ESTIMATION ENTROPY for NONLINEAR and SWITCHED SYSTEMS

Daniel Liberzon

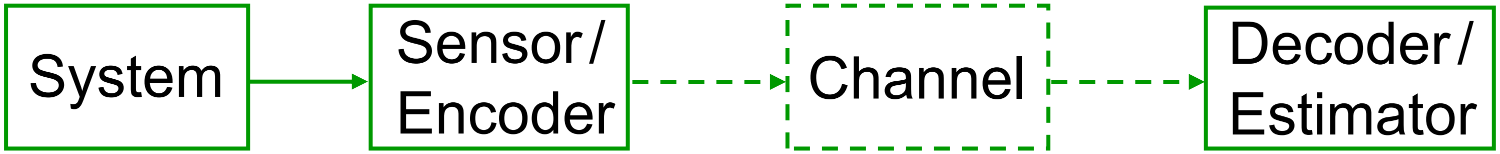
Joint work with Sayan Mitra and Guosong Yang (now at UCSB)

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

Online Seminar on Interaction of Control and Information, Apr 19, 2021

MOTIVATION and PROBLEM FORMULATION

MOTIVATION and PROBLEM FORMULATION



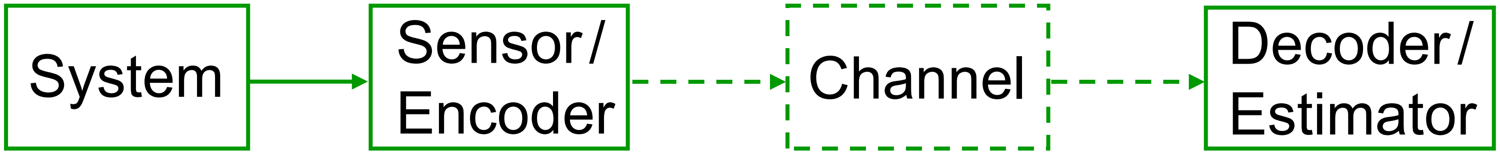
System

Sensor/ Encoder

Channel

Decoder/ Estimator

MOTIVATION and PROBLEM FORMULATION



System

Sensor/ Encoder

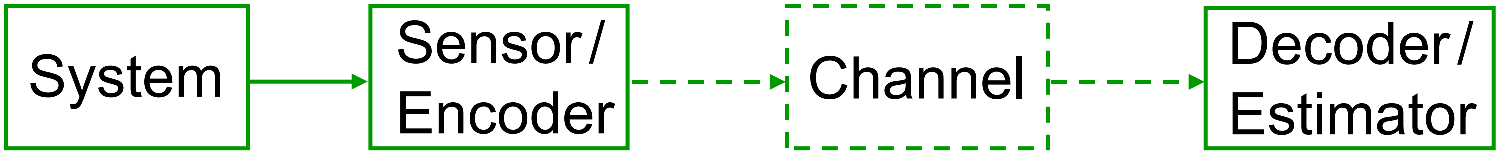
Channel

Decoder/ Estimator

State estimation:

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

MOTIVATION and PROBLEM FORMULATION


System

Sensor/ Encoder

Channel

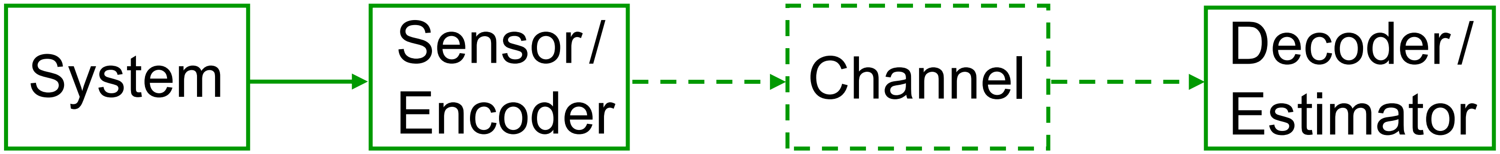
Decoder/ Estimator

State estimation:

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

State estimates can then be used for feedback control

MOTIVATION and PROBLEM FORMULATION



State estimation:

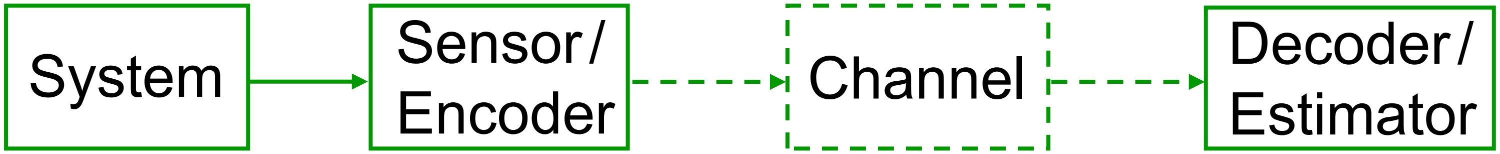
How much data rate is needed to estimate system state with error converging to 0 at desired exponential rate?

State estimates can then be used for feedback control

Model detection:

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

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?

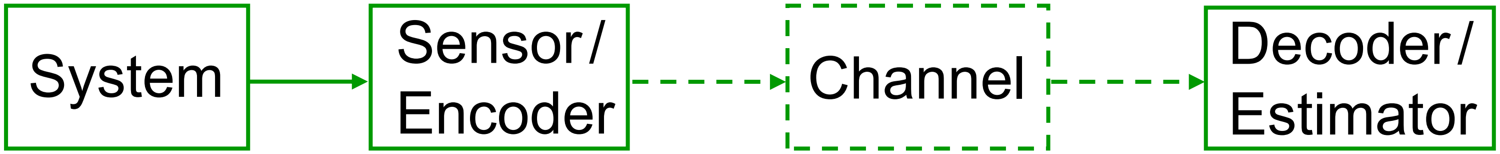
State estimates can then be used for feedback control

Model detection:

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

(Can apply state estimation scheme if trajectories of different models are sufficiently different)

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?

State estimates can then be used for feedback control

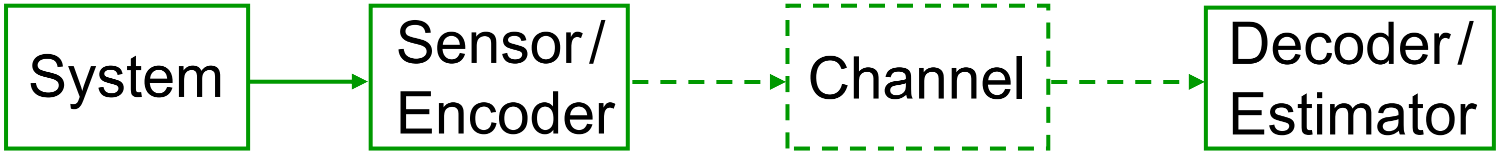
Model detection:

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

(Can apply state estimation scheme if trajectories of different models are sufficiently different)

Desired data rate is described by estimation entropy [L–Mitra ’18]

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?

State estimates can then be used for feedback control

Model detection:

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

(Can apply state estimation scheme if trajectories of different models are sufficiently different)

Desired data rate is described by estimation entropy [L–Mitra ’18]

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

ENTROPY DEFINITION

ENTROPY DEFINITION

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

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

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

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

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

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

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

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

$\alpha \geq 0$ – 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 initial states $x_1, \dots, x_N \in K$
(T, ε)-spanning if $\forall x \in K \; \exists \, x_i$:

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

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

$\alpha \geq 0$ – 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 initial states $x_1, \dots, x_N \in K$
(T, ε)-spanning if $\forall x \in K \ \exists\, x_i$:

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

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

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

$\alpha \geq 0$ – 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 initial states $x_1, \ldots, x_N \in K$
(T,ε)-spanning if $\forall x \in K \; \exists \, x_i$:

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

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

Estimation entropy:

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

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

$\alpha \geq 0$ – 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 initial states $x_1, \dots, x_N \in K$
(T,ε)-spanning if $\forall x \in K \; \exists \, x_i$:

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

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

Estimation entropy:

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

Intuition: $x_1, \dots, x_N$ are quantization points, $h_{\text{est}} =$ bit rate

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

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

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

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

Equivalent definition via (T, ε)-separated sets:

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

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

Equivalent definition via (T,ε)-separated sets:

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

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

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

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

Equivalent definition via (T,ε)-separated sets:

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

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

$n_{\text{est}}(T,\varepsilon) :=$ cardinality M of largest (T,ε)-separated set

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

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

Equivalent definition via (T,ε)-separated sets:

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

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

$n_{\text{est}}(T,\varepsilon) :=$ cardinality M of largest (T,ε)-separated set

Can define entropy by $h_{\text{est}} := \lim_{\varepsilon\to 0} \limsup_{T\to\infty} \frac{1}{T} \log n_{\text{est}}(T,\varepsilon)$

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

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

Equivalent definition via (T,ε)-separated sets:

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

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

$n_{\text{est}}(T,\varepsilon) :=$ cardinality M of largest (T,ε)-separated set

Can define entropy by $h_{\text{est}} := \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log n_{\text{est}}(T,\varepsilon)$

Can show (using standard techniques):

$$n_{\text{est}}(T,2\varepsilon) \leq s_{\text{est}}(T,\varepsilon) \leq n_{\text{est}}(T,\varepsilon)$$

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

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

Equivalent definition via (T, ε)-separated sets:

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

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

$n_{\text{est}}(T, \varepsilon) :=$ cardinality M of largest (T, ε)-separated set

Can define entropy by $h_{\text{est}} := \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log n_{\text{est}}(T, \varepsilon)$

Can show (using standard techniques):

$$n_{\text{est}}(T, 2\varepsilon) \leq s_{\text{est}}(T, \varepsilon) \leq n_{\text{est}}(T, \varepsilon)$$

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

TOY EXAMPLE

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \ x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$: • divide K into N (equal) subintervals with centers x_i

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \ x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period ↘

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T}\log N$ per time unit

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T} \log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell} e^{\ell a T}$

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \ x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T} \log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell} e^{\ell a T}$

Hence to estimate $x(t)$ with error converging to 0 as $e^{-\alpha t}$ we need data rate of $\boxed{a + \alpha \text{ bits}}$ (or nats)

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \ x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period ↘

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T}\log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell} e^{\ell a T}$

Hence to estimate $x(t)$ with error converging to 0 as $e^{-\alpha t}$
we need data rate of $\boxed{a + \alpha \text{ bits}}$ (or nats)

Entropy: the set $C := \{x_1, ..., x_N\}$ is (T, ε)-spanning if

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T}\log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell} e^{\ell a T}$

Hence to estimate $x(t)$ with error converging to 0 as $e^{-\alpha t}$ we need data rate of $\boxed{a + \alpha \text{ bits}}$ (or nats)

Entropy: the set $C := \{x_1, ..., x_N\}$ is (T, ε)-spanning if $|x_i - x_j| < \varepsilon e^{-(a+\alpha)T}$

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \ x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T}\log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell}e^{\ell aT}$

Hence to estimate $x(t)$ with error converging to 0 as $e^{-\alpha t}$ we need data rate of $\boxed{a + \alpha \text{ bits}}$ (or nats)

Entropy: the set $C := \{x_1, ..., x_N\}$ is (T, ε)-spanning if

$$|x_i - x_j| < \varepsilon e^{-(a+\alpha)T} \Rightarrow \#C = e^{(a+\alpha)T}|K|/\varepsilon$$

TOY EXAMPLE

$\dot{x} = ax, \quad a > 0, \; x(0) \in K \subset \mathbb{R}$ – known compact interval

Goal: estimate $x(t)$ using finite-data-rate encoding of x-values

$t = 0$:
- divide K into N (equal) subintervals with centers x_i
- send the index of the interval containing $x(0)$

sampling period

$t = T$:
- divide reachable set again into N (equal) subintervals
- repeat

This coding scheme uses data rate $\frac{1}{T}\log N$ per time unit

At $t = \ell T$, we know $x(t)$ is in an interval of length $\frac{|K|}{N^\ell} e^{\ell a T}$

Hence to estimate $x(t)$ with error converging to 0 as $e^{-\alpha t}$ we need data rate of $\boxed{a + \alpha \text{ bits}}$ (or nats)

Entropy: the set $C := \{x_1, ..., x_N\}$ is (T, ε)-spanning if

$|x_i - x_j| < \varepsilon e^{-(a+\alpha)T} \Rightarrow \#C = e^{(a+\alpha)T}|K|/\varepsilon$

Take $\limsup_{T\to\infty} \frac{1}{T}\log$ of this to get $\boxed{h_{\text{est}} = a + \alpha}$

BOUNDS on ENTROPY: PRELIMINARIES

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds}|x_0|$$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds}|x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \le e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \le \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \neq i} |a_{ij}|\}$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \leq \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \neq i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \leq \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \neq i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \;\; x(0) \in K, \; f \in C^1$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \leq \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \neq i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \quad x(0) \in K, \ f \in C^1$
the above LTV bound applied to its variational equation gives
$|\xi(x_1, t) - \xi(x_2, t)| \leq e^{Lt} |x_1 - x_2|$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \leq \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j\neq i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \quad x(0) \in K, \; f \in C^1$ the above LTV bound applied to its variational equation gives $|\xi(x_1, t) - \xi(x_2, t)| \leq e^{Lt} |x_1 - x_2|$ where $L := \sup \mu\big(\partial f / \partial x(x)\big)$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \leq e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \leq \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \neq i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \;\; x(0) \in K, \; f \in C^1$
the above LTV bound applied to its variational equation gives
$|\xi(x_1, t) - \xi(x_2, t)| \leq e^{Lt} |x_1 - x_2|$ where $L := \sup \mu\big(\partial f / \partial x(x)\big)$
and the sup is taken over all x reachable from $\mathrm{conv}(K)$

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \le e^{\int_{t_0}^{t} \mu(A(s))ds} |x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \le \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i \{a_{ii} + \sum_{j \ne i} |a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \;\; x(0) \in K, \; f \in C^1$
the above LTV bound applied to its variational equation gives
$|\xi(x_1,t) - \xi(x_2,t)| \le e^{Lt}|x_1 - x_2|$ where $L := \sup \mu\big(\partial f/\partial x(x)\big)$
and the sup is taken over all x reachable from $\mathrm{conv}(K)$

This provides a basis for constructing spanning sets (grids)

BOUNDS on ENTROPY: PRELIMINARIES

For LTV system $\dot{x} = A(t)x$ it is well known that

$|x(t)| \le e^{\int_{t_0}^{t} \mu(A(s))ds}|x_0|$

where $\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\|-1}{\varepsilon}$ is matrix measure

$\mu(A) \le \|A\|$; e.g., for ∞-norm, $\mu(A) = \max_i\{a_{ii} + \sum_{j\neq i}|a_{ij}|\}$

Can have $\mu(A) < 0$ but in all bounds $\mu(A)$ means $\max\{\mu(A), 0\}$

For a general nonlinear system $\dot{x} = f(x), \quad x(0) \in K, \ f \in C^1$
the above LTV bound applied to its variational equation gives
$|\xi(x_1,t) - \xi(x_2,t)| \le e^{Lt}|x_1 - x_2|$ where $L := \sup \mu\big(\partial f/\partial x(x)\big)$
and the sup is taken over all x reachable from $\mathrm{conv}(K)$

This provides a basis for constructing spanning sets (grids)

Note: we can instead take L to be the Lipschitz constant of f, which is more conservative but works if $\partial f/\partial x$ does not exist

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f / \partial x(x)\big)$$

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

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$$

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

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

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$$

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

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

We know: $|\xi(x_1,t) - \xi(x_2,t)| \leq |x_1 - x_2|e^{Lt}$

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$

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

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

We know: $|\xi(x_1, t) - \xi(x_2, t)| \leq |x_1 - x_2|e^{Lt}$

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

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$$

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

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

We know: $|\xi(x_1, t) - \xi(x_2, t)| \leq |x_1 - x_2| e^{Lt}$

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

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

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f / \partial x(x)\big)$

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

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

We know: $|\xi(x_1, t) - \xi(x_2, t)| \leq |x_1 - x_2| e^{Lt}$

If $|x_1 - x_2| \leq \varepsilon e^{-(L+\alpha)T}$ then $|\xi(x_1, t) - \xi(x_2, t)|$
$\leq \varepsilon e^{-(L+\alpha)T} e^{Lt} \leq \varepsilon e^{-\alpha t}$ $\forall t \in [0, T]$ 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 $\left(\dfrac{\ell}{\varepsilon e^{-(L+\alpha)T}}\right)^n = \dfrac{\ell^n e^{(L+\alpha)Tn}}{\varepsilon^n}$

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$$

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

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

We know: $|\xi(x_1,t) - \xi(x_2,t)| \leq |x_1 - x_2| e^{Lt}$

If $|x_1 - x_2| \leq \varepsilon e^{-(L+\alpha)T}$ then $|\xi(x_1,t) - \xi(x_2,t)|$ $\leq \varepsilon e^{-(L+\alpha)T} e^{Lt} \leq \varepsilon e^{-\alpha t}$ $\forall t \in [0,T]$ 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 $(L+\alpha)n$ ■

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$

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

BOUNDS on ENTROPY

$$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$$

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

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

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

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$

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

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

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

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

BOUNDS on ENTROPY

$\dot{x} = f(x),\ x(0) \in K \subset \mathbb{R}^n,\ L := \sup \mu\big(\partial f/\partial x(x)\big)$

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

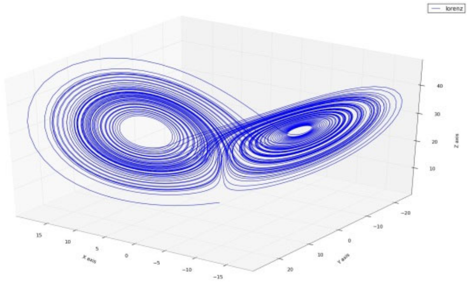
For linear system $\dot{x} = Ax$ this result can be refined to

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

Lower bound is obtained by propagating $\text{vol}(K)$ 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]): $\boxed{h_{\text{est}} \geq \inf_x \text{tr}\, \partial f/\partial x(x) + \alpha n}$

EXAMPLE: LORENZ SYSTEM



EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

can compute r s.t. $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall t \geq 0$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

can compute r s.t. $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

can compute r s.t. $x(t) \in B_r\big((0,0,\sigma+\theta)\big) \;\; \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Its matrix measure is $\mu(J(x)) = \max\limits_{i=1,2,3} \Big\{ J_{ii}(x) + \sum\limits_{j \neq i} |J_{ij}(x)| \Big\}$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

can compute r s.t. $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Its matrix measure is $\mu(J(x)) = \max\limits_{i=1,2,3} \Big\{ J_{ii}(x) + \sum\limits_{j \neq i} |J_{ij}(x)| \Big\}$

hence $L = \sup\limits_{x \in B_r} \mu(J(x)) = \max\big\{0, -1+\sigma+2r, -\beta+2r\big\}$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For initial set $K = B_{r_0}\big((0,0,0)\big)$

can compute r s.t. $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Its matrix measure is $\mu(J(x)) = \max\limits_{i=1,2,3} \Big\{ J_{ii}(x) + \sum\limits_{j \neq i} |J_{ij}(x)| \Big\}$

hence $L = \sup\limits_{x \in B_r} \mu(J(x)) = \max\big\{0, -1 + \sigma + 2r, -\beta + 2r\big\}$

and so $h_{\text{est}} \leq 3(L + \alpha)$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For $K = B_{r_0}\big((0,0,0)\big)$ we have $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Alternative viewpoint: interconnection of 3 scalar subsystems

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For $K = B_{r_0}\big((0,0,0)\big)$ we have $x(t) \in B_r\big((0,0,\sigma{+}\theta)\big)\ \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Alternative viewpoint: interconnection of 3 scalar subsystems

As in [Arcak–Maidens ’18], take matrix A s.t. $\forall\, x \in B_r$:

$\mu(J_{ii}(x)) \leq A_{ii}\,,\ \ \|J_{ij}(x)\| \leq A_{ij}$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For $K = B_{r_0}\big((0,0,0)\big)$ we have $x(t) \in B_r\big((0,0,\sigma+\theta)\big) \;\; \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Alternative viewpoint: interconnection of 3 scalar subsystems

As in [Arcak–Maidens ’18], take matrix A s.t. $\forall\, x \in B_r$:

$\mu(J_{ii}(x)) \leq A_{ii}\,, \;\; \|J_{ij}(x)\| \leq A_{ij}$

Can take this matrix to be $A = \begin{pmatrix} -\sigma & \sigma & 0 \\ \sigma + r & -1 & r \\ r & r & -\beta \end{pmatrix}$

EXAMPLE: LORENZ SYSTEM

$$\boxed{\begin{aligned}\dot{x}_1 &= \sigma x_2 - \sigma x_1\\ \dot{x}_2 &= \theta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\beta x_3 + x_1 x_2\end{aligned}}$$

For $K = B_{r_0}\big((0,0,0)\big)$ we have $x(t) \in B_r\big((0,0,\sigma+\theta)\big)\ \forall\, t \geq 0$

Jacobian is $J(x) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \theta - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$

Alternative viewpoint: interconnection of 3 scalar subsystems

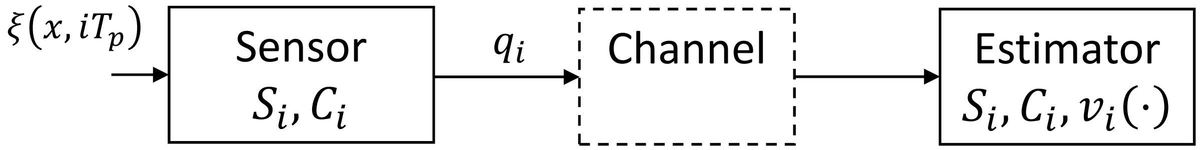
As in [Arcak–Maidens '18], take matrix A s.t. $\forall\, x \in B_r$:

$\mu(J_{ii}(x)) \leq A_{ii}\,,\ \ \|J_{ij}(x)\| \leq A_{ij}$

Can take this matrix to be $A = \begin{pmatrix} -\sigma & \sigma & 0 \\ \sigma + r & -1 & r \\ r & r & -\beta \end{pmatrix}$

Bound in [L '21] gives $h_{\text{est}} \leq 3(\max\{\lambda_{\max}(A), 0\} + \alpha)$

ESTIMATION PROCEDURE



$\xi(x, iT_p)$

Sensor

S_i, C_i

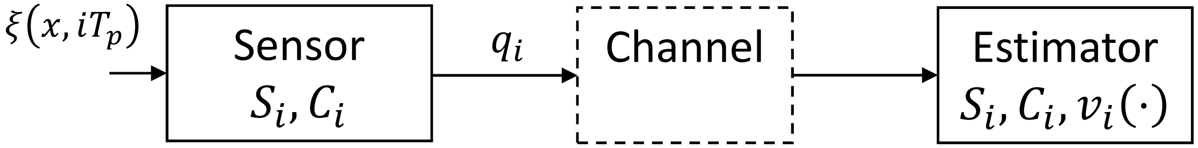
q_i

Channel

Estimator

$S_i, C_i, v_i(\cdot)$

ESTIMATION PROCEDURE



$\xi(x, iT_p)$

Sensor

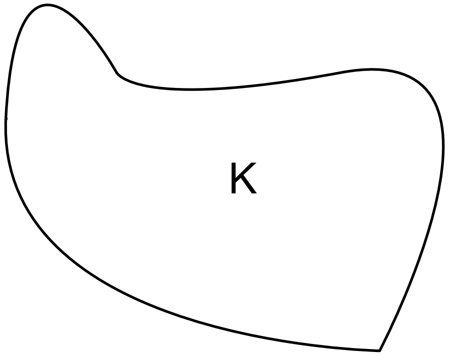
S_i, C_i

q_i

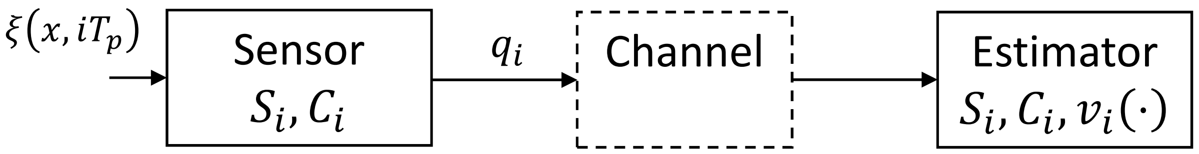
Channel

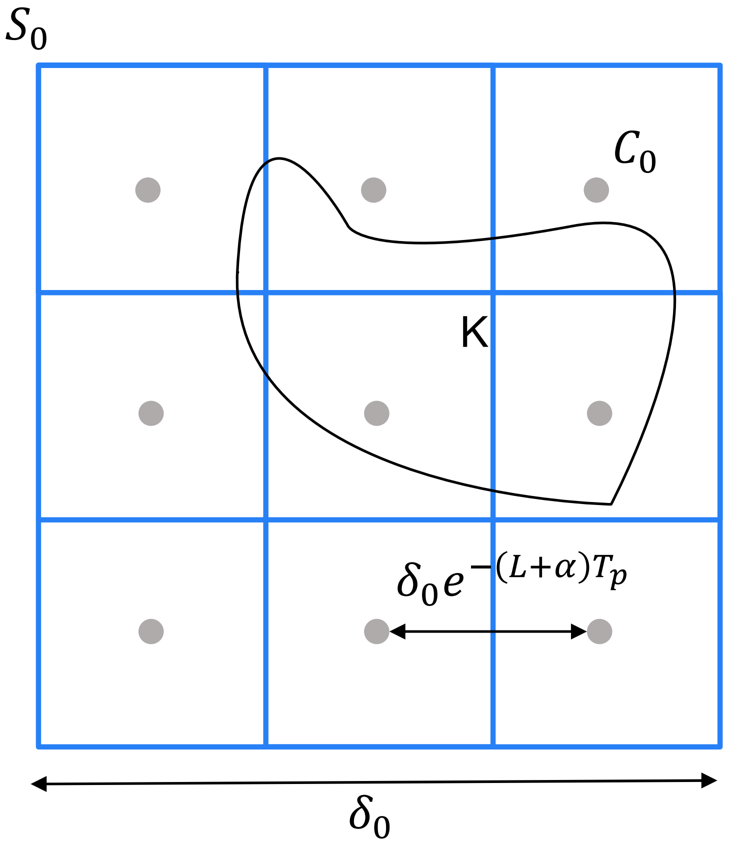
Estimator

$S_i, C_i, v_i(\cdot)$

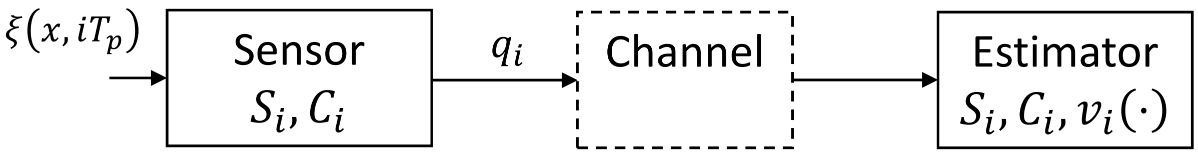


ESTIMATION PROCEDURE





ESTIMATION PROCEDURE



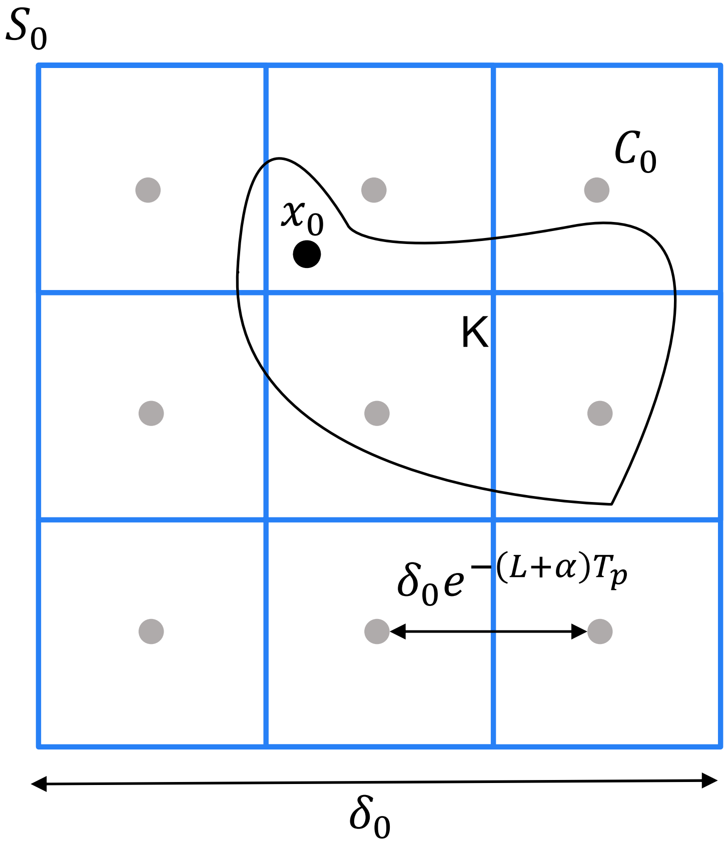
$\xi(x, iT_p)$

Sensor
S_i, C_i

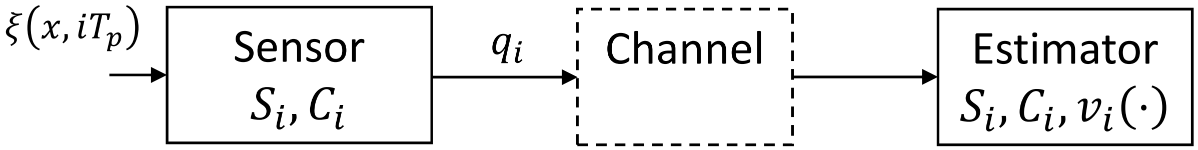
q_i

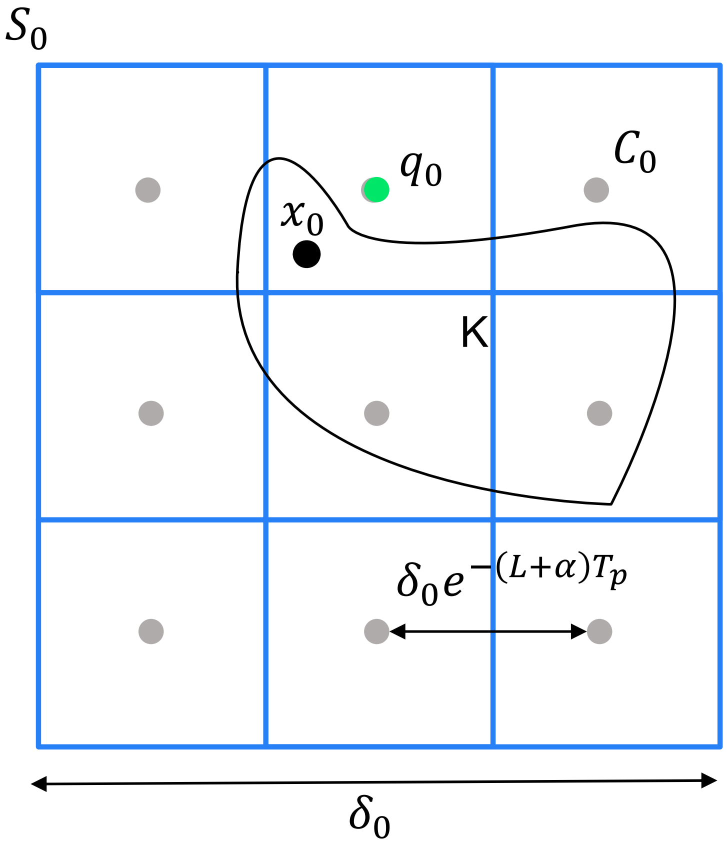
Channel

Estimator
$S_i, C_i, v_i(\cdot)$

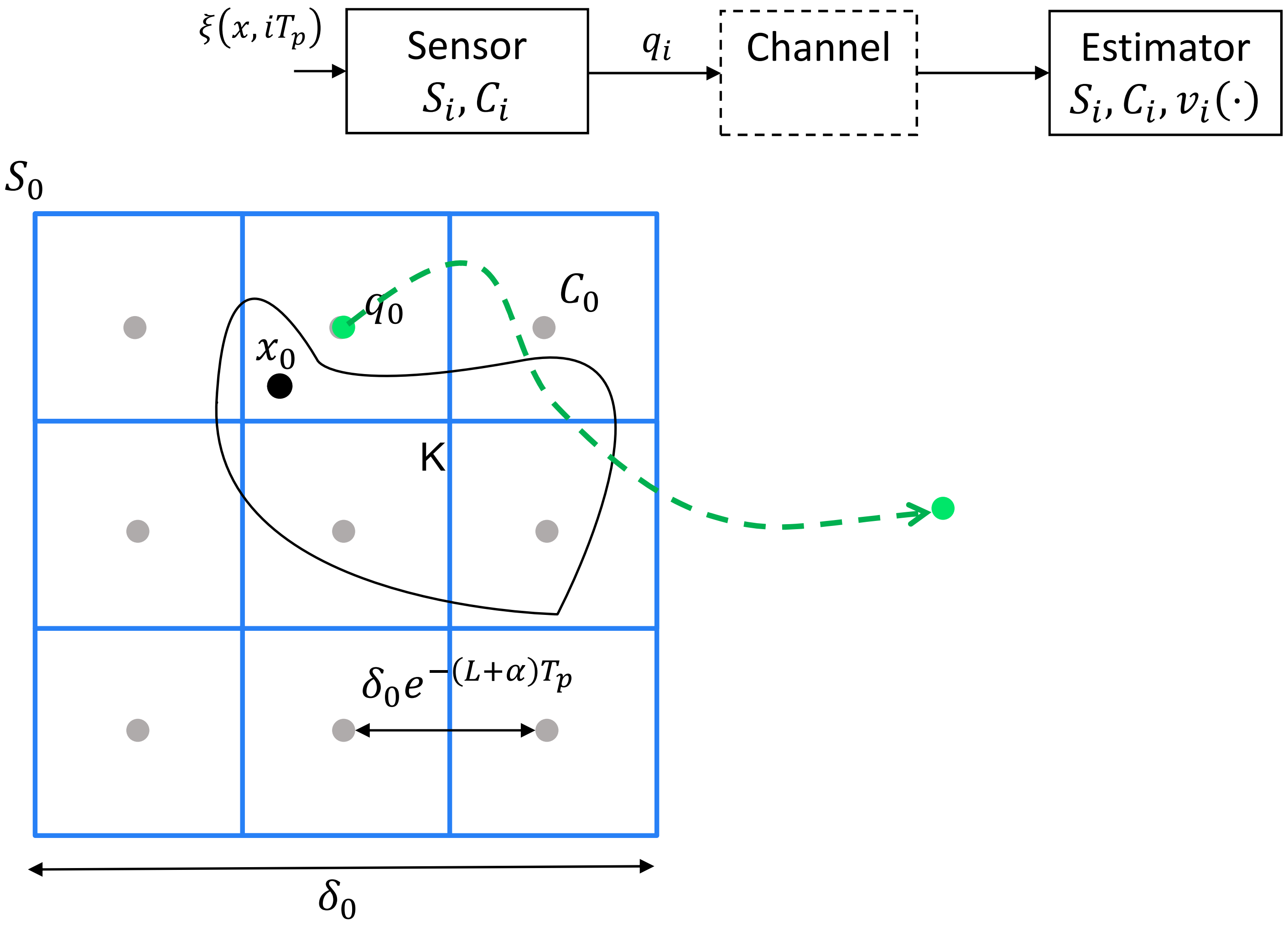


ESTIMATION PROCEDURE

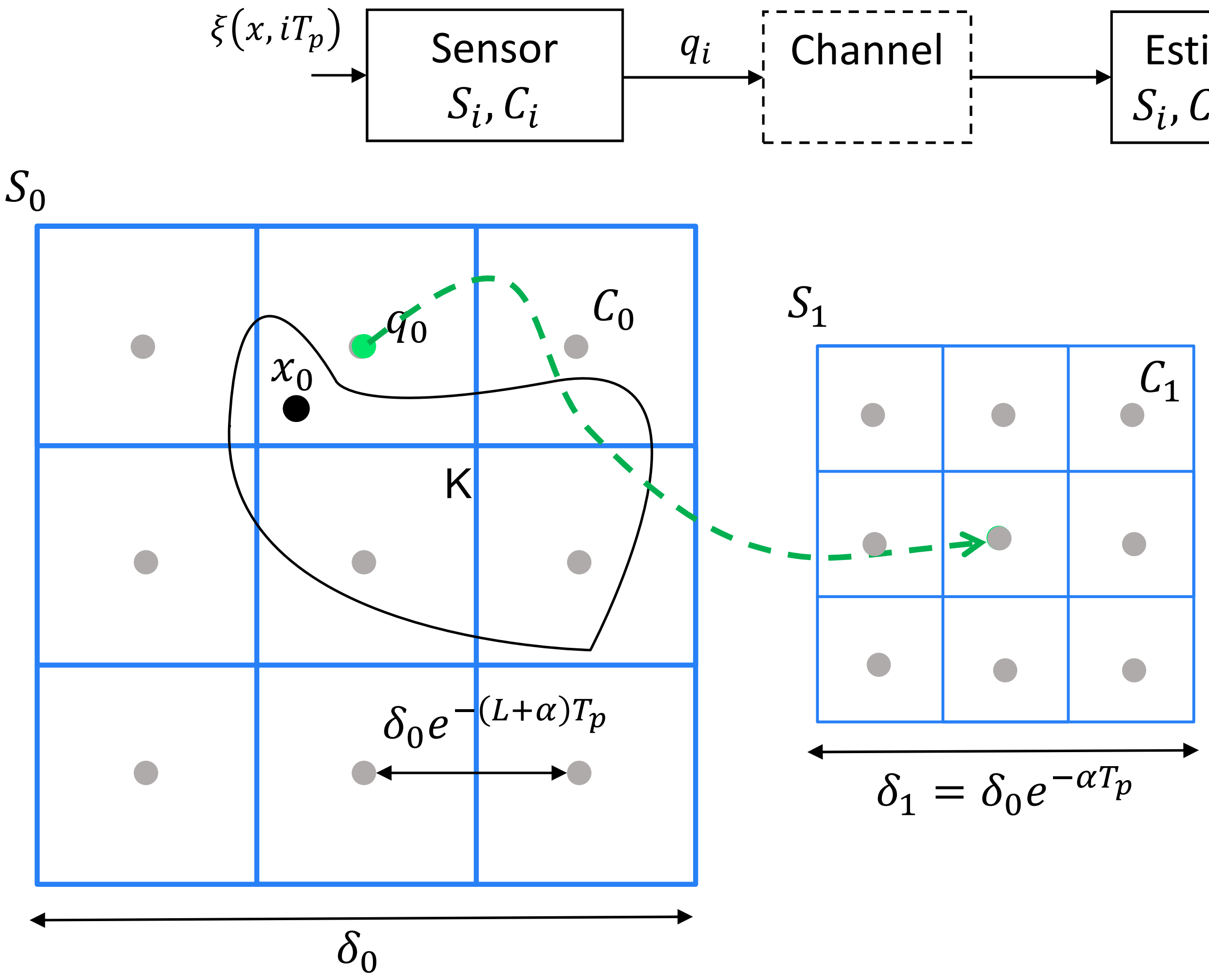




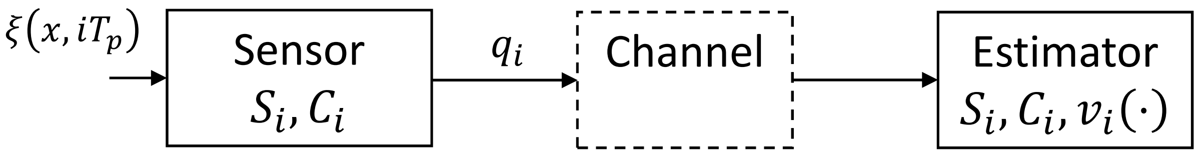
ESTIMATION PROCEDURE



ESTIMATION PROCEDURE



ESTIMATION PROCEDURE



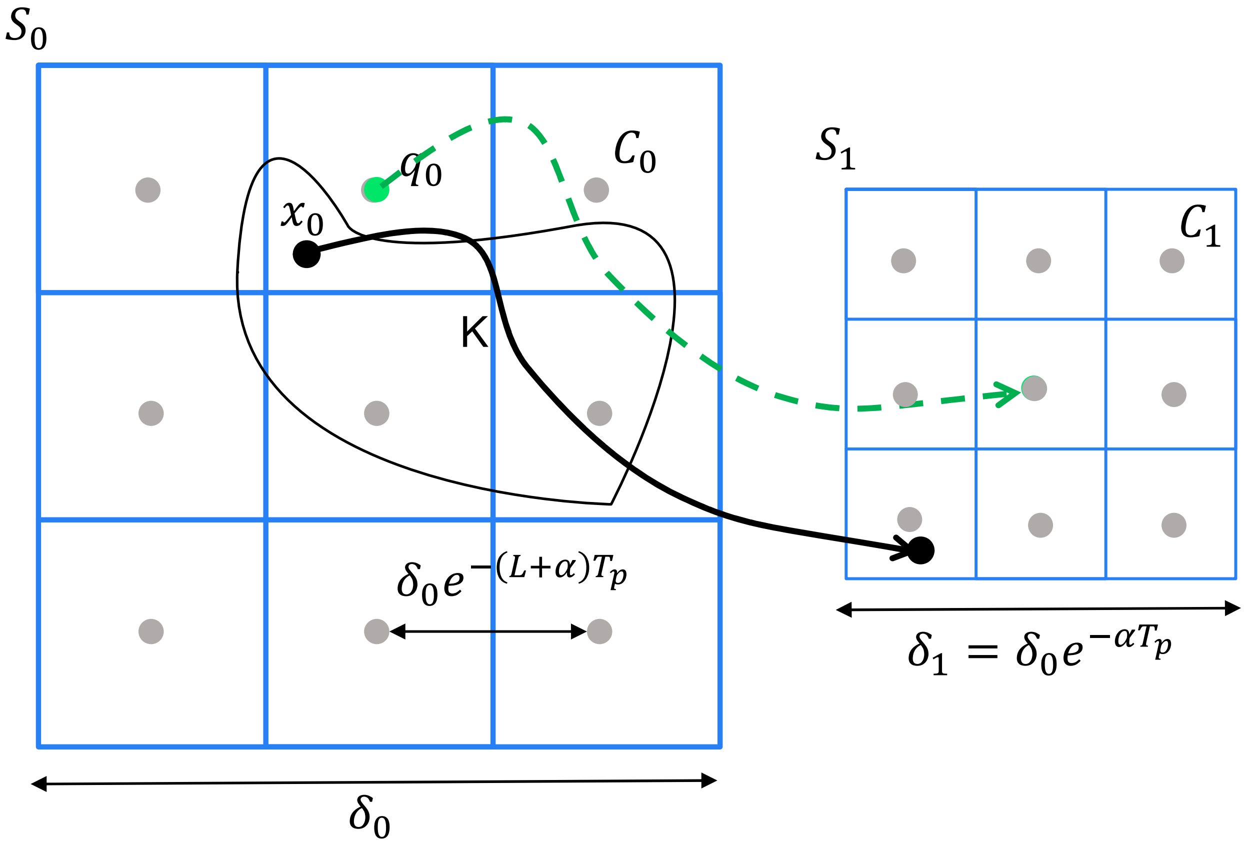
$\xi(x, iT_p)$

Sensor S_i, C_i

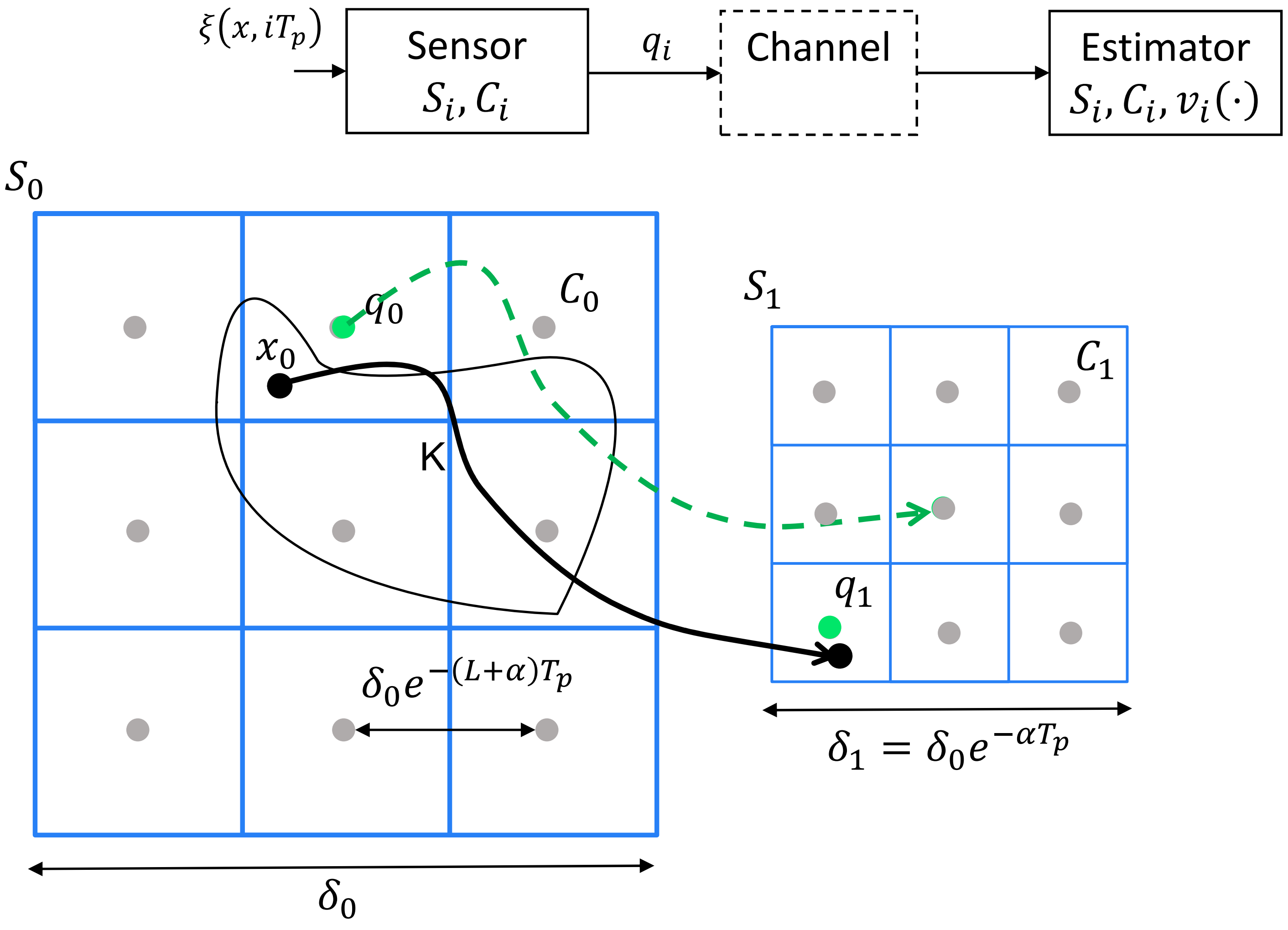
q_i

Channel

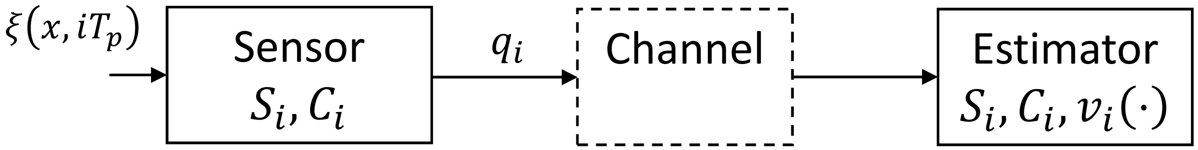
Estimator $S_i, C_i, v_i(\cdot)$



ESTIMATION PROCEDURE



ESTIMATION PROCEDURE



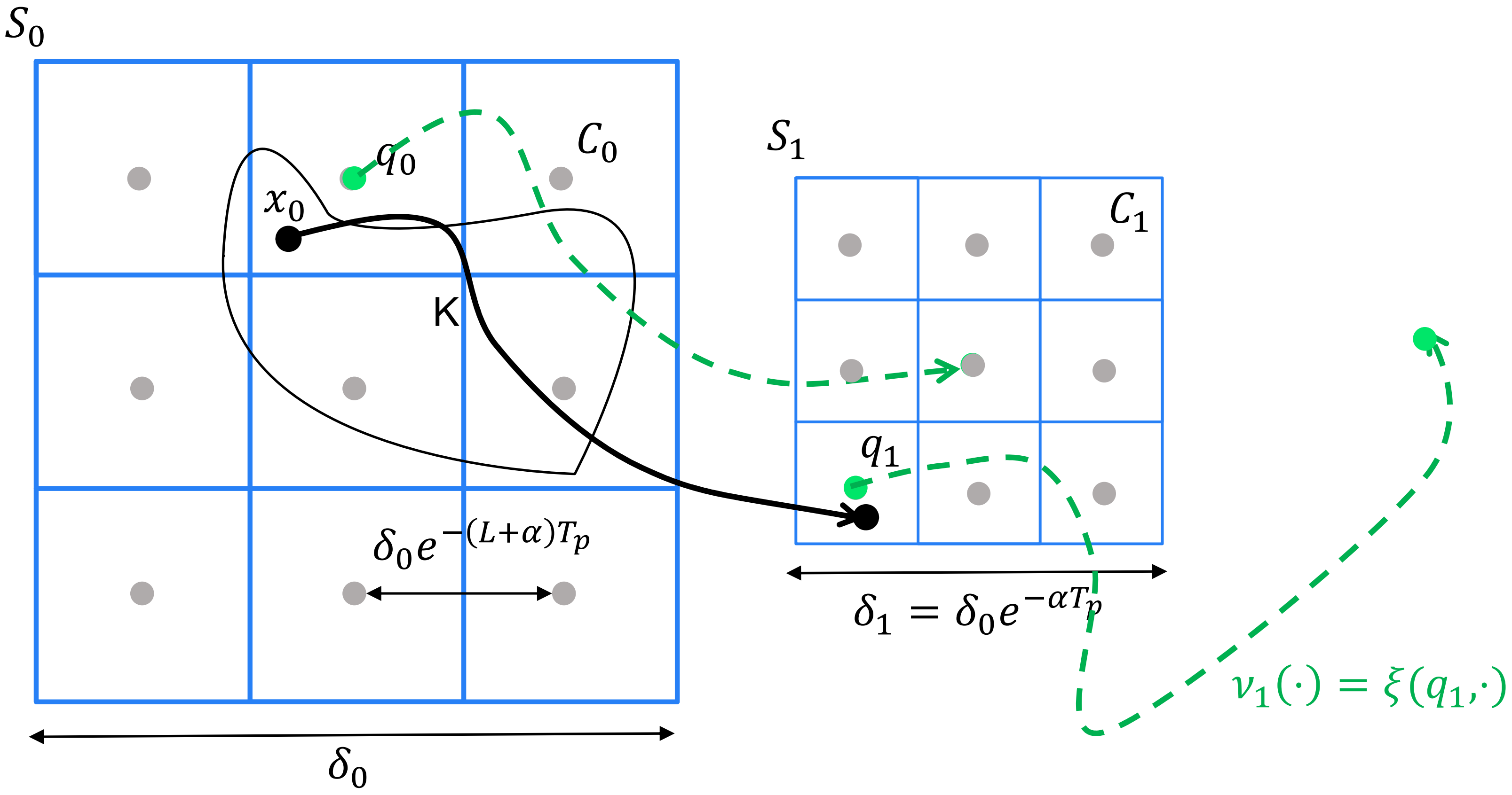
$\xi(x, iT_p)$

Sensor
S_i, C_i

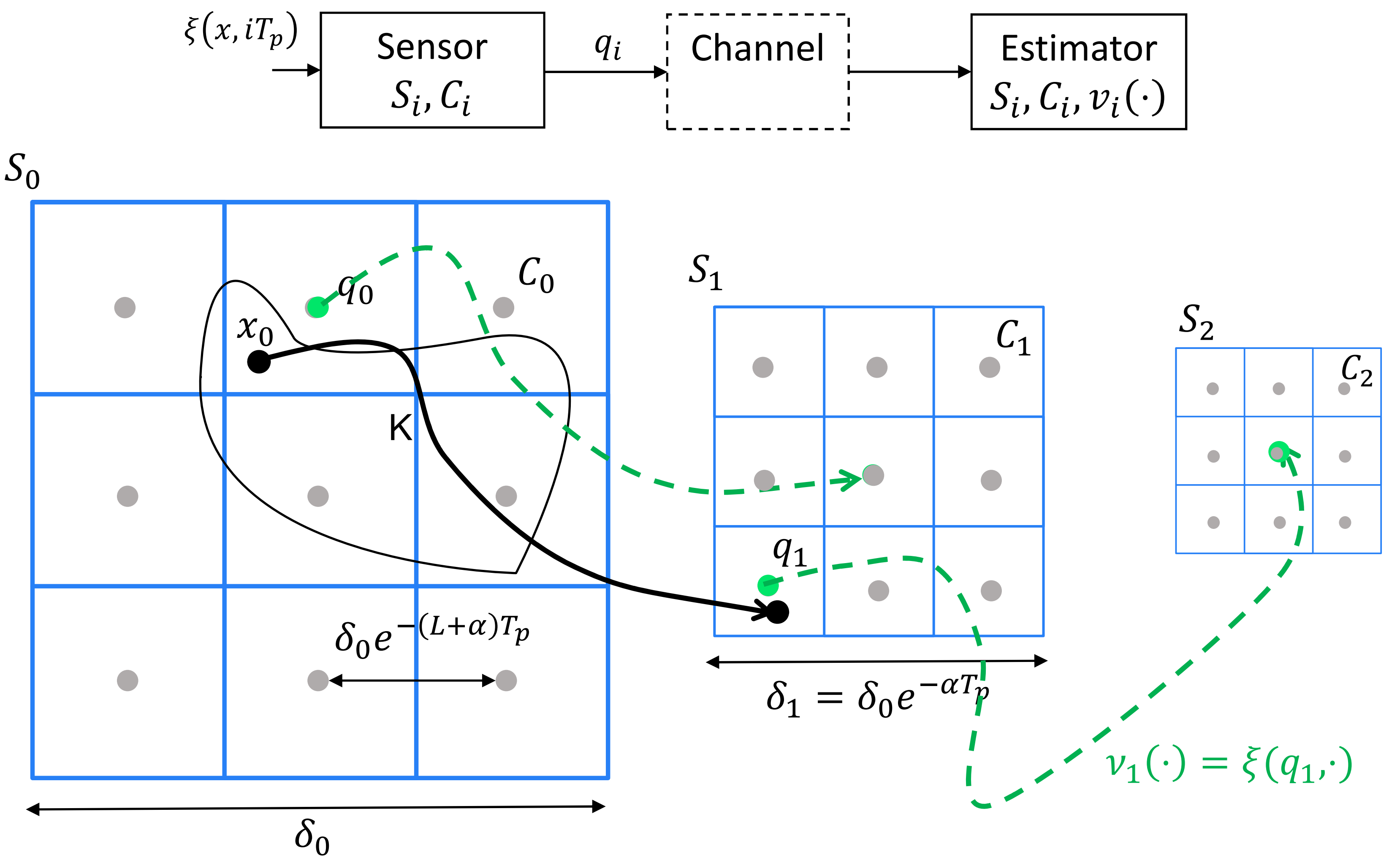
q_i

Channel

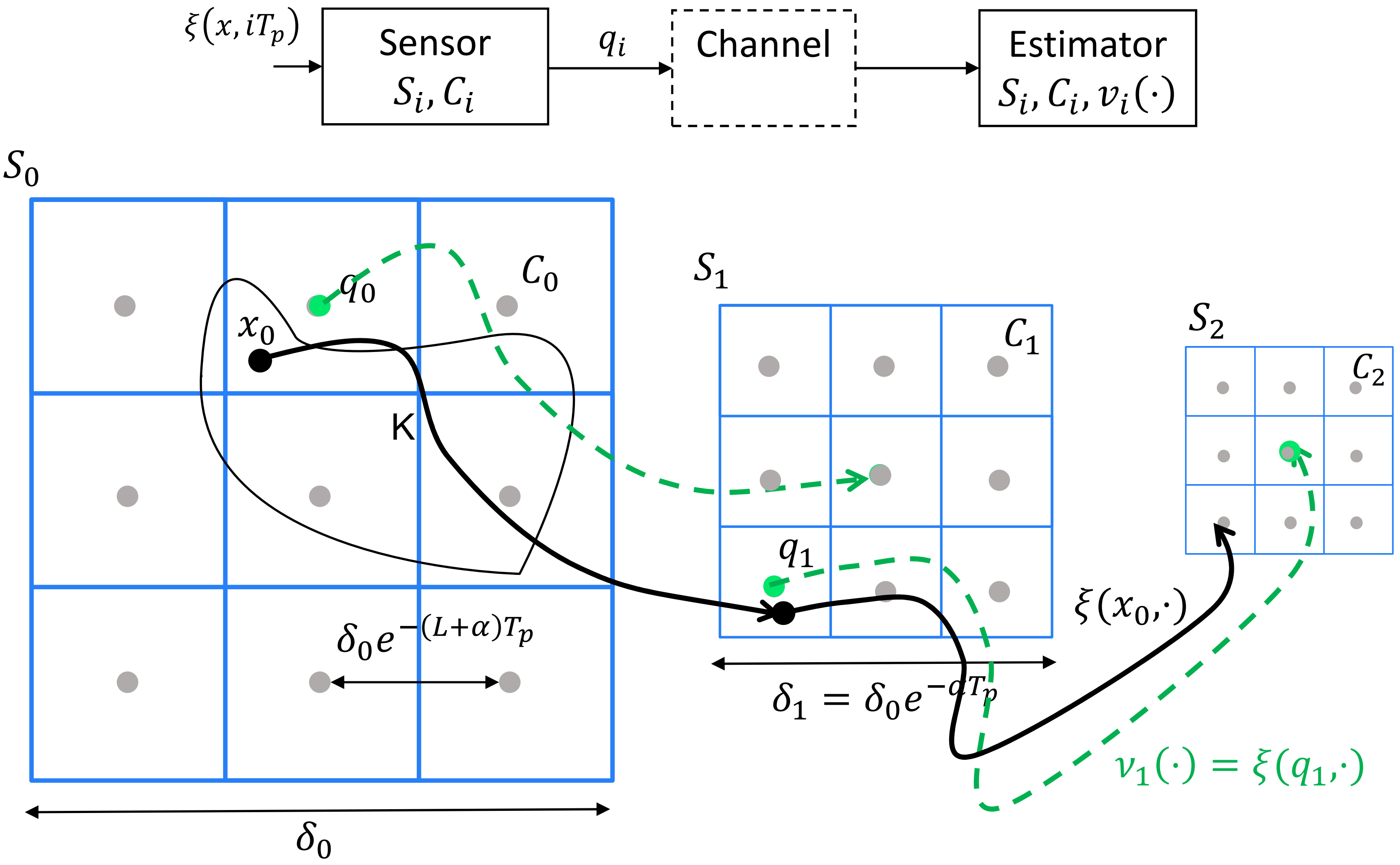
Estimator
$S_i, C_i, v_i(\cdot)$



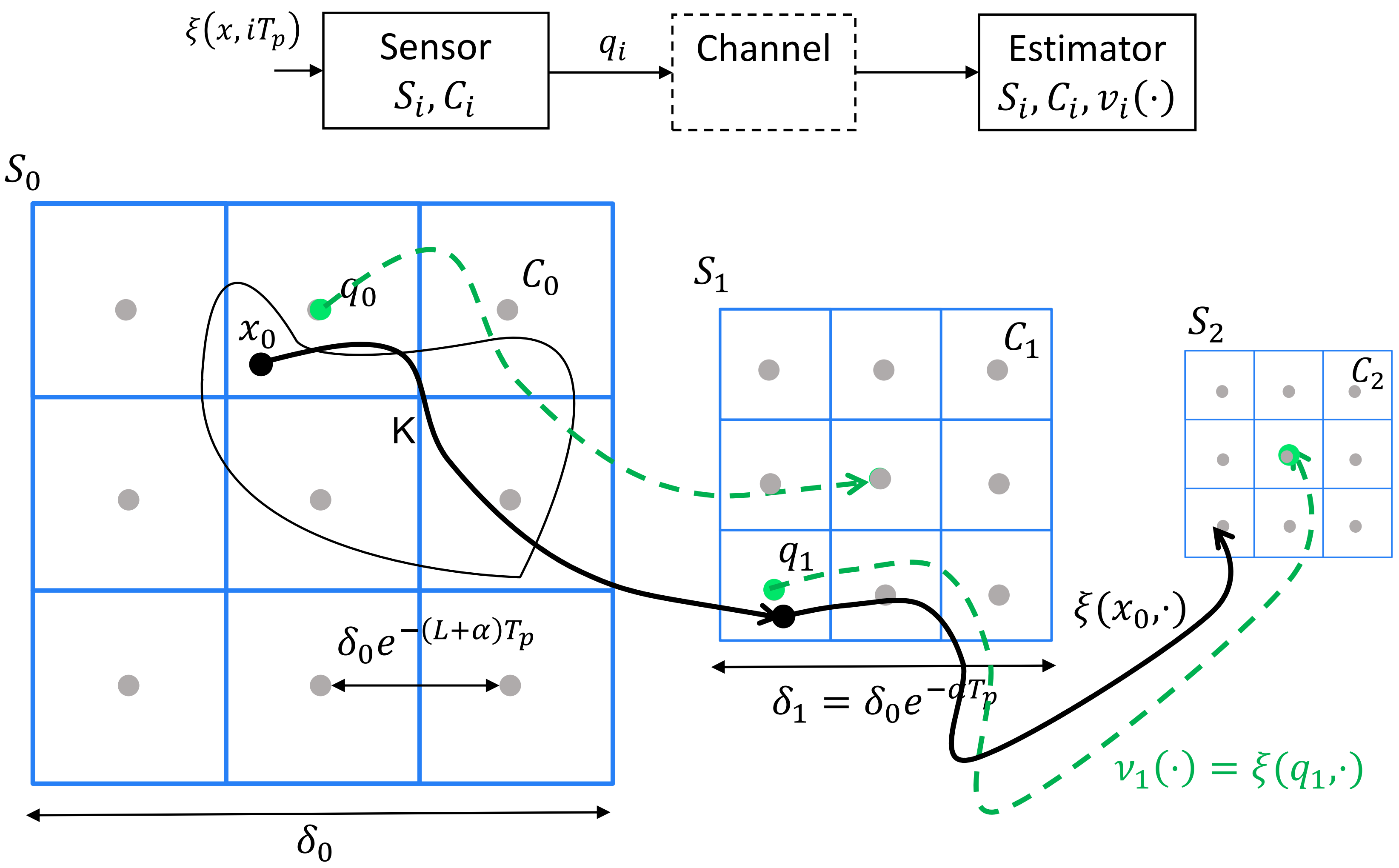
ESTIMATION PROCEDURE



ESTIMATION PROCEDURE



ESTIMATION PROCEDURE



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

DATA RATE and EFFICIENCY GAP

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L + \alpha)n$, upper bound on h_{est}

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L + \alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\mathrm{est}}$ bits

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L + \alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\mathrm{est}}$ bits

Proof idea: recall $h_{\text{est}} = \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log n_{\text{est}}(T, 2\varepsilon)$

where $n_{\text{est}}(T, 2\varepsilon) =$ cardinality of largest $(T, 2\varepsilon)$-separated set

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L+\alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\mathrm{est}}$ bits

Proof idea: recall $h_{\mathrm{est}} = \lim_{\varepsilon \to 0} \limsup_{T \to \infty} \frac{1}{T} \log n_{\mathrm{est}}(T, 2\varepsilon)$

where $n_{\mathrm{est}}(T, 2\varepsilon) =$ cardinality of largest $(T, 2\varepsilon)$-separated set

Two initial states from a $(T, 2\varepsilon)$-separated set cannot generate the same codeword (or they'd be within $\varepsilon e^{-\alpha t}$ of same estimate)

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L+\alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\mathrm{est}}$ bits

Proof idea: recall $h_{\mathsf{est}} = \lim_{\varepsilon\to 0} \limsup_{T\to\infty} \frac{1}{T} \log n_{\mathsf{est}}(T, 2\varepsilon)$

where $n_{\mathsf{est}}(T, 2\varepsilon) =$ cardinality of largest $(T, 2\varepsilon)$-separated set

Two initial states from a $(T, 2\varepsilon)$-separated set cannot generate the same codeword (or they'd be within $\varepsilon e^{-\alpha t}$ of same estimate)

This gives a lower bound on # of possible codewords in terms of cardinality of largest separated set, which leads to the claim

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L+\alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\text{est}}$ bits

Proof idea: recall $h_{\text{est}} = \lim_{\varepsilon\to 0} \limsup_{T\to\infty} \frac{1}{T} \log n_{\text{est}}(T, 2\varepsilon)$

where $n_{\text{est}}(T, 2\varepsilon) =$ cardinality of largest $(T, 2\varepsilon)$-separated set

Two initial states from a $(T, 2\varepsilon)$-separated set cannot generate the same codeword (or they'd be within $\varepsilon e^{-\alpha t}$ of same estimate)

This gives a lower bound on # of possible codewords in terms of cardinality of largest separated set, which leads to the claim

Efficiency gap of our algorithm = $(L+\alpha)n - h_{\text{est}}$

DATA RATE and EFFICIENCY GAP

Average bit rate of this algorithm is $(L+\alpha)n$, upper bound on h_{est}

Claim: the same estimation task cannot be done with $< h_{\mathrm{est}}$ bits

Proof idea: recall $h_{\text{est}} = \lim\limits_{\varepsilon\to 0} \limsup\limits_{T\to\infty} \frac{1}{T} \log n_{\text{est}}(T, 2\varepsilon)$

where $n_{\text{est}}(T, 2\varepsilon) =$ cardinality of largest $(T, 2\varepsilon)$-separated set

Two initial states from a $(T, 2\varepsilon)$-separated set cannot generate the same codeword (or they'd be within $\varepsilon e^{-\alpha t}$ of same estimate)

This gives a lower bound on # of possible codewords in terms of cardinality of largest separated set, which leads to the claim

Efficiency gap of our algorithm = $(L+\alpha)n - h_{\mathrm{est}}$

Procedure in [Savkin] operates at arbitrary bit rate > h_{est}, but does block coding using sequences from suitable spanning set – not 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. $\forall \varepsilon \leq \varepsilon_{\min}$:

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

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

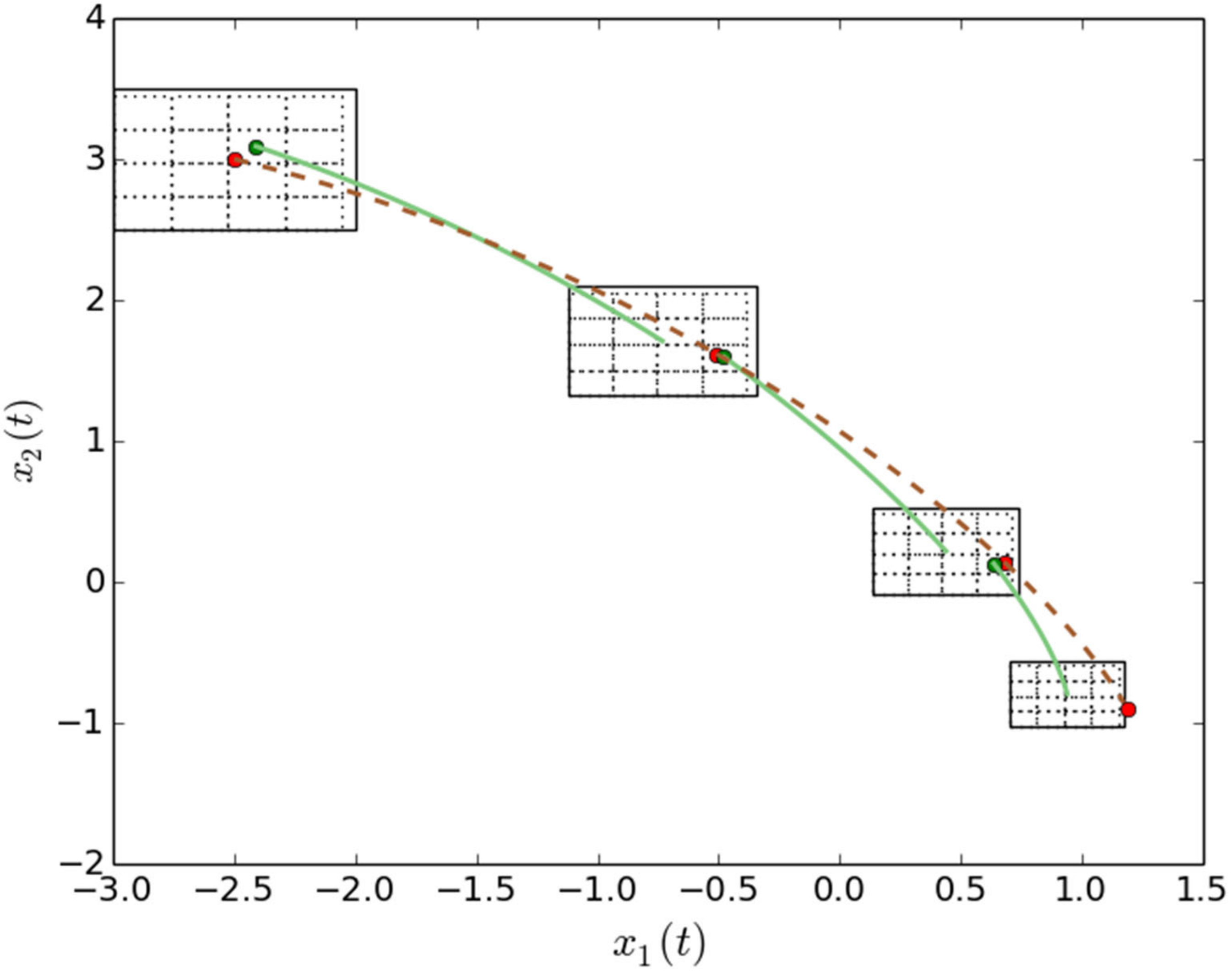
```
If $\xi(x_0, t) \notin S_{i-1}$ output "2"; break
Else  $q_i$:= Quantized measurement of $\xi\left(x_0, iT_p\right)$ w.r.t $C_{i-1}$
      $v(t)$:= $\xi_1\left(q_i, t - (i-1)T_p\right)$ for $t \in [(i-1)T_p, iT_p]$
      $\delta_i \coloneqq e^{-\alpha T_p}\delta_{i-1}$
      $S_i \coloneqq$ hypercube with center $v(iT_p)$ and radius $\delta_i$
      $C_i \coloneqq$ 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



ENTROPY of SWITCHED SYSTEMS

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before
for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$
and active rate $\rho_p(t) := \tau_p(t)/t$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

Solution: $\xi_\sigma(x, t) = \cdots \xi_q(\xi_p(x, t_1), t_2) \cdots$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

Solution: $\xi_\sigma(x,t) = \cdots \xi_q(\xi_p(x,t_1),t_2)\cdots$

$t \in [t_0, t_1]$: $|\xi(x_1,t) - \xi(x_2,t)| \leq e^{L_p(t_1-t_0)}|x_1 - x_2|$

where $L_p := \sup \mu\big(\partial f_p/\partial x(x)\big)$ and the sup is over x reachable from $\operatorname{conv}(K)$ when $\sigma = p$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

Solution: $\xi_\sigma(x, t) = \cdots \xi_q(\xi_p(x, t_1), t_2) \cdots$

$t \in [t_0, t_1]$: $|\xi(x_1, t) - \xi(x_2, t)| \le e^{L_p(t_1 - t_0)}|x_1 - x_2|$

where $L_p := \sup \mu\big(\partial f_p/\partial x(x)\big)$ and

the sup is over x reachable from $\operatorname{conv}(K)$ when $\sigma = p$

$t \in [t_1, t_2]$: $|\xi(x_1, t) - \xi(x_2, t)| \le e^{L_q(t_2 - t_1)}e^{L_p(t_1 - t_0)}|x_1 - x_2|$

where $L_q := \sup \mu\big(\partial f_q/\partial x(x)\big)$ and

the sup is over x reachable from $\operatorname{conv}(K)$ when $\sigma = q$, and so on…

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

modes

For simplicity, take convergence rate $\alpha = 0$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

Solution: $\xi_\sigma(x,t) = \cdots \xi_q(\xi_p(x,t_1),t_2)\cdots$

$t \in [t_0, t_1]$: $|\xi(x_1,t) - \xi(x_2,t)| \le e^{L_p(t_1 - t_0)}|x_1 - x_2|$

where $L_p := \sup \mu\big(\partial f_p/\partial x(x)\big)$ and

the sup is over x reachable from $\mathrm{conv}(K)$ when $\sigma = p$

$t \in [t_1, t_2]$: $|\xi(x_1,t) - \xi(x_2,t)| \le e^{L_q(t_2 - t_1)}e^{L_p(t_1 - t_0)}|x_1 - x_2|$

where $L_q := \sup \mu\big(\partial f_q/\partial x(x)\big)$ and

the sup is over x reachable from $\mathrm{conv}(K)$ when $\sigma = q$, and so on…

$$\boxed{|\xi(x_1,t) - \xi(x_2,t)| \le e^{\sum_p L_p \tau_p(t)}|x_1 - x_2|}$$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \le e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \le e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

$$\boxed{\limsup_{t\to\infty} \sum_p \operatorname{tr}(A_p)\rho_p(t) \le h(A_\sigma) \le \limsup_{t\to\infty} \sum_p n\mu(A_p)\rho_p(t)}$$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \le e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

$$\boxed{\limsup_{t\to\infty} \textstyle\sum_p \operatorname{tr}(A_p)\rho_p(t) \le h(A_\sigma) \le \limsup_{t\to\infty} \sum_p n\mu(A_p)\rho_p(t)}$$

Looser upper bounds depend only on asymptotic active rates or don't depend on $\sigma(\cdot)$ at all, e.g.: $h(A_\sigma) \le n \max_p \mu(A_p)$

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \leq e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

$$\boxed{\limsup_{t\to\infty} \sum_p \operatorname{tr}(A_p)\rho_p(t) \leq h(A_\sigma) \leq \limsup_{t\to\infty} \sum_p n\mu(A_p)\rho_p(t)}$$

Looser upper bounds depend only on asymptotic active rates or don't depend on $\sigma(\cdot)$ at all, e.g.: $h(A_\sigma) \leq n \max_p \mu(A_p)$

Sharper bounds if commutation structure on A_p: [Yang–L–Hespanha '19]

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \le e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

$$\boxed{\limsup_{t\to\infty} \textstyle\sum_p \operatorname{tr}(A_p)\rho_p(t) \le h(A_\sigma) \le \limsup_{t\to\infty} \textstyle\sum_p n\mu(A_p)\rho_p(t)}$$

Looser upper bounds depend only on asymptotic active rates or don’t depend on $\sigma(\cdot)$ at all, e.g.: $h(A_\sigma) \le n \max_p \mu(A_p)$

Sharper bounds if commutation structure on A_p: [Yang–L–Hespanha ’19]

Connections with Lyapunov exponents: [Vicinansa–L ’19]

ENTROPY of SWITCHED SYSTEMS

For switched system $\dot{x} = f_\sigma(x)$, define entropy as before for each fixed switching signal $\sigma : [0, \infty) \mapsto \{1, ..., P\}$

For each mode p, define active time $\tau_p(t) := \int_0^t \mathbf{1}_p(\sigma(s))ds$ and active rate $\rho_p(t) := \tau_p(t)/t$

$$\boxed{|\xi(x_1, t) - \xi(x_2, t)| \le e^{\sum_p L_p \tau_p(t)} |x_1 - x_2|}$$

For switched linear system $\dot{x} = A_\sigma x$, entropy satisfies

$$\boxed{\limsup_{t\to\infty} \sum_p \operatorname{tr}(A_p)\rho_p(t) \le h(A_\sigma) \le \limsup_{t\to\infty} \sum_p n\mu(A_p)\rho_p(t)}$$

Looser upper bounds depend only on asymptotic active rates or don't depend on $\sigma(\cdot)$ at all, e.g.: $h(A_\sigma) \le n \max_p \mu(A_p)$

Sharper bounds if commutation structure on A_p: [Yang–L–Hespanha '19]

Connections with Lyapunov exponents: [Vicinansa–L '19]

Switched nonlinear systems: [Yang–L–Hespanha '21]

OPEN QUESTIONS and ONGOING WORK

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite
- When entropy is infinite, stabilization is in a weaker sense

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite
- When entropy is infinite, stabilization is in a weaker sense
- Finite-data-rate stabilizing controllers for switched systems exist [L ’14, Wakaiki–Yamamoto ’16, Yang–L ’18, Berger–Jungers ’21] but we need to bridge the gap between their data rate and entropy

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite
- When entropy is infinite, stabilization is in a weaker sense
- Finite-data-rate stabilizing controllers for switched systems exist [L ’14, Wakaiki–Yamamoto ’16, Yang–L ’18, Berger–Jungers ’21] but we need to bridge the gap between their data rate and entropy

- Unknown switching signal can be treated as disturbance input

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite
- When entropy is infinite, stabilization is in a weaker sense
- Finite-data-rate stabilizing controllers for switched systems exist [L ’14, Wakaiki–Yamamoto ’16, Yang–L ’18, Berger–Jungers ’21] but we need to bridge the gap between their data rate and entropy

- Unknown switching signal can be treated as disturbance input
- What is a proper definition of entropy for systems with inputs? Preliminary results in [Sibai–Mitra ’18].

OPEN QUESTIONS and ONGOING WORK

- In finite-data-rate scenario, exact switching times are not known
- This increases entropy, making it hard to bound or even infinite
- When entropy is infinite, stabilization is in a weaker sense
- Finite-data-rate stabilizing controllers for switched systems exist [L ’14, Wakaiki–Yamamoto ’16, Yang–L ’18, Berger–Jungers ’21] but we need to bridge the gap between their data rate and entropy

- Unknown switching signal can be treated as disturbance input
- What is a proper definition of entropy for systems with inputs? Preliminary results in [Sibai–Mitra ’18].
- How does entropy behave under input/output interconnections? [Kawan–Delvenne ’16, Matveev et al. ’19, Tomar–Zamani ’20, L ’21].