Ergodicity and Asymptotic Stationarity of Controlled Stochastic Nonlinear Systems under Information Constraints

Nicolas Garcia - Joint work with Christoph Kawan and Serdar Yüksel

Queen's University, Dept. of Mathematics and Statistics

April 26, 2021

Control Under Communication Constraints

- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.

Control Under Communication Constraints

- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.
- We will focus on stochastic non-linear discrete time systems.

Control Under Communication Constraints

- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.
- We will focus on stochastic non-linear discrete time systems.
- The stability notions considered will be asymptotic ergodicity, and asymptotic mean stationarity (AMS).

Control Under Communication Constraints

- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.
- We will focus on stochastic non-linear discrete time systems.
- The stability notions considered will be asymptotic ergodicity, and asymptotic mean stationarity (AMS).
- We will discuss necessary lower bounds on channel capacity required for stochastic stability.

Control Under Communication Constraints

- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.
- We will focus on stochastic non-linear discrete time systems.
- The stability notions considered will be asymptotic ergodicity, and asymptotic mean stationarity (AMS).
- We will discuss necessary lower bounds on channel capacity required for stochastic stability.
- The techniques used build on the notion of invariance entropy for noiseless systems.

Control Under Communication Constraints

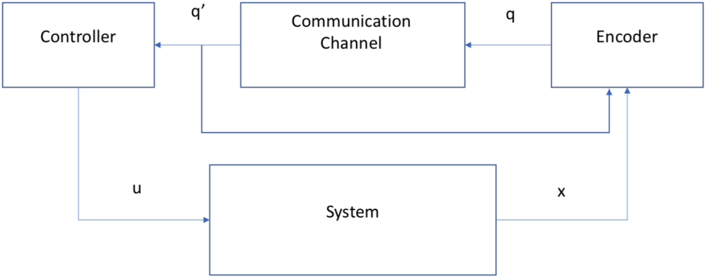
- As per the theme of this seminar, this talk deals with the problem of stabilizing a control system over a rate-limited channel.
- We will focus on stochastic non-linear discrete time systems.
- The stability notions considered will be asymptotic ergodicity, and asymptotic mean stationarity (AMS).
- We will discuss necessary lower bounds on channel capacity required for stochastic stability.
- The techniques used build on the notion of invariance entropy for noiseless systems.
- I will be presenting some joint work with my supervisors Serdar Yüksel and Christoph Kawan, as well as some of their own past work.

Introduction

- A common way to model communication constraints is to impose that the state information travel through a (possibly noisy) information channel to reach the controller.

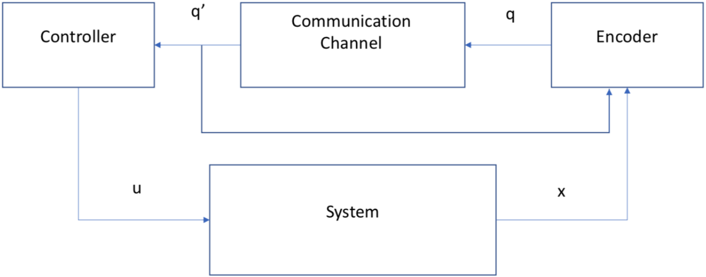
Introduction

- A common way to model communication constraints is to impose that the state information travel through a (possibly noisy) information channel to reach the controller.
- The following diagram provides a graphical representation.



Introduction

- A common way to model communication constraints is to impose that the state information travel through a (possibly noisy) information channel to reach the controller.
- The following diagram provides a graphical representation.



- The diagram depicts a channel with feedback. Of course, not all channels have this feature.

Channel Capacity

For now, we will focus on finite alphabet noiseless channels.

Channel Capacity

For now, we will focus on finite alphabet noiseless channels.

Definition 1

A noiseless channel with finite alphabet $\mathcal{M}$ has capacity $C := \log_2(|\mathcal{M}|)$ bits, where $|\mathcal{M}|$ denotes cardinality of the set $\mathcal{M}$.

Channel Capacity

For now, we will focus on finite alphabet noiseless channels.

Definition 1

A noiseless channel with finite alphabet $\mathcal{M}$ has capacity $C := \log_2(|\mathcal{M}|)$ bits, where $|\mathcal{M}|$ denotes cardinality of the set $\mathcal{M}$.

- When a system is controlled over such a channel, a controller determines its action based on a *reliable estimate* of the system state.

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively.

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively. State and control are $\mathbb{R}^N$-valued and the i.i.d. noise takes values in some standard probability space.

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively. State and control are $\mathbb{R}^N$-valued and the i.i.d. noise takes values in some standard probability space. We will assume that the

random variables $x_0, w_0, w_1, w_2, ...$ are all defined on a common probability space $(\Omega, \mathcal{F}, P)$.

Information Constraints: The above system is controlled over a finite alphabet noiseless channel. I.e.

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively. State and control are $\mathbb{R}^N$-valued and the i.i.d. noise takes values in some standard probability space. We will assume that the

random variables $x_0, w_0, w_1, w_2, \dots$ are all defined on a common probability space $(\Omega, \mathcal{F}, P)$.

Information Constraints: The above system is controlled over a finite alphabet noiseless channel. I.e.

- At each time step, an encoder with perfect state knowledge sends $q_t \in \mathcal{M}$ to the controller over the channel.

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively. State and control are $\mathbb{R}^N$-valued and the i.i.d. noise takes values in some standard probability space. We will assume that the

random variables $x_0, w_0, w_1, w_2, ...$ are all defined on a common probability space $(\Omega, \mathcal{F}, P)$.

Information Constraints: The above system is controlled over a finite alphabet noiseless channel. I.e.

- At each time step, an encoder with perfect state knowledge sends $q_t \in \mathcal{M}$ to the controller over the channel.
- The estimate and control decisions are determined according to a sequence of encoding and control functions $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$ where

$$q_t = \gamma_t^e(x_0, .., x_t), \qquad u_t = \gamma_t^c(q_0, .., q_t)$$

Setup

System: We consider discrete-time non-linear systems of the form

$$x_{t+1} = f(x_t, w_t) + u_t \tag{1}$$

where (x_t), (u_t), and (w_t) are the state, control, and noise processes respectively. State and control are $\mathbb{R}^N$-valued and the i.i.d. noise takes values in some standard probability space. We will assume that the

random variables $x_0, w_0, w_1, w_2, \ldots$ are all defined on a common probability space $(\Omega, \mathcal{F}, P)$.

Information Constraints: The above system is controlled over a finite alphabet noiseless channel. I.e.

- At each time step, an encoder with perfect state knowledge sends $q_t \in \mathcal{M}$ to the controller over the channel.
- The estimate and control decisions are determined according to a sequence of encoding and control functions $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$ where

$$q_t = \gamma_t^e(x_0, .., x_t), \qquad u_t = \gamma_t^c(q_0, .., q_t)$$

- Such a pair of sequences of functions is called a coding and control

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \tag{2}$$

with randomly distributed initial state x_0.

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \tag{2}$$

with randomly distributed initial state x_0. Suppose we fix a coding and control policy for a given noiseless channel, say $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$.

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \tag{2}$$

with randomly distributed initial state x_0. Suppose we fix a coding and control policy for a given noiseless channel, say $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$.

Then $x_1, x_2, x_3, , , ,$ and $u_0, u_1, u_2, ...$ are well defined random variables.

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \tag{2}$$

with randomly distributed initial state x_0. Suppose we fix a coding and control policy for a given noiseless channel, say $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$.

Then $x_1, x_2, x_3, , , ,$ and $u_0, u_1, u_2, ...$ are well defined random variables.Indeed:

$$u_0(\omega) = \gamma_0^c(\gamma_0^e(x_0(\omega))), \quad x_1(\omega) = f(x_0(\omega), w_0(\omega)) + u_0(\omega) \tag{3}$$

with similar definitions for larger time indices.

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \qquad (2)$$

with randomly distributed initial state x_0. Suppose we fix a coding and control policy for a given noiseless channel, say $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$. Then $x_1, x_2, x_3, , , ,$ and $u_0, u_1, u_2, ...$ are well defined random variables.Indeed:

$$u_0(\omega) = \gamma_0^c(\gamma_0^e(x_0(\omega))), \quad x_1(\omega) = f(x_0(\omega), w_0(\omega)) + u_0(\omega) \qquad (3)$$

with similar definitions for larger time indices. If we consider the system above without fixing a causal coding and control policy, then $x_1(\omega), x_2(\omega), x_3(\omega), ...$ and $u_0(\omega), u_1(\omega), ...$ are note well defined.

Setup (Continued)

Digression: Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t. \tag{2}$$

with randomly distributed initial state x_0. Suppose we fix a coding and control policy for a given noiseless channel, say $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$. Then $x_1, x_2, x_3, , , ,$ and $u_0, u_1, u_2, ...$ are well defined random variables.Indeed:

$$u_0(\omega) = \gamma_0^c(\gamma_0^e(x_0(\omega))), \quad x_1(\omega) = f(x_0(\omega), w_0(\omega)) + u_0(\omega) \tag{3}$$

with similar definitions for larger time indices. If we consider the system above without fixing a causal coding and control policy, then $x_1(\omega), x_2(\omega), x_3(\omega), ...$ and $u_0(\omega), u_1(\omega), ...$ are note well defined. Given that we have fixed a causal coding and control policy, we will use

$$x(\omega) \quad u(\omega) \quad w(\omega) \tag{4}$$

to denote the resulting state, control, and noise sequences given $\omega \in \Omega$.

Setup (Continued)

Control Objective: Render the state process stochastically stable (precise definition to come).

Setup (Continued)

Control Objective: Render the state process stochastically stable (precise definition to come).

Results:

- *Ideally:* Provide an exact characterization of the necessary and sufficient channel capacity required for stochastic stabilization.

Setup (Continued)

Control Objective: Render the state process stochastically stable (precise definition to come).

Results:

- *Ideally:* Provide an exact characterization of the necessary and sufficient channel capacity required for stochastic stabilization.
  - This seems intractable with current techniques.

Setup (Continued)

Control Objective: Render the state process stochastically stable (precise definition to come).

Results:

- *Ideally:* Provide an exact characterization of the necessary and sufficient channel capacity required for stochastic stabilization.
  - This seems intractable with current techniques.
- *Reality:* Under some technical assumptions, we can provide lower bounds on channel capacity *necessary* for stochastic stabilization.

Setup (Continued)

Control Objective: Render the state process stochastically stable (precise definition to come).

Results:

- *Ideally:* Provide an exact characterization of the necessary and sufficient channel capacity required for stochastic stabilization.
  - This seems intractable with current techniques.
- *Reality:* Under some technical assumptions, we can provide lower bounds on channel capacity *necessary* for stochastic stabilization.

Before proceeding, some history.

- For linear systems such as

$$x_{t+1} = Ax_t + Bu_t \quad \text{and} \quad \dot{x}(t) = Ax(t) + Bu(t) \quad A \in \mathbb{R}^{n\times n} \tag{5}$$

it has been established (under many different assumptions and stability criteria) that the minimum data rate required for stabilization is the sum of logarithms of unstable eigenvalues of A.

[1] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback, IEEE Transactions on Automatic Control, 44 (1999), pp. 1049–1053.

[2] J. Baillieul, Feedback designs for controlling device arrays with communication channel bandwidth constraints, in ARO workshop on smart structures, University Park PA, 1999, pp. 16– 18.

[3] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control, 49 (2004), pp. 1585– 1597.

[4] F. Colonius and C. Kawan, Invariance entropy for control systems, SIAM Journal on Control and Optimization, 48 (2009), pp. 1701–1721.

Some Brief History

- For linear systems such as

$$x_{t+1} = Ax_t + Bu_t \quad \text{and} \quad \dot{x}(t) = Ax(t) + Bu(t) \quad A \in \mathbb{R}^{n \times n} \qquad (5)$$

  it has been established (under many different assumptions and stability criteria) that the minimum data rate required for stabilization is the sum of logarithms of unstable eigenvalues of A.
- Some early results for such systems include [1] and [2].

[1] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback, IEEE Transactions on Automatic Control, 44 (1999), pp. 1049–1053.

[2] J. Baillieul, Feedback designs for controlling device arrays with communication channel bandwidth constraints, in ARO workshop on smart structures, University Park PA, 1999, pp. 16– 18.

[3] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control, 49 (2004), pp. 1585– 1597.

[4] F. Colonius and C. Kawan, Invariance entropy for control systems, SIAM Journal on Control and Optimization, 48 (2009), pp. 1701–1721.

Some Brief History

- For linear systems such as

$$x_{t+1} = Ax_t + Bu_t \quad \text{and} \quad \dot{x}(t) = Ax(t) + Bu(t) \quad A \in \mathbb{R}^{n \times n} \quad (5)$$

  it has been established (under many different assumptions and stability criteria) that the minimum data rate required for stabilization is the sum of logarithms of unstable eigenvalues of A.
- Some early results for such systems include [1] and [2].
- In [3], the notion of topological feedback entropy (TFE) was introduced for the study of data rates of non-linear discrete time deterministic systems.

[1] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback, IEEE Transactions on Automatic Control, 44 (1999), pp. 1049–1053.

[2] J. Baillieul, Feedback designs for controlling device arrays with communication channel bandwidth constraints, in ARO workshop on smart structures, University Park PA, 1999, pp. 16– 18.

[3] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control, 49 (2004), pp. 1585– 1597.

[4] F. Colonius and C. Kawan, Invariance entropy for control systems, SIAM Journal on Control and Optimization, 48 (2009), pp. 1701–1721.

Some Brief History

- For linear systems such as

$$x_{t+1} = Ax_t + Bu_t \quad \text{and} \quad \dot{x}(t) = Ax(t) + Bu(t) \quad A \in \mathbb{R}^{n\times n} \quad (5)$$

  it has been established (under many different assumptions and stability criteria) that the minimum data rate required for stabilization is the sum of logarithms of unstable eigenvalues of A.
- Some early results for such systems include [1] and [2].
- In [3], the notion of topological feedback entropy (TFE) was introduced for the study of data rates of non-linear discrete time deterministic systems.
- Invariance entropy was introduced in [4] for the same problem, but in continuous time. When modified to the discrete time setting, this notion coincides with TFE.

1 2 3 4

[1] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback, IEEE Transactions on Automatic Control, 44 (1999), pp. 1049–1053.

[2] J. Baillieul, Feedback designs for controlling device arrays with communication channel bandwidth constraints, in ARO workshop on smart structures, University Park PA, 1999, pp. 16– 18.

[3] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control, 49 (2004), pp. 1585– 1597.

[4] F. Colonius and C. Kawan, Invariance entropy for control systems, SIAM Journal on Control and Optimization, 48 (2009), pp. 1701–1721.

Some Brief History (Continued)

- To the best of our knowledge, [5], [6], and [7] are the only works which have established necessary lower bounds on channel capacity for stochastic stability of random non-linear systems.

[5] S. Yüksel , Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels, SIAM Journal on Control and Optimization, 54 (2016), pp. 2844– 2871.

[6] C. Kawan and S. Yüksel, Invariance properties of nonlinear stochastic dynamical systems under information constraints, IEEE Transactions on Automatic Control, to appear (arXiv: 1901.02825), (2020).

[7] N. Garcia, C. Kawan, and S. Yüksel, "Ergodicity conditions for con- trolled stochastic nonlinear systems under information constraints: A volume growth approach," SIAM Journal on Control and Optimization, vol. 59, no. 1, pp. 534–560, 2021.

Some Brief History (Continued)

- To the best of our knowledge, [5], [6], and [7] are the only works which have established necessary lower bounds on channel capacity for stochastic stability of random non-linear systems.
- In the first, information-theoretic methods were used. Stability notions considered were asymptotic mean stationarity, ergodicity, and positive Harris recurrence.

---

[5] S. Yüksel , Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels, SIAM Journal on Control and Optimization, 54 (2016), pp. 2844– 2871.

[6] C. Kawan and S. Yüksel, Invariance properties of nonlinear stochastic dynamical systems under information constraints, IEEE Transactions on Automatic Control, to appear (arXiv: 1901.02825), (2020).

[7] N. Garcia, C. Kawan, and S. Yüksel, "Ergodicity conditions for con- trolled stochastic nonlinear systems under information constraints: A volume growth approach," SIAM Journal on Control and Optimization, vol. 59, no. 1, pp. 534–560, 2021.

Some Brief History (Continued)

- To the best of our knowledge, [5], [6], and [7] are the only works which have established necessary lower bounds on channel capacity for stochastic stability of random non-linear systems.
- In the first, information-theoretic methods were used. Stability notions considered were asymptotic mean stationarity, ergodicity, and positive Harris recurrence.
- In the latter two, stabilization entropy was used. This notion is a modification of invariance entropy for discrete-time stochastic systems and will be explained in detail later in the talk.

5 6 7

[5] S. Yüksel , Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels, SIAM Journal on Control and Optimization, 54 (2016), pp. 2844– 2871.

[6] C. Kawan and S. Yüksel, Invariance properties of nonlinear stochastic dynamical systems under information constraints, IEEE Transactions on Automatic Control, to appear (arXiv: 1901.02825), (2020).

[7] N. Garcia, C. Kawan, and S. Yüksel, "Ergodicity conditions for con- trolled stochastic nonlinear systems under information constraints: A volume growth approach," SIAM Journal on Control and Optimization, vol. 59, no. 1, pp. 534–560, 2021.

Stochastic Stability

We now give a precise definition of the stability notions considered.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n \in \mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.
- Σ inherits a Borel σ-algebra $\mathcal{B}(\Sigma)$ from $\mathbb{R}^N$ using the product topology.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n \in \mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.
- Σ inherits a Borel σ-algebra $\mathcal{B}(\Sigma)$ from $\mathbb{R}^N$ using the product topology.
- To define a probability measure on $\mathcal{B}(\Sigma)$ it suffices to define it on finite dimensional rectangles, i.e. sets of the form

$$(B_0, B_1, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots) \tag{6}$$

  for $B_0, .., B_m \in \mathcal{B}(\mathbb{R}^N)$.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.
- Σ inherits a Borel σ-algebra $\mathcal{B}(\Sigma)$ from $\mathbb{R}^N$ using the product topology.
- To define a probability measure on $\mathcal{B}(\Sigma)$ it suffices to define it on finite dimensional rectangles, i.e. sets of the form

$$(B_0, B_1, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots) \tag{6}$$

  for $B_0, .., B_m \in \mathcal{B}(\mathbb{R}^N)$.
- This is true since sets of the above form generate $\mathcal{B}(\Sigma)$.

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.
- Σ inherits a Borel σ-algebra $\mathcal{B}(\Sigma)$ from $\mathbb{R}^N$ using the product topology.
- To define a probability measure on $\mathcal{B}(\Sigma)$ it suffices to define it on finite dimensional rectangles, i.e. sets of the form

$$(B_0, B_1, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots) \tag{6}$$

  for $B_0, .., B_m \in \mathcal{B}(\mathbb{R}^N)$.
- This is true since sets of the above form generate $\mathcal{B}(\Sigma)$.
- The stochastic process $(X_n)_{n\in\mathbb{N}}$ induces a measure μ on $\mathcal{B}(\Sigma)$ defined by:

$$\mu((B_0, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots)) = P(\{\omega \in \Omega : X_i(\omega) \in B_i \text{ for } i = 0, .., m\}).$$

Stochastic Stability

We now give a precise definition of the stability notions considered.

- Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued stochastic process defined on a common probability space $(\Omega, \mathcal{F}, P)$.
- Let $\Sigma := (\mathbb{R}^N)^{\mathbb{N}}$ denote the set of sequences with elements in $\mathbb{R}^N$.
- Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra of $\mathbb{R}^N$.
- Σ inherits a Borel σ-algebra $\mathcal{B}(\Sigma)$ from $\mathbb{R}^N$ using the product topology.
- To define a probability measure on $\mathcal{B}(\Sigma)$ it suffices to define it on finite dimensional rectangles, i.e. sets of the form

$$(B_0, B_1, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots) \tag{6}$$

  for $B_0, .., B_m \in \mathcal{B}(\mathbb{R}^N)$.
- This is true since sets of the above form generate $\mathcal{B}(\Sigma)$.
- The stochastic process $(X_n)_{n\in\mathbb{N}}$ induces a measure μ on $\mathcal{B}(\Sigma)$ defined by:

$$\mu((B_0, .., B_m, \mathbb{R}^N, \mathbb{R}^N, \dots)) = P(\{\omega \in \Omega : X_i(\omega) \in B_i \text{ for } i = 0, .., m\}).$$

- μ is known as the process measure corresponding to $(X_n)_{n\in\mathbb{N}}$.

A few more definitions:

Stochastic Stability (Continued)

A few more definitions:

- Let $\theta : \Sigma \to \Sigma$ denote the left shift map on the sequence space (I.e. $\theta((x_t))_n = x_{n+1}$ for $(x_t) \in \Sigma$).

Stochastic Stability (Continued)

A few more definitions:

- Let $\theta : \Sigma \to \Sigma$ denote the left shift map on the sequence space (I.e. $\theta((x_t))_n = x_{n+1}$ for $(x_t) \in \Sigma$).
- Let $\mathcal{F}_{inv(\theta)}$ denote the collection of $\mathcal{B}(\Sigma)$-measurable sets which are θ-invariant. I.e

$$\mathcal{F}_{inv(\theta)} := \{A \in \mathcal{B}(\Sigma) : A = \theta^{-1}(A)\}. \quad (7)$$

Stochastic Stability (Continued)

A few more definitions:

- Let $\theta : \Sigma \to \Sigma$ denote the left shift map on the sequence space (I.e. $\theta((x_t))_n = x_{n+1}$ for $(x_t) \in \Sigma$).
- Let $\mathcal{F}_{inv(\theta)}$ denote the collection of $\mathcal{B}(\Sigma)$-measurable sets which are θ-invariant. I.e

$$\mathcal{F}_{inv(\theta)} := \{A \in \mathcal{B}(\Sigma) : A = \theta^{-1}(A)\}. \tag{7}$$

- It is not hard to check that $\mathcal{F}_{inv(\theta)}$ is itself a σ-algebra on Σ.

Stochastic Stability (Continued)

A few more definitions:

- Let $\theta : \Sigma \to \Sigma$ denote the left shift map on the sequence space (I.e. $\theta((x_t))_n = x_{n+1}$ for $(x_t) \in \Sigma$).
- Let $\mathcal{F}_{inv(\theta)}$ denote the collection of $\mathcal{B}(\Sigma)$-measurable sets which are θ-invariant. I.e

$$\mathcal{F}_{inv(\theta)} := \{A \in \mathcal{B}(\Sigma) : A = \theta^{-1}(A)\}. \tag{7}$$

- It is not hard to check that $\mathcal{F}_{inv(\theta)}$ is itself a σ-algebra on Σ.
- We are ready to define distinct notions of stochastic stability.

Stationary Processes

Definition 2

(Stationarity) Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.

Stationary Processes

Definition 2

(Stationarity) Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *stationary (or measure-preserving)* iff $\mu(B) = \mu(\theta^{-1}(B))$ for all $B \in \mathcal{B}(\Sigma)$.

Definition 2

(Stationarity) Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *stationary (or measure-preserving)* iff $\mu(B) = \mu(\theta^{-1}(B))$ for all $B \in \mathcal{B}(\Sigma)$.

A consequence of $(X_n)_{n\in\mathbb{N}}$ being stationary is that if we fix an arbitrary collection of indices $k_1, .., k_n \in \mathbb{N}$ and a collection $B_1, .., B_n \in \mathcal{B}(\mathbb{R}^N)$ of Borel sets, then

$$P(\bigcap_{i=1}^{n}\{X_{k_i} \in B_i\}) = P(\bigcap_{i=1}^{n}\{X_{k_i+l} \in B_i\}) \tag{8}$$

for any integer l.

Stationary Processes

Definition 2

(Stationarity) Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *stationary (or measure-preserving)* iff $\mu(B) = \mu(\theta^{-1}(B))$ for all $B \in \mathcal{B}(\Sigma)$.

A consequence of $(X_n)_{n\in\mathbb{N}}$ being stationary is that if we fix an arbitrary collection of indices $k_1, .., k_n \in \mathbb{N}$ and a collection $B_1, .., B_n \in \mathcal{B}(\mathbb{R}^N)$ of Borel sets, then

$$P(\bigcap_{i=1}^{n}\{X_{k_i} \in B_i\}) = P(\bigcap_{i=1}^{n}\{X_{k_i+l} \in B_i\}) \tag{8}$$

for any integer l. In other words, the finite dimensional distributions are invariant under shifts.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, ...)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, ...)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, ...)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged. From now on we write Q for both the projected and sequence space measures.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, \ldots)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged. From now on we write Q for both the projected and sequence space measures.

Definition 3

Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, ...)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged. From now on we write Q for both the projected and sequence space measures.

Definition 3

Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *asymptotically mean stationary (AMS)* iff there exists a measure Q (called the Asymptotic Mean of the process) on $(\Sigma, \mathcal{B}(\Sigma))$ such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mu(\theta^{-k}(B)) = Q(B) \quad \text{for all } B \in \mathcal{B}(\Sigma).$$

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, \ldots)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged. From now on we write Q for both the projected and sequence space measures.

Definition 3

Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *asymptotically mean stationary (AMS)* iff there exists a measure Q (called the Asymptotic Mean of the process) on $(\Sigma, \mathcal{B}(\Sigma))$ such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mu(\theta^{-k}(B)) = Q(B) \quad \text{for all } B \in \mathcal{B}(\Sigma).$$

- It is immediate that stationarity implies AMS.

Asymptotic Mean Stationarity (AMS)

Remark: Suppose Q is a stationary measure on $\mathcal{B}(\Sigma)$.We can unambiguously project it to a measure on $\mathcal{B}(\mathbb{R}^N)$ by defining

$$Q'(B) = Q((B, \mathbb{R}^N, \mathbb{R}^N, \mathbb{R}^N, \ldots)) \quad \text{for any } B \in \mathcal{B}(\mathbb{R}^N). \tag{9}$$

If we were to put B in any other position, stationarity would result in $Q'(B)$ remaining unchanged. From now on we write Q for both the projected and sequence space measures.

Definition 3

Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in $\mathbb{R}^N$ and let μ be its process measure on $\mathcal{B}(\Sigma)$.We say that the process $(x_n)_{n\in\mathbb{N}}$ is:

- *asymptotically mean stationary (AMS)* iff there exists a measure Q (called the Asymptotic Mean of the process) on $(\Sigma, \mathcal{B}(\Sigma))$ such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mu(\theta^{-k}(B)) = Q(B) \quad \text{for all } B \in \mathcal{B}(\Sigma).$$

- It is immediate that stationarity implies AMS.
- Not hard to show that the AMS mean Q is stationary.

Ergodicity and Asymptotic Ergodicity

Definition 4

Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued process with process measure μ.

Ergodicity and Asymptotic Ergodicity

Definition 4

Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued process with process measure μ. We say it is:

- *ergodic* iff it is stationary and $\mu(A) \in \{0,1\}$ for every $A \in \mathcal{F}_{inv(\theta)}$

Ergodicity and Asymptotic Ergodicity

Definition 4

Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued process with process measure μ. We say it is:

- *ergodic* iff it is stationary and $\mu(A) \in \{0, 1\}$ for every $A \in \mathcal{F}_{inv(\theta)}$
- *AMS ergodic (or asymptotically ergodic)* iff $(X_n)_{n\in\mathbb{N}}$ is AMS with asymptotic mean Q, and Q is additionally ergodic.

Ergodicity and Asymptotic Ergodicity

Definition 4

Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued process with process measure μ. We say it is:

- *ergodic* iff it is stationary and $\mu(A) \in \{0, 1\}$ for every $A \in \mathcal{F}_{inv(\theta)}$
- *AMS ergodic (or asymptotically ergodic)* iff $(X_n)_{n\in\mathbb{N}}$ is AMS with asymptotic mean Q, and Q is additionally ergodic.

Interpretation of Ergodicity

- Ergodicity is an non-decomposability condition.

Ergodicity and Asymptotic Ergodicity

Definition 4

Let $(X_n)_{n\in\mathbb{N}}$ be an $\mathbb{R}^N$-valued process with process measure μ. We say it is:

- *ergodic* iff it is stationary and $\mu(A) \in \{0, 1\}$ for every $A \in \mathcal{F}_{inv(\theta)}$
- *AMS ergodic (or asymptotically ergodic)* iff $(X_n)_{n\in\mathbb{N}}$ is AMS with asymptotic mean Q, and Q is additionally ergodic.

Interpretation of Ergodicity

- Ergodicity is an non-decomposability condition.
- It tells us that in a certain sense, the long term behavior of all sample paths is the same.

Example: Frequency of visits to a set.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \tag{10}$$

for some constant $c \in [0, 1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \tag{10}$$

for some constant $c \in [0,1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

- A contains exactly the sequences which asymptotically spend $100c$ percent of the time in B.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \qquad (10)$$

for some constant $c \in [0, 1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

- A contains exactly the sequences which asymptotically spend $100c$ percent of the time in B.
- The set A is easily seen to be θ-invariant.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \tag{10}$$

for some constant $c \in [0, 1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

- A contains exactly the sequences which asymptotically spend $100c$ percent of the time in B.
- The set A is easily seen to be θ-invariant.
- If (X_n) is ergodic, then $\mu(A)$ is either zero or one. I.e. either (almost) all sample paths visit a region at a given frequency, or (almost) none of them do.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \tag{10}$$

for some constant $c \in [0, 1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

- A contains exactly the sequences which asymptotically spend $100c$ percent of the time in B.
- The set A is easily seen to be θ-invariant.
- If (X_n) is ergodic, then $\mu(A)$ is either zero or one. I.e. either (almost) all sample paths visit a region at a given frequency, or (almost) none of them do.

Asymptotic ergodicity is a relaxation of ergodicity which still provides almost-sure guarantees on asymptotic sample path behavior.

Ergodicity and Asymptotic Ergodicity (Continued)

Example: Frequency of visits to a set. Consider the event

$$A := \{(x_n) \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k) = c\} \quad (10)$$

for some constant $c \in [0,1]$ and some Borel set $B \subseteq \mathbb{R}^N$.

- A contains exactly the sequences which asymptotically spend $100c$ percent of the time in B.
- The set A is easily seen to be θ-invariant.
- If (X_n) is ergodic, then $\mu(A)$ is either zero or one. I.e. either (almost) all sample paths visit a region at a given frequency, or (almost) none of them do.

Asymptotic ergodicity is a relaxation of ergodicity which still provides almost-sure guarantees on asymptotic sample path behavior.

We have all the background to state the main theorems discussed in this talk.

Results

The following are two theorems which provide lower bounds on channel capacity for ergodic and AMS stabilization.

Theorem 5

Consider the control system

$$x_{t+1} = f(x_t) + w_t + u_t \tag{11}$$

with state, i.i.d noise, and control taking values in $\mathbb{R}^N$.

Results

The following are two theorems which provide lower bounds on channel capacity for ergodic and AMS stabilization.

Theorem 5

Consider the control system

$$x_{t+1} = f(x_t) + w_t + u_t \tag{11}$$

with state, i.i.d noise, and control taking values in $\mathbb{R}^N$. *Suppose that*

- *the system is controlled over a finite alphabet channel with capacity* C,

Results

The following are two theorems which provide lower bounds on channel capacity for ergodic and AMS stabilization.

Theorem 5

Consider the control system

$$x_{t+1} = f(x_t) + w_t + u_t \tag{11}$$

with state, i.i.d noise, and control taking values in $\mathbb{R}^N$. *Suppose that*

- *the system is controlled over a finite alphabet channel with capacity* C,
- *there exists a causal coding and control policy which renders the state process AMS with asymptotic mean* Q.

Results

The following are two theorems which provide lower bounds on channel capacity for ergodic and AMS stabilization.

Theorem 5

Consider the control system

$$x_{t+1} = f(x_t) + w_t + u_t \tag{11}$$

with state, i.i.d noise, and control taking values in $\mathbb{R}^N$*. Suppose that*

- *the system is controlled over a finite alphabet channel with capacity* C*,*
- *there exists a causal coding and control policy which renders the state process AMS with asymptotic mean* Q*.*

Then if some technical assumptions are satisfied, we must have that

$$Q(B)\log_2\left(\inf_{x\in B}|\det Df(x)|\right) \leq C \tag{12}$$

for any $B \in \mathbb{R}^N$ *with finite and non-zero Lebesgue measure.*

Results (Continued)

Theorem 6

Consider the control system

$$x_{t+1} = f(x_t, w_t) + u_t \tag{13}$$

with state and control taking values in $\mathbb{R}^N$, *and i.i.d noise* w_t *with law* ν.

Theorem 6

Consider the control system

$$x_{t+1} = f(x_t, w_t) + u_t \tag{13}$$

with state and control taking values in $\mathbb{R}^N$, *and i.i.d noise* w_t *with law* ν.*Suppose that*

- *the system state is made asymptotically ergodic over a noiseless channel of capacity* C *with ergodic AMS mean* Q.

Theorem 6

Consider the control system

$$x_{t+1} = f(x_t, w_t) + u_t \tag{13}$$

with state and control taking values in $\mathbb{R}^N$, *and i.i.d noise* w_t *with law* ν*.Suppose that*

- *the system state is made asymptotically ergodic over a noiseless channel of capacity* C *with ergodic AMS mean* Q.

Under technical assumptions, we must have that

$$\int \int \log_2 |\det Df_w(x)| \, \mathrm{d}Q(x) \, \mathrm{d}\nu(w) \leq C. \tag{14}$$

where f_w *denotes the map* $x \mapsto f(x, w)$ *for a fixed noise symbol* w.

Theorem 6

Consider the control system

$$x_{t+1} = f(x_t, w_t) + u_t \tag{13}$$

with state and control taking values in $\mathbb{R}^N$, *and i.i.d noise* w_t *with law* ν*.Suppose that*

- *the system state is made asymptotically ergodic over a noiseless channel of capacity* C *with ergodic AMS mean* Q*.*

Under technical assumptions, we must have that

$$\int\int \log_2 |\det Df_w(x)| \,\mathrm{d}Q(x)\,\mathrm{d}\nu(w) \leq C. \tag{14}$$

where f_w *denotes the map* $x \mapsto f(x, w)$ *for a fixed noise symbol* w*.*

Note: For both of the stated theorems, an identical result can be proven for scalar systems controlled over Discrete Memoryless Channels.

Intuition for the Bound

Consider the linear system

$$x_{t+1} = 5x_t + u_t \tag{15}$$

and imagine x_0 is uniformly distributed on $[-1, 1]$.

Intuition for the Bound

Consider the linear system

$$x_{t+1} = 5x_t + u_t \tag{15}$$

and imagine x_0 is uniformly distributed on $[-1, 1]$.

- Suppose we wish to render the interval $[-1, 1]$ invariant.

Intuition for the Bound

Consider the linear system

$$x_{t+1} = 5x_t + u_t \tag{15}$$

and imagine x_0 is uniformly distributed on $[-1, 1]$.

- Suppose we wish to render the interval $[-1, 1]$ invariant.
- We know $x_0 \in [-1, 1]$ therefore we can quantize the state in bins

$$[-1, -0.6), [-0.6, -0.2), [-0.2, 0.2), [0.2, 0.6), [0.6, 1] \tag{16}$$

and apply the respective control decisions:

$$4, 2, 0, -2, -4. \tag{17}$$

This ensures that $x_1 \in [-1, 1]$.

Intuition for the Bound

Consider the linear system

$$x_{t+1} = 5x_t + u_t \tag{15}$$

and imagine x_0 is uniformly distributed on $[-1, 1]$.

- Suppose we wish to render the interval $[-1, 1]$ invariant.
- We know $x_0 \in [-1, 1]$ therefore we can quantize the state in bins

$$[-1, -0.6), [-0.6, -0.2), [-0.2, 0.2), [0.2, 0.6), [0.6, 1] \tag{16}$$

and apply the respective control decisions:

$$4, 2, 0, -2, -4. \tag{17}$$

This ensures that $x_1 \in [-1, 1]$. We repeat this process for $t = 2, 3, \ldots$

Thus if controlled over a noiseless channel, we need a capacity of at least $\log_2(5)$ bits to accomplish this.

Intuition for the Bound

Consider the linear system

$$x_{t+1} = 5x_t + u_t \tag{15}$$

and imagine x_0 is uniformly distributed on $[-1, 1]$.

- Suppose we wish to render the interval $[-1, 1]$ invariant.
- We know $x_0 \in [-1, 1]$ therefore we can quantize the state in bins

$$[-1, -0.6), [-0.6, -0.2), [-0.2, 0.2), [0.2, 0.6), [0.6, 1] \tag{16}$$

  and apply the respective control decisions:

$$4, 2, 0, -2, -4. \tag{17}$$

  This ensures that $x_1 \in [-1, 1]$. We repeat this process for $t = 2, 3, \ldots$

Thus if controlled over a noiseless channel, we need a capacity of at least $\log_2(5)$ bits to accomplish this.For a non-linear system whose state trajectories are ergodic with measure Q, we might expect that we should average the logarithm of the Jacobian determinant of the system dynamics function w.r.t. the measure Q.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

- Suppose a causal coding and control policy exists which maintains the state process (x_t) contained in a compact set $B \in \mathcal{B}(\mathbb{R}^N)$, and suppose $x_0 \in B$.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

- Suppose a causal coding and control policy exists which maintains the state process (x_t) contained in a compact set $B \in \mathcal{B}(\mathbb{R}^N)$, and suppose $x_0 \in B$.
- At a given time, the controller receives C bits thus can generate at most 2^C different control decisions.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

- Suppose a causal coding and control policy exists which maintains the state process (x_t) contained in a compact set $B \in \mathcal{B}(\mathbb{R}^N)$, and suppose $x_0 \in B$.
- At a given time, the controller receives C bits thus can generate at most 2^C different control decisions.
- During the first T time steps, the controller can thus generate no more than 2^{CT} distinct control sequences.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

- Suppose a causal coding and control policy exists which maintains the state process (x_t) contained in a compact set $B \in \mathcal{B}(\mathbb{R}^N)$, and suppose $x_0 \in B$.
- At a given time, the controller receives C bits thus can generate at most 2^C different control decisions.
- During the first T time steps, the controller can thus generate no more than 2^{CT} distinct control sequences.
- Suppose S_T is the smallest cardinality set of open-loop control sequences of length T with which we can render B invariant for the first T time steps.

Invariance Entropy

The notion of stabilization entropy was inspired by Invariance Entropy [Colonius-Kawan 09'] which we now briefly discuss. Consider a noiseless system $x_{t+1} = f(x_t, u_t)$ controlled over a noiseless channel of capacity C.

- Suppose a causal coding and control policy exists which maintains the state process (x_t) contained in a compact set $B \in \mathcal{B}(\mathbb{R}^N)$, and suppose $x_0 \in B$.
- At a given time, the controller receives C bits thus can generate at most 2^C different control decisions.
- During the first T time steps, the controller can thus generate no more than 2^{CT} distinct control sequences.
- Suppose S_T is the smallest cardinality set of open-loop control sequences of length T with which we can render B invariant for the first T time steps.
- By the assumption that B can be rendered invariant using closed-loop control, we must have that

$$|S_T| \leq 2^{CT}. \tag{18}$$

Invariance Entropy

It follows from $|S_T| \leq 2^{CT}$ that

$$\frac{1}{T}\log_2|S_T| \leq \frac{1}{T}\log_2 2^{CT} = C. \tag{19}$$

Invariance Entropy

It follows from $|S_T| \leq 2^{CT}$ that

$$\frac{1}{T}\log_2|S_T| \leq \frac{1}{T}\log_2 2^{CT} = C. \tag{19}$$

- The smallest number of control sequences required to accomplish a control task may be therefore be used to obtain lower bounds on channel capacity.

It follows from $|S_T| \leq 2^{CT}$ that

$$\frac{1}{T}\log_2|S_T| \leq \frac{1}{T}\log_2 2^{CT} = C. \tag{19}$$

- The smallest number of control sequences required to accomplish a control task may be therefore be used to obtain lower bounds on channel capacity.
- The quantity $\limsup_{T\to\infty}\frac{1}{T}\log_2|S_T|$ is known as the Invariance Entropy for the set B.

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.
- For this reason, stabilization entropy [Kawan-Yüksel] was introduced.

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.
- For this reason, stabilization entropy [Kawan-Yüksel] was introduced.
- For this notion, it is only required that the rate at which a set B is visited be above some threshold.

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.
- For this reason, stabilization entropy [Kawan-Yüksel] was introduced.
- For this notion, it is only required that the rate at which a set B is visited be above some threshold.

Notation: if we fix an initial state x_0, a noise sequence $w := (w_t)$, and control sequence $u := (u_t)$, the state trajectory (x_t) is uniquely defined. I.e.

$$x_1 = f(x_0, w_0, u_0), \quad x_2 = f(x_1, w_1, u_1), \quad x_3 = f(x_2, w_2, u_2), \ldots. \tag{20}$$

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.
- For this reason, stabilization entropy [Kawan-Yüksel] was introduced.
- For this notion, it is only required that the rate at which a set B is visited be above some threshold.

Notation: if we fix an initial state x_0, a noise sequence $w := (w_t)$, and control sequence $u := (u_t)$, the state trajectory (x_t) is uniquely defined. I.e.

$$x_1 = f(x_0, w_0, u_0), \quad x_2 = f(x_1, w_1, u_1), \quad x_3 = f(x_2, w_2, u_2), \ldots. \tag{20}$$

To make explicit the fixed parameters, we use the notation

$$\varphi(t, x_0, u, w) := x_t. \tag{21}$$

Stabilization Entropy

- Consider again the stochastic system $x_{t+1} = f(x_t, w_t, u_t)$ with $\mathbb{R}^N$-valued state and control.
- Set invariance is too much to ask due to the presence of noise.
- For this reason, stabilization entropy [Kawan-Yüksel] was introduced.
- For this notion, it is only required that the rate at which a set B is visited be above some threshold.

Notation: if we fix an initial state x_0, a noise sequence $w := (w_t)$, and control sequence $u := (u_t)$, the state trajectory (x_t) is uniquely defined. I.e.

$$x_1 = f(x_0, w_0, u_0), \quad x_2 = f(x_1, w_1, u_1), \quad x_3 = f(x_2, w_2, u_2), \dots \tag{20}$$

To make explicit the fixed parameters, we use the notation

$$\varphi(t, x_0, u, w) := x_t. \tag{21}$$

Notation: Given $\omega \in \Omega$, the initial state and noise sequence are deterministic. We denote them by $x_0(\omega)$ and $w(\omega)$.

Stabilization Entropy

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

Stabilization Entropy

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

- a Borel set $B \subseteq \mathbb{R}^N$, and a time horizon $T \in \mathbb{N}$.

Stabilization Entropy

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

- a Borel set $B \subseteq \mathbb{R}^N$, and a time horizon $T \in \mathbb{N}$.
- a rate $r \in (0, 1)$ and a probability $\rho \in (0, 1)$.

Stabilization Entropy

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

- a Borel set $B \subseteq \mathbb{R}^N$, and a time horizon $T \in \mathbb{N}$.
- a rate $r \in (0,1)$ and a probability $\rho \in (0,1)$.

A subset $S \subseteq (\mathbb{R}^N)^T$ of control sequences of length T is called (B, r, ρ, T)-spanning for the above system iff there exists a set $\tilde{\Omega} \in \mathcal{F}$ such that $P(\tilde{\Omega}) > \rho$ and such that

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

- a Borel set $B \subseteq \mathbb{R}^N$, and a time horizon $T \in \mathbb{N}$.
- a rate $r \in (0,1)$ and a probability $\rho \in (0,1)$.

A subset $S \subseteq (\mathbb{R}^N)^T$ of control sequences of length T is called (B, r, ρ, T)-spanning for the above system iff there exists a set $\tilde{\Omega} \in \mathcal{F}$ such that $P(\tilde{\Omega}) > \rho$ and such that

- for any $\omega \in \tilde{\Omega}$ there exists $u \in S$ with the property that

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega))\} \in B| \geq r. \quad (22)$$

Stabilization Entropy

Consider the control system $x_{t+1} = f(x_t, w_t, u_t)$. For now, we *do not* fix a coding and control policy.

Definition 7

(Spanning Sets) Fix the following objects:

- a Borel set $B \subseteq \mathbb{R}^N$, and a time horizon $T \in \mathbb{N}$.
- a rate $r \in (0,1)$ and a probability $\rho \in (0,1)$.

A subset $S \subseteq (\mathbb{R}^N)^T$ of control sequences of length T is called (B, r, ρ, T)-spanning for the above system iff there exists a set $\tilde{\Omega} \in \mathcal{F}$ such that $P(\tilde{\Omega}) > \rho$ and such that

- for any $\omega \in \tilde{\Omega}$ there exists $u \in S$ with the property that

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega))\} \in B| \geq r. \tag{22}$$

Digression: The second condition tells us the following. Let $\omega \in \tilde{\Omega}$. Then there exists $u \in S$ such that when we iterate the initial state $x_0(\omega)$ together with the sequences u and $w(\omega)$, the rate of state visits to B is no smaller than r on the time interval $0, 1, .., T-1$.

Stabilization Entropy

Interpretation of spanning set: Let S be a (B, r, ρ, T) spanning set. Then using only open loop control with sequences from the set S, we can, with probability ρ, ensure that the rate of visits to B is no smaller than r (on the time interval $0, .., T-1$).

Interpretation of spanning set: Let S be a (B, r, ρ, T) spanning set. Then using only open loop control with sequences from the set S, we can, with probability ρ, ensure that the rate of visits to B is no smaller than r (on the time interval $0, .., T-1$).

Definition 8

For arbitrary $T \in \mathbb{N}$, let S_T denote a minimum cardinality (B, r, ρ, T)-spanning set.

Interpretation of spanning set: Let S be a (B, r, ρ, T) spanning set. Then using only open loop control with sequences from the set S, we can, with probability ρ, ensure that the rate of visits to B is no smaller than r (on the time interval $0, .., T-1$).

Definition 8

For arbitrary $T \in \mathbb{N}$, let S_T denote a minimum cardinality (B, r, ρ, T)-spanning set. We define the (B, r, ρ)-stabilization entropy as

$$h(B, r, \rho) = \limsup_{T \to \infty} \frac{1}{T} \log_2 |S_T|. \tag{23}$$

Stabilization Entropy

Interpretation of spanning set: Let S be a (B, r, ρ, T) spanning set. Then using only open loop control with sequences from the set S, we can, with probability ρ, ensure that the rate of visits to B is no smaller than r (on the time interval $0, .., T-1$).

Definition 8

For arbitrary $T \in \mathbb{N}$, let S_T denote a minimum cardinality (B, r, ρ, T)-spanning set. We define the (B, r, ρ)-stabilization entropy as

$$h(B, r, \rho) = \limsup_{T \to \infty} \frac{1}{T} \log_2 |S_T|. \tag{23}$$

The notion of stabilization entropy is crucial in the proofs of the two main theorems stated earlier.

Proof Techniques

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Proof Techniques

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Theorem 9

Suppose that the $\mathbb{R}^N$-valued system,

$$x_{t+1} = f(x_t) + w_t + u_t, \quad (w_t) \text{ i.i.d. with measure } v$$

is made ergodic with process measure Q via a causal coding and control policy over a noiseless channel with capacity $C = \log_2 |\mathcal{M}|$.

Proof Techniques

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Theorem 9

Suppose that the $\mathbb{R}^N$-valued system,

$$x_{t+1} = f(x_t) + w_t + u_t, \quad (w_t) \text{ i.i.d. with measure } \nu$$

is made ergodic with process measure Q via a causal coding and control policy over a noiseless channel with capacity $C = \log_2 |\mathcal{M}|$. Suppose also that:

(A1) *The map $f : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective.*

Proof Techniques

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Theorem 9

Suppose that the $\mathbb{R}^N$-valued system,

$$x_{t+1} = f(x_t) + w_t + u_t, \quad (w_t) \text{ i.i.d. with measure } v$$

is made ergodic with process measure Q via a causal coding and control policy over a noiseless channel with capacity $C = \log_2 |\mathcal{M}|$. Suppose also that:

(A1) *The map $f : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective.*

(A2) *The initial state x_0 is random and admits a bounded density π_0.*

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Theorem 9

Suppose that the $\mathbb{R}^N$-valued system,

$$x_{t+1} = f(x_t) + w_t + u_t, \quad (w_t) \text{ i.i.d. with measure } v$$

is made ergodic with process measure Q via a causal coding and control policy over a noiseless channel with capacity $C = \log_2 |\mathcal{M}|$. Suppose also that:

(A1) *The map $f : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective.*

(A2) *The initial state x_0 is random and admits a bounded density π_0.*

(A3) *We have that $|\det Df(x)| > 1$ for all $x \in \mathbb{R}^N$.*

Proof Techniques

We now state and sketch a proof for a simplified version of the theorems presented earlier.

Theorem 9

Suppose that the $\mathbb{R}^N$-valued system,

$$x_{t+1} = f(x_t) + w_t + u_t, \quad (w_t) \text{ i.i.d. with measure } v$$

is made ergodic with process measure Q via a causal coding and control policy over a noiseless channel with capacity $C = \log_2 |\mathcal{M}|$. Suppose also that:

(A1) *The map $f : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective.*

(A2) *The initial state x_0 is random and admits a bounded density π_0.*

(A3) *We have that $|\det Df(x)| > 1$ for all $x \in \mathbb{R}^N$.*

Then for any $B \in \mathcal{B}(\mathbb{R}^N)$ with finite and non-zero Lebesgue measure, we have that

$$Q(B) \log_2(\inf_{x \in B} |\det Df(x)|) \leq C.$$

Proof Technique

Proof Technique

- Fix a Borel set $B \subseteq \mathbb{R}^N$ with finite and nonzero Lebesgue measure for which we will prove the bounds. WLOG we can assume $0 < Q(B) < 1$.

Proof Technique

- Fix a Borel set $B \subseteq \mathbb{R}^N$ with finite and nonzero Lebesgue measure for which we will prove the bounds. WLOG we can assume $0 < Q(B) < 1$.
- By assumption, there exists coding and control policy which renders the state process ergodic with ergodic (thus stationary!) process measure Q. Fix such a coding and control policy.

Proof Technique

- Fix a Borel set $B \subseteq \mathbb{R}^N$ with finite and nonzero Lebesgue measure for which we will prove the bounds. WLOG we can assume $0 < Q(B) < 1$.
- By assumption, there exists coding and control policy which renders the state process ergodic with ergodic (thus stationary!) process measure Q. Fix such a coding and control policy.
- Note that given $\omega \in \Omega$, the state sequence $x(\omega)$, noise sequence $w(\omega)$ and control sequence $u(\omega)$ are all well defined (since we have fixed a coding and control policy).

Proof Technique

- Fix a Borel set $B \subseteq \mathbb{R}^N$ with finite and nonzero Lebesgue measure for which we will prove the bounds. WLOG we can assume $0 < Q(B) < 1$.
- By assumption, there exists coding and control policy which renders the state process ergodic with ergodic (thus stationary!) process measure Q. Fix such a coding and control policy.
- Note that given $\omega \in \Omega$, the state sequence $x(\omega)$, noise sequence $w(\omega)$ and control sequence $u(\omega)$ are all well defined (since we have fixed a coding and control policy).

Lemma 10

For every $\epsilon > 0$ sufficiently small and any probability $\rho \in (0,1)$ we have that

$$h(B, Q(B) - \epsilon, \rho) \leq C. \tag{24}$$

Proof Technique

Proof of Lemma:

Proof Technique

Proof of Lemma: By the pointwise ergodic theorem, one obtains that

$$P\Big(\{\omega \in \Omega : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) = Q(B)\}\Big) = 1.$$

Proof Technique

Proof of Lemma: By the pointwise ergodic theorem, one obtains that

$$P\Big(\{\omega \in \Omega : \lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) = Q(B)\}\Big) = 1.$$

In Words: Almost surely, each sample path $x(\omega)$ will asymptotically visit B at a rate of $Q(B)$.

Proof Technique

Proof of Lemma: By the pointwise ergodic theorem, one obtains that

$$P\Big(\{\omega \in \Omega : \lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) = Q(B)\}\Big) = 1.$$

In Words: Almost surely, each sample path $x(\omega)$ will asymptotically visit B at a rate of $Q(B)$.

By taking T sufficiently large and through continuity of probability arguments, one can show that:

Proof Technique

Proof of Lemma: By the pointwise ergodic theorem, one obtains that

$$P\left(\{\omega \in \Omega : \lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) = Q(B)\}\right) = 1.$$

In Words: Almost surely, each sample path $x(\omega)$ will asymptotically visit B at a rate of $Q(B)$.

By taking T sufficiently large and through continuity of probability arguments, one can show that:

- there exists a set $\tilde{\Omega} \in \mathcal{F}$ with $P(\tilde{\Omega}) > \rho$,

Proof Technique

Proof of Lemma: By the pointwise ergodic theorem, one obtains that

$$P\left(\{\omega \in \Omega : \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) = Q(B)\}\right) = 1.$$

In Words: Almost surely, each sample path $x(\omega)$ will asymptotically visit B at a rate of $Q(B)$.

By taking T sufficiently large and through continuity of probability arguments, one can show that:

- there exists a set $\tilde{\Omega} \in \mathcal{F}$ with $P(\tilde{\Omega}) > \rho$,
- for any $\omega \in \tilde{\Omega}$ we have that

$$\frac{1}{T} \sum_{k=0}^{T-1} \mathbb{1}_B(x_k(\omega)) > Q(B) - \epsilon. \tag{25}$$

Proof Technique

Define the set

$$S_T = \{u(\omega) \in (\mathbb{R}^N)^T : \omega \in \tilde{\Omega}\}. \tag{26}$$

Proof Technique

Define the set

$$S_T = \{u(\omega) \in (\mathbb{R}^N)^T : \omega \in \tilde{\Omega}\}. \tag{26}$$

Then for any $\omega \in \tilde{\Omega}$, there exists $u \in S_T$ (just take $u(\omega)$) with

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega)) \in B\}| \geq Q(B) - \epsilon.$$

Proof Technique

Define the set

$$S_T = \{u(\omega) \in (\mathbb{R}^N)^T : \omega \in \tilde{\Omega}\}. \tag{26}$$

Then for any $\omega \in \tilde{\Omega}$, there exists $u \in S_T$ (just take $u(\omega)$) with

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega)) \in B\}| \geq Q(B) - \epsilon.$$

This establishes that S_T is a $(B, Q(B) - \epsilon, \rho, T)$-spanning set.

Proof Technique

Define the set

$$S_T = \{u(\omega) \in (\mathbb{R}^N)^T : \omega \in \tilde{\Omega}\}. \tag{26}$$

Then for any $\omega \in \tilde{\Omega}$, there exists $u \in S_T$ (just take $u(\omega)$) with

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega)) \in B\}| \geq Q(B) - \epsilon.$$

This establishes that S_T is a $(B, Q(B) - \epsilon, \rho, T)$-spanning set.
Note that at time T, the coding and control policy can generate no more than $|\mathcal{M}|^T$ distinct control sequences.

Proof Technique

Define the set

$$S_T = \{u(\omega) \in (\mathbb{R}^N)^T : \omega \in \tilde{\Omega}\}. \tag{26}$$

Then for any $\omega \in \tilde{\Omega}$, there exists $u \in S_T$ (just take $u(\omega)$) with

$$\frac{1}{T}|\{t \in \{0, 1, .., T-1\} : \varphi(t, x_0(\omega), u, w(\omega)) \in B\}| \geq Q(B) - \epsilon.$$

This establishes that S_T is a $(B, Q(B) - \epsilon, \rho, T)$-spanning set.
Note that at time T, the coding and control policy can generate no more than $|\mathcal{M}|^T$ distinct control sequences.As such, $|S_T| \leq |\mathcal{M}|^T$ and we obtain

$$h(B, Q(B) - \epsilon, \rho) \leq \limsup_{T\to\infty}(\frac{1}{T}\log_2 |S_T|) \tag{27}$$

$$\leq \limsup_{T\to\infty}(\frac{1}{T}\log_2 |\mathcal{M}|^T) = \log_2 |\mathcal{M}| = C \tag{28}$$

which completes the proof of the lemma.

Proof Technique

We now sketch the proof of the simplified theorem.

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set.

Proof Technique

We now sketch the proof of the simplified theorem.

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set.
- Imagine that we can obtain a lower bound

$$L^T \leq |S_T| \tag{29}$$

  for some $L > 0$ which does not depend on the specific spanning set S_T chosen.
- Taking S_T to be a minimum cardinality spanning set, we would obtain

Proof Technique

We now sketch the proof of the simplified theorem.

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set.
- Imagine that we can obtain a lower bound

$$L^T \leq |S_T| \tag{29}$$

  for some $L > 0$ which does not depend on the specific spanning set S_T chosen.
- Taking S_T to be a minimum cardinality spanning set, we would obtain

$$C \geq h(B, Q(B) - \epsilon, \rho) = \lim_{T \to \infty} \frac{1}{T} \log_2 |S_T| \geq \lim_{T \to \infty} \frac{1}{T} \log_2 L^T = \log_2 L.$$

Proof Technique

We now sketch the proof of the simplified theorem.

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set.
- Imagine that we can obtain a lower bound

$$L^T \leq |S_T| \tag{29}$$

  for some $L > 0$ which does not depend on the specific spanning set S_T chosen.
- Taking S_T to be a minimum cardinality spanning set, we would obtain

$$C \geq h(B, Q(B) - \epsilon, \rho) = \lim_{T\to\infty} \frac{1}{T} \log_2 |S_T| \geq \lim_{T\to\infty} \frac{1}{T} \log_2 L^T = \log_2 L.$$

We will indeed find a suitable L which results in

$$Q(B) \log_2(\inf_{x\in B} |\det Df(x)|) \leq C. \tag{30}$$

for the set $B \subseteq \mathbb{R}^N$ that was previously fixed.

Proof Technique

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set and let $\tilde{\Omega}$ denote the associated subset of $\mathcal{F}$ with $P(\tilde{\Omega}) > \rho$.

Proof Technique

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set and let $\tilde{\Omega}$ denote the associated subset of $\mathcal{F}$ with $P(\tilde{\Omega}) > \rho$.
- Let also m denote the N-dimensional Lebesgue measure, and let $M > 0$ be an upper bound for the density of π_0.

Proof Technique

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set and let $\tilde{\Omega}$ denote the associated subset of $\mathcal{F}$ with $P(\tilde{\Omega}) > \rho$.
- Let also m denote the N-dimensional Lebesgue measure, and let $M > 0$ be an upper bound for the density of π_0.
- Consider the set $\{(w(\omega), x_0(\omega)) : \omega \in \tilde{\Omega}\}$.

Proof Technique

- Let S_T be a finite $(B, Q(B)-\epsilon, \rho, T)$-spanning set and let $\tilde{\Omega}$ denote the associated subset of $\mathcal{F}$ with $P(\tilde{\Omega}) > \rho$.
- Let also m denote the N-dimensional Lebesgue measure, and let $M > 0$ be an upper bound for the density of π_0.
- Consider the set $\{(w(\omega), x_0(\omega)) : \omega \in \tilde{\Omega}\}$.
- Decomposing it into 'fibers' and using Fubini-Tonelli, one can obtain

$$0 < \frac{\rho}{M} \leq |S_T| \max_{u \in S_T} \int m(A(u,w)) d\nu(w). \tag{31}$$

where

$$A(u,w) := \{x \in \mathbb{R}^N : \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_B(\varphi(t,x,u,w)) \geq Q(B) - \epsilon\}.$$

Proof Technique

- Let S_T be a finite $(B, Q(B) - \epsilon, \rho, T)$-spanning set and let $\tilde{\Omega}$ denote the associated subset of $\mathcal{F}$ with $P(\tilde{\Omega}) > \rho$.
- Let also m denote the N-dimensional Lebesgue measure, and let $M > 0$ be an upper bound for the density of π_0.
- Consider the set $\{(w(\omega), x_0(\omega)) : \omega \in \tilde{\Omega}\}$.
- Decomposing it into 'fibers' and using Fubini-Tonelli, one can obtain

$$0 < \frac{\rho}{M} \leq |S_T| \max_{u \in S_T} \int m(A(u, w)) d\nu(w). \tag{31}$$

where

$$A(u, w) := \{x \in \mathbb{R}^N : \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_B(\varphi(t, x, u, w)) \geq Q(B) - \epsilon\}.$$

In words: $A(u, w)$ consists of initial conditions that, when paired with sequences u and w, visit the B at a rate no smaller than $Q(B) - \epsilon$ during the first T time steps.

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u, w))$ when iterated up to time $T - 1$.

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u,w))$ when iterated up to time $T-1$. More precisely, define the function

$$\varphi_{t,u,w}(\cdot) := \varphi(t,\cdot,u,w), \tag{32}$$

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u,w))$ when iterated up to time $T-1$. More precisely, define the function

$$\varphi_{t,u,w}(\cdot) := \varphi(t,\cdot,u,w), \tag{32}$$

and attempt to bound the volume of.

$$\varphi_{T-1,u,w}(A(u,w)) \tag{33}$$

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u,w))$ when iterated up to time $T-1$. More precisely, define the function

$$\varphi_{t,u,w}(\cdot) := \varphi(t,\cdot,u,w), \tag{32}$$

and attempt to bound the volume of.

$$\varphi_{T-1,u,w}(A(u,w)) \tag{33}$$

Idea: Consider a set $V \subseteq B$. We have that

$$m(V)(\inf_{x\in V} |\det Df(x)|) \leq m(f(V)). \tag{34}$$

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u,w))$ when iterated up to time $T-1$. More precisely, define the function

$$\varphi_{t,u,w}(\cdot) := \varphi(t,\cdot,u,w), \tag{32}$$

and attempt to bound the volume of.

$$\varphi_{T-1,u,w}(A(u,w)) \tag{33}$$

Idea: Consider a set $V \subseteq B$. We have that

$$m(V)(\inf_{x\in V}|\det Df(x)|) \leq m(f(V)). \tag{34}$$

We know that initial states in $A(u,w)$ will visit the set B at a frequency no less than $Q(B)-\epsilon$. We also have that by assumption $m(A) \leq m(f(A))$ for any $A \subseteq R^N$.

Proof Technique

Goal: Obtain a 'small' upper bound for the volume $m(A(u,w))$ when iterated up to time $T-1$. More precisely, define the function

$$\varphi_{t,u,w}(\cdot) := \varphi(t,\cdot,u,w), \tag{32}$$

and attempt to bound the volume of.

$$\varphi_{T-1,u,w}(A(u,w)) \tag{33}$$

Idea: Consider a set $V \subseteq B$. We have that

$$m(V)(\inf_{x\in V} |\det Df(x)|) \leq m(f(V)). \tag{34}$$

We know that initial states in $A(u,w)$ will visit the set B at a frequency no less than $Q(B)-\epsilon$. We also have that by assumption $m(A) \leq m(f(A))$ for any $A \subseteq R^N$.We might therefore be tempted to say that

$$m(A(u,w))(\inf_{x\in B} |\det Df(x)|)^{(Q(B)-\epsilon)T} \leq m(\varphi_{T-1,u,w}(A(u,w))). \tag{35}$$

Proof Technique

Problem: This does not quite work however:

Proof Technique

Problem: This does not quite work however:

- Even though trajectories obtained by iterating elements in $A(u, w)$ necessarily visit B at least $(Q(B) - \epsilon)T$ times, they need not visit B at the same time!

Proof Technique

Problem: This does not quite work however:

- Even though trajectories obtained by iterating elements in $A(u, w)$ necessarily visit B at least $(Q(B) - \epsilon)T$ times, they need not visit B at the same time!
- As such, at any given time, part of the iterated volume may not be in B and we cannot apply the previous bound.

Proof Technique

Problem: This does not quite work however:

- Even though trajectories obtained by iterating elements in $A(u, w)$ necessarily visit B at least $(Q(B) - \epsilon)T$ times, they need not visit B at the same time!
- As such, at any given time, part of the iterated volume may not be in B and we cannot apply the previous bound.

Solution:

- We can instead write $A(u, w)$ as a disjoint union of sets, each of which contains initial states that when iterated, visit B at the desired frequency, and *at the same times*.

Proof Technique

Problem: This does not quite work however:

- Even though trajectories obtained by iterating elements in $A(u, w)$ necessarily visit B at least $(Q(B) - \epsilon)T$ times, they need not visit B at the same time!
- As such, at any given time, part of the iterated volume may not be in B and we cannot apply the previous bound.

Solution:

- We can instead write $A(u, w)$ as a disjoint union of sets, each of which contains initial states that when iterated, visit B at the desired frequency, and *at the same times*.
- This allows us to bound the volumes using the quantity $(\inf_{x \in B} |\det Df(x)|)$.

Proof Technique

Problem: This does not quite work however:

- Even though trajectories obtained by iterating elements in $A(u, w)$ necessarily visit B at least $(Q(B) - \epsilon)T$ times, they need not visit B at the same time!
- As such, at any given time, part of the iterated volume may not be in B and we cannot apply the previous bound.

Solution:

- We can instead write $A(u, w)$ as a disjoint union of sets, each of which contains initial states that when iterated, visit B at the desired frequency, and *at the same times*.
- This allows us to bound the volumes using the quantity $(\inf_{x \in B} |\det Df(x)|)$.
- By doing this with as few partitions and bounding each one, we obtain

$$0 < \frac{\rho}{M} \leq |S_T| \max_{u \in S_T} \int m(A(u, w)) d\nu(w)$$

$$\leq |S_T| \cdot m(B)((Q(B) - \epsilon)T)^{-(Q(B)-\epsilon)T}.$$

Proof Technique

Rearranging the inequality on the previous slide results in

$$\frac{\rho}{M \cdot m(B)} \big(\inf_{x \in B} |\det Df(x)| \big)^{(Q(B)-\epsilon)T} \leq |S_T|.$$

Proof Technique

Rearranging the inequality on the previous slide results in

$$\frac{\rho}{M \cdot m(B)}\left(\inf_{x \in B} |\det Df(x)|\right)^{(Q(B)-\epsilon)T} \leq |S_T|.$$

As per the discussion a few slides back, we have our lower bound on the minimum-cardinality spanning set $|S_T|$.

Proof Technique

Rearranging the inequality on the previous slide results in

$$\frac{\rho}{M \cdot m(B)}(\inf_{x \in B} |\det Df(x)|)^{(Q(B)-\epsilon)T} \leq |S_T|.$$

As per the discussion a few slides back, we have our lower bound on the minimum-cardinality spanning set $|S_T|$.Taking logs, normalizing by $1/T$, taking a limit as $T \to \infty$, and noting that $\epsilon > 0$ was arbitrary, we obtain

$$Q(B)\log_2(\inf_{x \in B} |\det Df(x)|) \leq \lim_{T \to \infty} \frac{1}{T}\log_2 |S_T| \leq C$$

as desired.

Proof Technique

Rearranging the inequality on the previous slide results in

$$\frac{\rho}{M \cdot m(B)} (\inf_{x \in B} |\det Df(x)|)^{(Q(B)-\epsilon)T} \leq |S_T|.$$

As per the discussion a few slides back, we have our lower bound on the minimum-cardinality spanning set $|S_T|$.Taking logs, normalizing by $1/T$, taking a limit as $T \to \infty$, and noting that $\epsilon > 0$ was arbitrary, we obtain

$$Q(B) \log_2(\inf_{x \in B} |\det Df(x)|) \leq \lim_{T \to \infty} \frac{1}{T} \log_2 |S_T| \leq C$$

as desired. This completes the proof of the simplified theorem.

Generalization

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) dv(w) \leq C \qquad (36)$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$.

Generalization

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) dv(w) \leq C \tag{36}$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$. This bound is obtained using the above approach, however:

Generalization

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int \int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \qquad (36)$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$. This bound is obtained using the above approach, however:

- We view (x_t, w_t) as an 'augmented' ergodic process.

Generalization

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) dv(w) \leq C \tag{36}$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$. This bound is obtained using the above approach, however:

- We view (x_t, w_t) as an 'augmented' ergodic process.
- We define stabilization entropy for disjoint collections of sets of the form $(B_i \times W_j)$.

Generalization

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int\int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{36}$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$. This bound is obtained using the above approach, however:

- We view (x_t, w_t) as an 'augmented' ergodic process.
- We define stabilization entropy for disjoint collections of sets of the form $(B_i \times W_j)$.
- We obtain the bound

$$\sum_{i,j}(Q \times v)(B_i \times W_j)\log_2(\inf_{(x,w)\in D_i \times W_j}|\det Df_w(x)|) \leq C. \tag{37}$$

Recall that we previously claimed that under asymptotic ergodicity, we obtain the bound

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) dv(w) \leq C \tag{36}$$

for the system $x_{t+1} = f(x_t, w_t) + u_t$. This bound is obtained using the above approach, however:

- We view (x_t, w_t) as an 'augmented' ergodic process.
- We define stabilization entropy for disjoint collections of sets of the form $(B_i \times W_j)$.
- We obtain the bound

$$\sum_{i,j} (Q \times v)(B_i \times W_j) \log_2(\inf_{(x,w) \in D_i \times W_j} |\det Df_w(x)|) \leq C. \tag{37}$$

- As this holds for arbitrarily fine partitions, we can approximate the integral from below with the above bounds, from which the result follows.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.
- One can characterize capacity in terms of the mutual information between its input and outputs.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.
- One can characterize capacity in terms of the mutual information between its input and outputs.
- In the proofs involving DMCs, we combine stabilization entropy techniques with the strong converse for the channel theorems for DMCs.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.
- One can characterize capacity in terms of the mutual information between its input and outputs.
- In the proofs involving DMCs, we combine stabilization entropy techniques with the strong converse for the channel theorems for DMCs.
- The idea is to view the control decisions as a code with which to estimate x_0 at the controller end.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.
- One can characterize capacity in terms of the mutual information between its input and outputs.
- In the proofs involving DMCs, we combine stabilization entropy techniques with the strong converse for the channel theorems for DMCs.
- The idea is to view the control decisions as a code with which to estimate x_0 at the controller end.
- If the desired channel capacity inequalities do not hold, one shows that ergodic/AMS stabilization is not possible, since if it were, one would be able to reconstruct x_0 with non-vanishing probability.

DMC Case

- As mentioned earlier, the theorems stated in this talk also hold for scalar systems controlled over Discrete Memoryless Channels (DMCs).
- A DMC is a channel in which the channel output at time t is a random variable that is independent from all other random variables except for the channel input at time t.
- One can characterize capacity in terms of the mutual information between its input and outputs.
- In the proofs involving DMCs, we combine stabilization entropy techniques with the strong converse for the channel theorems for DMCs.
- The idea is to view the control decisions as a code with which to estimate x_0 at the controller end.
- If the desired channel capacity inequalities do not hold, one shows that ergodic/AMS stabilization is not possible, since if it were, one would be able to reconstruct x_0 with non-vanishing probability.
- This results in a contradiction with the strong converse theorem for DMCs.

Concluding Remarks

- As mentioned in the history section, bounds of the form

$$\int \int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{38}$$

have been obtained using information theoretic methods.

Concluding Remarks

- As mentioned in the history section, bounds of the form

$$\int \int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{38}$$

  have been obtained using information theoretic methods.
- Such bounds rely on inequalities involving differential entropy and mutual information.

Concluding Remarks

- As mentioned in the history section, bounds of the form

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) d\nu(w) \leq C \qquad (38)$$

  have been obtained using information theoretic methods.
- Such bounds rely on inequalities involving differential entropy and mutual information.
- This approach is better suited for dealing with a larger class of communication channels, but cannot deal with systems for which state variables do not admit differential entropies.

Concluding Remarks

- Finally, we note that a recent refinement (using stabilization entropy tools) to the bound

$$\int \int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{39}$$

has been obtained, where we allow for the integration to be done over any arbitrary subset of the coordinates of the system state space.

Concluding Remarks

- Finally, we note that a recent refinement (using stabilization entropy tools) to the bound

$$\int\int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{39}$$

has been obtained, where we allow for the integration to be done over any arbitrary subset of the coordinates of the system state space.
- This allows one to eliminate 'volume contracting' directions, and results in a tighter bound.

Concluding Remarks

- Finally, we note that a recent refinement (using stabilization entropy tools) to the bound

$$\int\int \log_2(|\det Df_w(x)|)dQ(x)dv(w) \leq C \tag{39}$$

has been obtained, where we allow for the integration to be done over any arbitrary subset of the coordinates of the system state space.
- This allows one to eliminate 'volume contracting' directions, and results in a tighter bound.
- This refinement is not possible with information theoretic techniques, where one cannot 'decouple' distinct coordinates of the system state-space due to non-zero mutual information between the distinct coordinates.

Concluding Remarks

- Finally, we note that a recent refinement (using stabilization entropy tools) to the bound

$$\int \int \log_2(|\det Df_w(x)|) dQ(x) dv(w) \leq C \tag{39}$$

has been obtained, where we allow for the integration to be done over any arbitrary subset of the coordinates of the system state space.
- This allows one to eliminate 'volume contracting' directions, and results in a tighter bound.
- This refinement is not possible with information theoretic techniques, where one cannot 'decouple' distinct coordinates of the system state-space due to non-zero mutual information between the distinct coordinates.
- For a state space decomposing into 'stable' and 'unstable' components, the bound becomes

$$\int \int \log_2(|\det D_{x^u} f_w(x^u, x^s)|) dQ(x^u, x^s) dv(w) \leq C \tag{40}$$

where the Jacobian determinant is a square matrix of size $< N$.

THE END