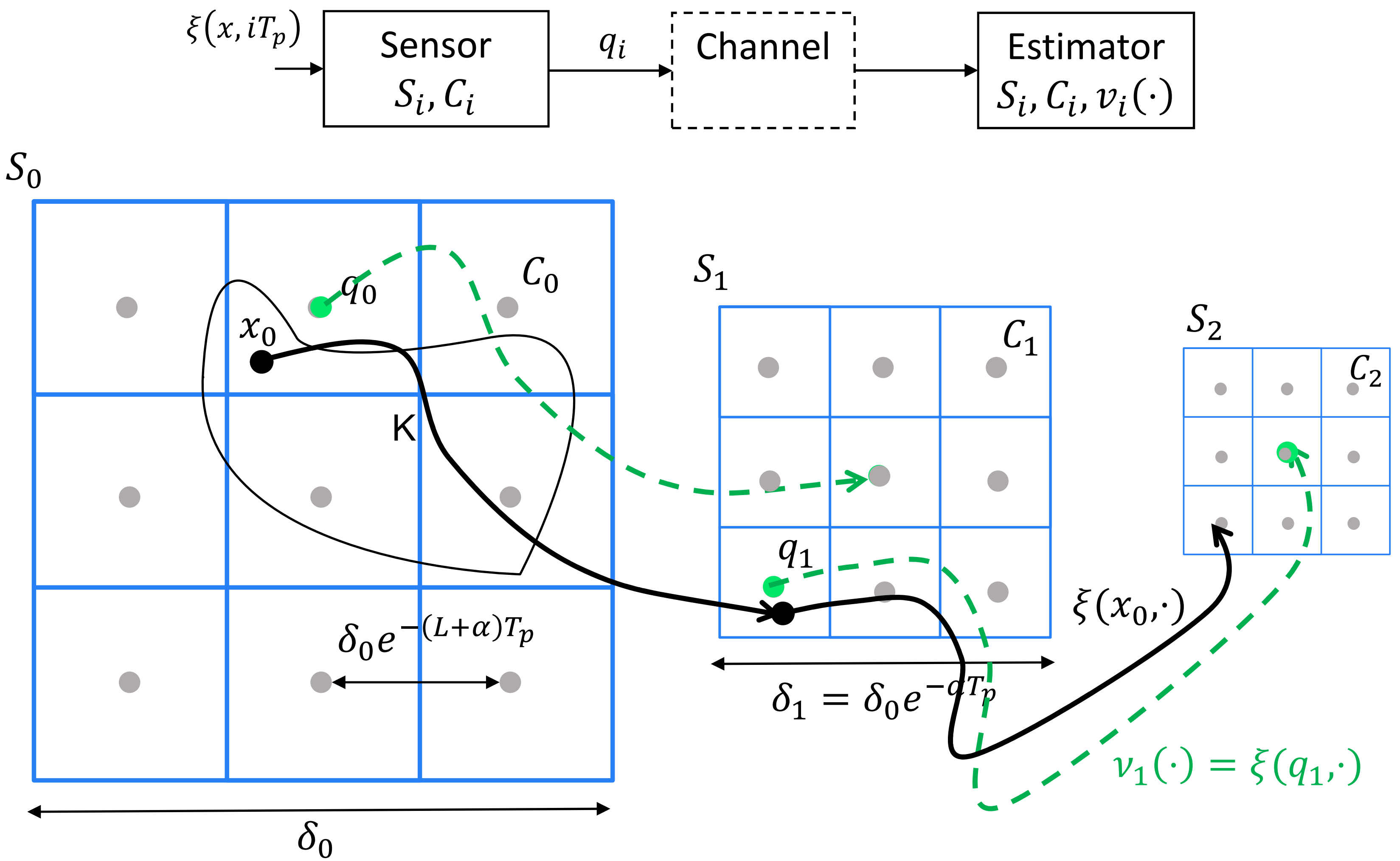
For linear system $\dot{x} = Ax$ this result can be refined to

$$\boxed{h_{\text{est}} = \sum_{i=1}^{n} \max\{\operatorname{Re} \lambda_i(A) + \alpha, 0\}}$$

Lower bound is obtained by propagating $\operatorname{vol}(K)$ along flow (Liouville's trace formula) and counting # of spanning $\varepsilon-$balls needed to cover this volume (cf. [Savkin] or [Schmidt '16])

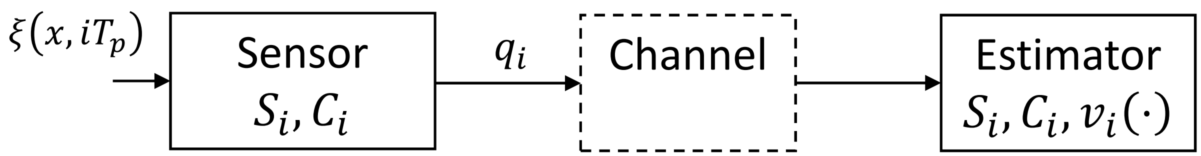
Similar argument gives lower bound for nonlinear system (cf. [Colonius]): $\boxed{h_{\text{est}} \geq \inf_x \operatorname{tr} \partial f / \partial x(x) + \alpha n}$

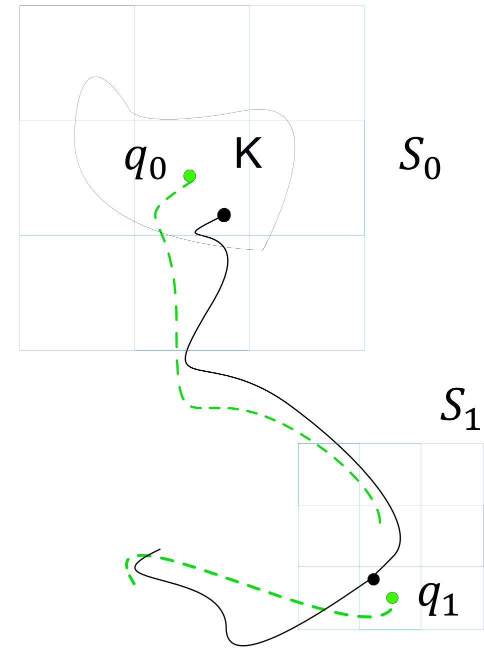
ESTIMATION PROCEDURE



Properties: $\xi(x, iT_p) \in S_i \; \forall i$ and $\|\xi(x,t) - v(t)\|_\infty \leq \delta_0 e^{-\alpha t} \; \forall t$

BIT RATE





The only information to send is q_i

Size of S_i is δ_i, resolution of C_i is $e^{-(L+\alpha)T_p}\delta_i$

hence $\#C_i$ = $e^{(L+\alpha)T_p n}$ (for $i > 0$)

Number of bits sent for q_i is $\log(\#C_i) = (L+\alpha)T_p n$

Long-term average bit rate: $(L+\alpha)n$

Matches entropy upper bound $h_{\text{est}} \leq (L+\alpha)n$

EFFICIENCY GAP

Question: Can the same estimation task be done with $< h_{\text{est}}$ bits?

Answer: No

Proof idea: Suppose another estimator uses $< h_{\text{est}}$ bit rate

Recall: $h_{\text{est}} = \sup_{\varepsilon>0} \limsup_{T\to\infty} \frac{1}{T} \log n^*_{\text{est}}(T, 2\varepsilon)$

Then for ε small enough & ℓ large enough, # of possible codeword sequences over ℓ rounds $<$ size of largest $(\ell T_p, 2\varepsilon)$-separated set

$\Rightarrow$ two initial states from $(\ell T_p, 2\varepsilon)$-separated set generate same codeword $\Rightarrow$ they're within $\varepsilon e^{-\alpha t}$ of same estimate – contradiction

Efficiency gap of our algorithm is at most as large as the gap between the entropy and its upper bound

Procedure in [Savkin] operates at arbitrary bit rate $> h_{\text{est}}$, but does block coding using sequences from suitable spanning set – not as constructive

MODEL DETECTION PROBLEM

Want to distinguish between two competing models

$$\dot{x} = f_i(x),\ i \in \{1,2\},\ x \in \mathbb{R}^n,\ x(0) \in K$$

using finite-data-rate state measurements (as above)

Need solutions of two models to be "sufficiently different"

$\xi_i(x,t)$ – solution of model i from x after time t

L_i – Lipschitz constant of f_i (can use matrix measure instead)

Call models (L,T)-separated if $\exists \varepsilon_{\min} > 0$ s.t. $\forall \varepsilon \le \varepsilon_{\min}$:

$$|x_1 - x_2| \le \varepsilon \ \Rightarrow\ |\xi_1(x_1,T) - \xi_2(x_2,T)| > \varepsilon e^{LT}$$

Sufficient condition: exponential separation holds over a compact set of states D if $f_1(x) \neq f_2(x)\ \ \forall x \in D$ ("generically true")

MODEL DETECTION ALGORITHM PROBLEM

If $\xi(x_0, t) \notin S_{i-1}$ output "2"; break

Else q_i:= Quantized measurement of $\xi\left(x_0, iT_p\right)$ w.r.t C_{i-1}

$v(t)$:= $\xi_1\left(q_i, t - (i-1)T_p\right)$ for $t \in [(i-1)T_p, iT_p]$

$\delta_i \coloneqq e^{-\alpha T_p}\delta_{i-1}$

$S_i \coloneqq$ hypercube with center $v(iT_p)$ and radius δ_i

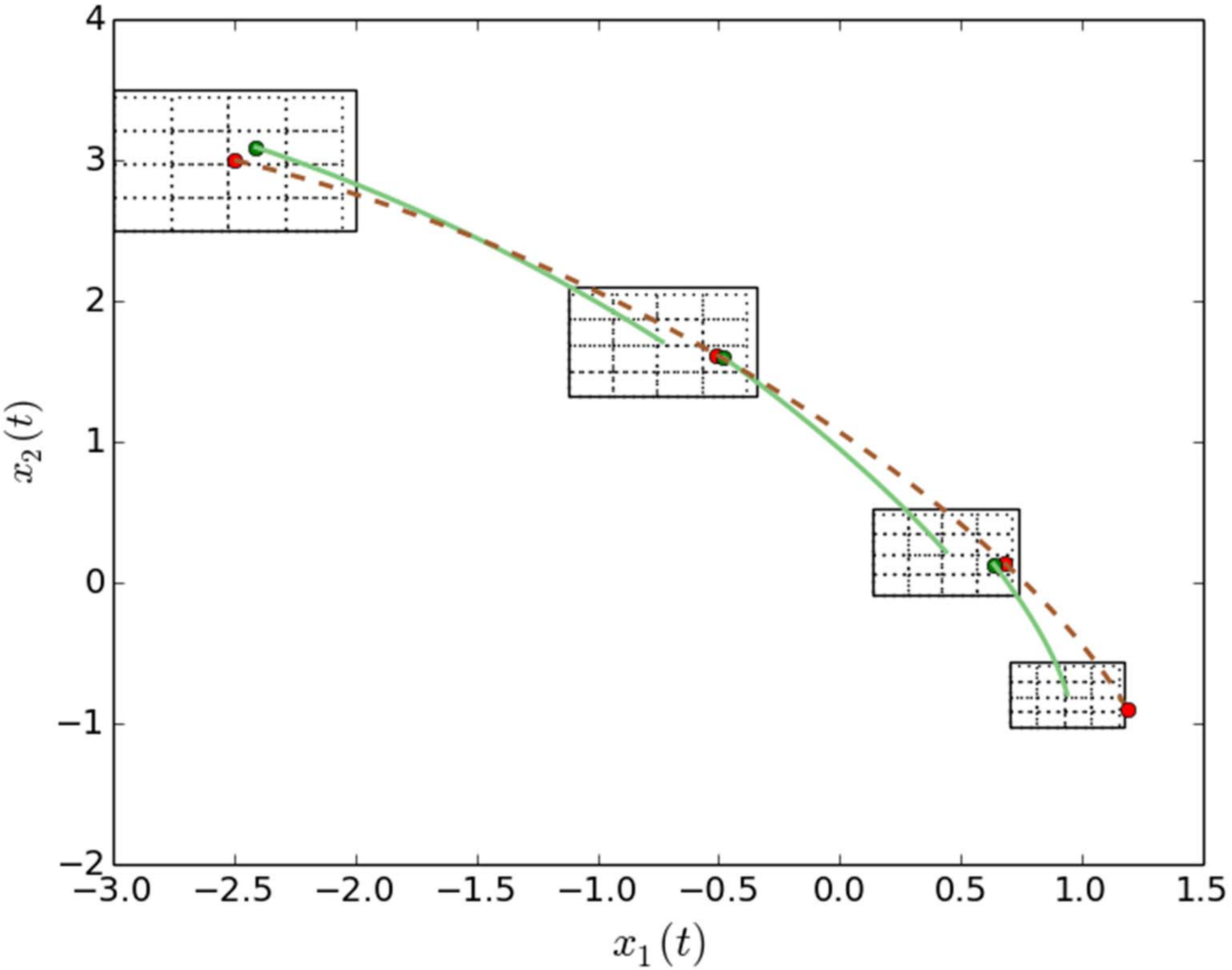
$C_i \coloneqq$ grid S_i with size $e^{-(L_1+\alpha)T_p}\delta_i$

Theorem: Under (L_1, T_p) separation, output "2" iff true model is f_2

If the true model is f_1: by correctness of estimation, actual state always stays in S_i, no output

If the true model is f_2: since δ_i decays geometrically, it will eventually become smaller than $\varepsilon_{\min}$. By exponential separation, at next iteration the actual state will exit S_i

MODEL DETECTION ALGORITHM



http://mitras.ece.illinois.edu/Detectdoc/

CONCLUSIONS

Contributions:

- Two equivalent notions of estimation entropy $h_{est} = h^*_{est}$
- Upper bound $h_{est} \leq (L + \alpha)n$ and lower bound
- State estimation and model detection algorithms with bit rate matching the upper bound on entropy

Future work:

- Better conditions for separation between two models
- Improved lower bounds on entropy
- Entropy for switched/hybrid systems and applications to finite-data-rate control [Schmidt '16, Yang–L]