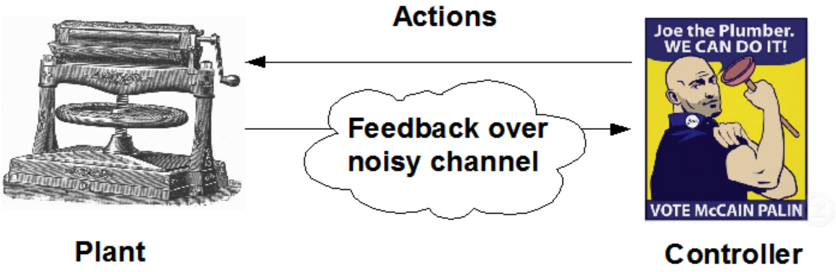
Anytime Capacity of Stabilization of a Linear System over Noisy Channel

Graduate Seminar in Area I (6.454)
October 26, 2011

Outline

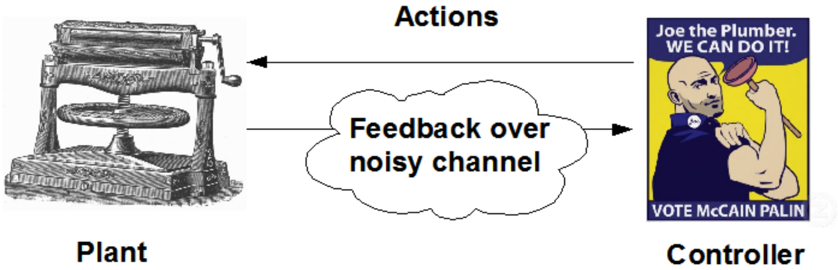
Outline

Control and Communications



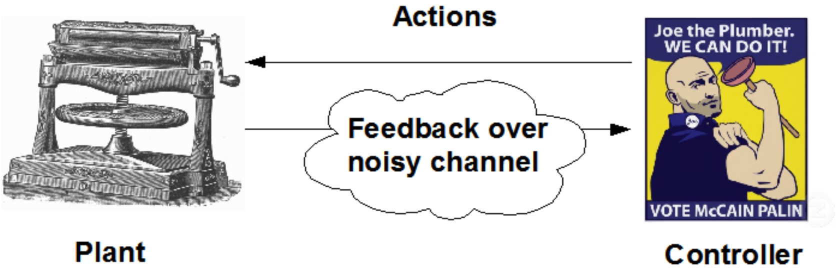
- **General Problem**: Stabilizing an unstable plant with noisy feedback.
- How much “information” do we need?
- What is the correct measure of “information”?

Control and Communications



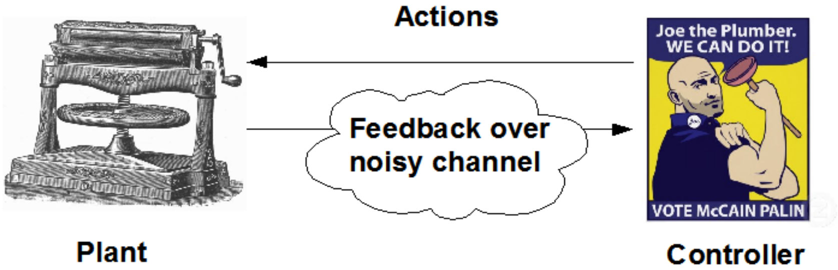
- **General Problem**: Stabilizing an unstable plant with noisy feedback.
- How much “information” do we need?
- What is the correct measure of “information”?

Control and Communications



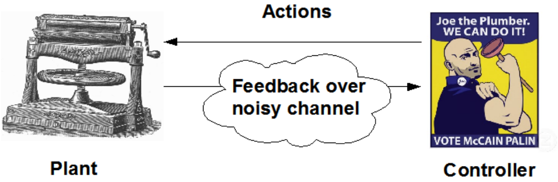
- **General Problem**: Stabilizing an unstable plant with noisy feedback.
- How much "information" do we need?
- What is the correct measure of "information"?

Control and Communications



- **General Problem**: Stabilizing an unstable plant with noisy feedback.
- How much "information" do we need?
- What is the correct measure of "information"?

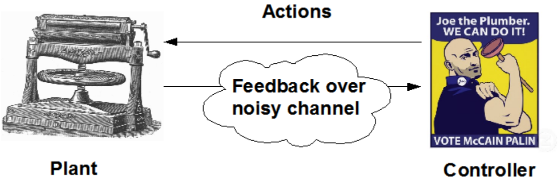
Control and Communications



Main insights

- How much "information" do we need?
  - No single answer. It depends on the degree of "stability" desirable.
- What is the correct measure of "information"?
  - *Shannon capacity* may not be adequate for stronger notions of stability. Need *anytime capacity*.

Control and Communications



Main insights

- How much "information" do we need?
  - No single answer. It depends on the degree of "stability" desirable.
- What is the correct measure of "information"?
  - *Shannon capacity* may not be adequate for stronger notions of stability. Need *anytime capacity*.

Plan of This talk

- A simple example to illustrate that Shannon capacity is not strong enough for control applications.
  - In particular, a plant can be *unstable* even if the Shannon capacity of the channel is *infinite*.
- A necessary condition for stability in terms of anytime capacity.

Plan of This talk

- A simple example to illustrate that Shannon capacity is not strong enough for control applications.
  - In particular, a plant can be *unstable* even if the Shannon capacity of the channel is *infinite*.
- A necessary condition for stability in terms of anytime capacity.

Main Reference

A. Sahai, S. K. Mitter, "The Necessity and Sufficiency of Anytime Capacity for Stabilization of a Linear System Over a Noisy Communication Link. Part I: Scalar Systems," IEEE Trans. Inform. Th., vol. 52, no. 8, pp. 3369-3395, Aug. 2006.

Outline

The Control Problem

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \in \mathbb{Z}^+.$$

- Time (discrete): $t \in \mathbb{Z}^+$.
- State: $X_t \in \mathbb{R}$.
- control: $U_t \in \mathbb{R}$.
- *Bounded* disturbance: $|W_t| < \frac{\Omega}{2}$, with probability 1.

To make things interesting:

- *unstable* gain: $\lambda > 1$.

The Control Problem

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \in \mathbb{Z}^+.$$

- **Goal**: choose good U_t to keep X_t "small".
- If feedback is perfect, simply set $U_t = -\lambda X_t$.
- What if feedback is sent through a noisy channel?

The Control Problem

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \in \mathbb{Z}^+.$$

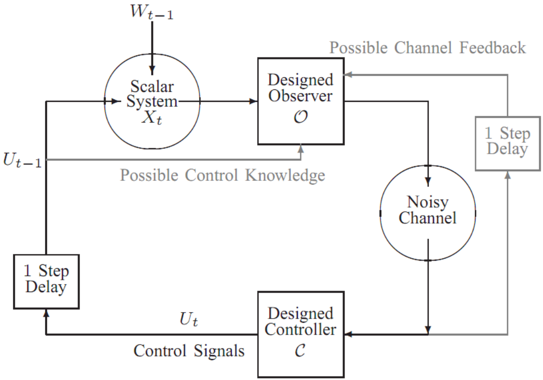
- **Goal**: choose good U_t to keep X_t "small".
- If feedback is perfect, simply set $U_t = -\lambda X_t$.
- What if feedback is sent through a noisy channel?

The Control Problem

$$X_{t+1} = \lambda X_t + U_t + W_t, \quad t \in \mathbb{Z}^+.$$

- **Goal**: choose good U_t to keep X_t "small".
- If feedback is perfect, simply set $U_t = -\lambda X_t$.
- What if feedback is sent through a noisy channel?

Definition of Stability



- **Observer** $\mathcal{O}$: sees X_t and generates channel input a_t.
- **Controller** $\mathcal{C}$: observes channel output B_t and generates control signal U_t.

The Control Problem

Definition: η-stability

A closed-loop system is η-stable if there exists $K < \infty$, such that

$$\mathbb{E}\left[|X_t|^\eta\right] < K$$

for all $t \geq 0$.

(More general notions of stability can be defined, but we will focus on η-stability for now.)

Counter Example in Real-Erasure Channel

When is Shannon capacity not sufficient in describing communications in control systems?

Real Erasure Channel (REC)

The *real packet erasure channel* has

- Input alphabet: $\mathcal{A} = \mathbb{R}$.
- Output alphabet: $\mathcal{B} = \mathbb{R}$.
- Transition probabilities

$$\begin{aligned} p(x|x) &= 1-\delta, \\ p(0|x) &= \delta. \end{aligned}$$

I.e., a symbol is either received *perfectly*, or received as *zero*.

Counter Example in Real-Erasure Channel

- What is the Shannon capacity of the channel?
- It is *infinite*, because a real number can carry as many bits as we want.

Counter Example in Real-Erasure Channel

- What is the Shannon capacity of the channel?
- It is *infinite*, because a real number can carry as many bits as we want.

Counter Example in Real-Erasure Channel

- What is the optimal communication / control policy?
- **Communication**: set
$$a_t = X_t.$$
- **Control**: set
$$U_t = -\lambda B_t.$$
- **Resulting dynamics**: X_t is reset to 0 every $Geo(\delta)$ steps.

Counter Example in Real-Erasure Channel

- Is the system η-stable under optimal control?
- It is 1-stable,

$$\mathbb{E}[|X_t|] = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right) < 1,$$

  for all t.
- However, it is *not* η-stable, for $\eta \geq 2$,

$$\mathbb{E}[|X_t|^2] > \frac{4\sigma^2}{5}\sum_{i=0}^{t}\left(\left(\frac{9}{8}\right)^{i+1} - \left(\frac{1}{2}\right)^{i+1}\right)$$

  which diverges as $t \to \infty$.

Counter Example in Real-Erasure Channel

Lesson learned: notion of information depends on the strength of stability required (e.g., values of η).

- Why was Shannon capacity insufficient?
- Need good information about the system state at *all times*, not just the end of a large block.
- Fix: define a stronger notion of capacity to guarantee good estimation of system state at any point in time ("anytime capacity").

Counter Example in Real-Erasure Channel

Lesson learned: notion of information depends on the strength of stability required (e.g., values of η).

- Why was Shannon capacity insufficient?
- Need good information about the system state at *all times*, not just the end of a large block.
- Fix: define a stronger notion of capacity to guarantee good estimation of system state at any point in time ("anytime capacity").

Counter Example in Real-Erasure Channel

Lesson learned: notion of information depends on the strength of stability required (e.g., values of η).

- Why was Shannon capacity insufficient?
- Need good information about the system state at *all times*, not just the end of a large block.
- Fix: define a stronger notion of capacity to guarantee good estimation of system state at any point in time ("anytime capacity").

Outline

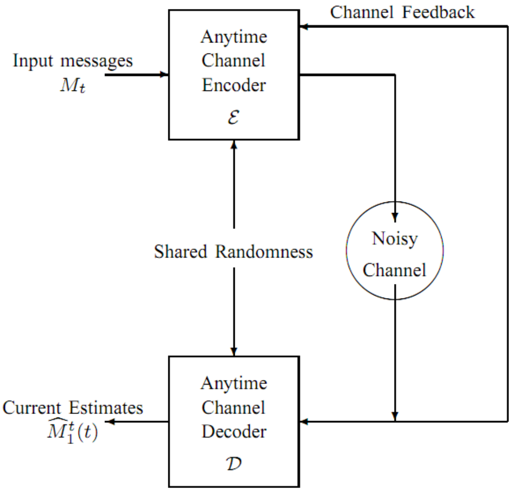
Anytime Reliability and Capacity

Communication System

A rate R communication system is

- Encoder receives R-bit message M_t in slot t. (details on whiteboard)
- Encoder produces channel input based on all past messages and possible feedback $B_1^{t-1-\theta}$ (with delay $1+\theta$).
- Decoder updates estimates of *all* past messages, $\hat{M}_i(t)$, for all $i \leq t$, based on *all* channel outputs till time t.

Anytime Reliability and Capacity



Anytime Reliability and Capacity

Anytime Reliability

A rate R communication system achieves *anytime reliability* α if there exists constant K such that

$$\mathbb{P}\left(\hat{M}_1^i(t) \neq M_1^i\right) \leq K 2^{-\alpha(t-i)}.$$

The system is *uniformly anytime reliable* if the above holds for all messages M.

- Comparing to Shannon reliability? Block versus sequential?
- Exercise: fix t or i and vary the other.

Anytime Reliability and Capacity

α-anytime Capacity

$C_{any}(\alpha)$ of a channel is the highest rate R, at which the channel can achieve uniform anytime reliability α.

More stringent than Shannon capacity, C:

$$C_{any}(\alpha) \leq C,$$

for any $\alpha > 0$.

Necessity of Anytime Capacity

Theorem: Necessity of Anytime Capacity

If there exists an observer / controller pair that achieves η-stability under bounded disturbance, then the channel's *feedback anytime capacity* satisfies

$$C_{any}(\eta \log_2 \lambda) \geq \log_2 \lambda,$$

Necessity of Anytime Capacity: Proof

Use the control system as a **black box** to construct a communication system with good anytime reliability. (sketch on white board)

1. Encoder sits with the plant; decoder with the controller.
2. Encode messages in the *disturbance*, W_t.
3. Controller must somehow know the disturbances, otherwise there is no way to stabilize the plant.
4. Decoder then reads off the *control actions* chosen by the controller to decode message.

Necessity of Anytime Capacity: Proof

Use the control system as a **black box** to construct a communication system with good anytime reliability. (sketch on white board)

1. Encoder sits with the plant; decoder with the controller.
2. Encode messages in the *disturbance*, W_t.
3. Controller must somehow know the disturbances, otherwise there is no way to stabilize the plant.
4. Decoder then reads off the *control actions* chosen by the controller to decode message.

Necessity of Anytime Capacity: Proof

Use the control system as a **black box** to construct a communication system with good anytime reliability. (sketch on white board)

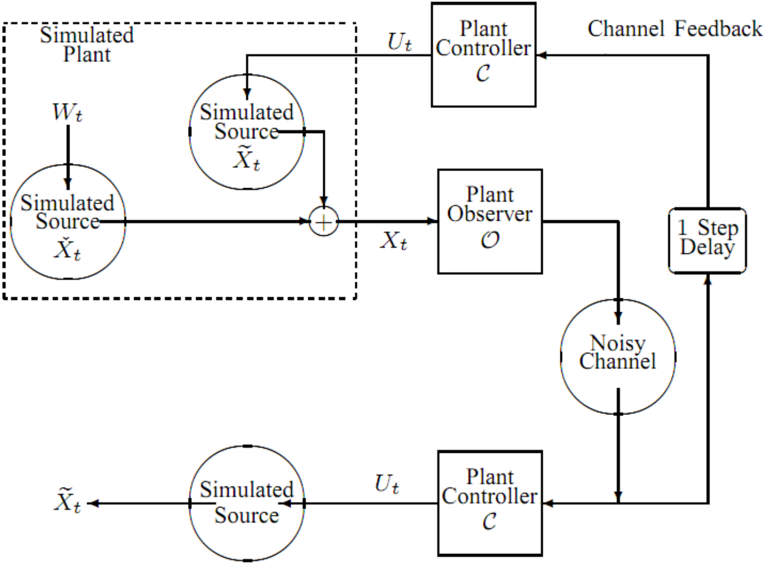
1. Encoder sits with the plant; decoder with the controller.
2. Encode messages in the *disturbance*, W_t.
3. Controller must somehow know the disturbances, otherwise there is no way to stabilize the plant.
4. Decoder then reads off the *control actions* chosen by the controller to decode message.

Necessity of Anytime Capacity: Proof

Use the control system as a **black box** to construct a communication system with good anytime reliability. (sketch on white board)

1. Encoder sits with the plant; decoder with the controller.
2. Encode messages in the *disturbance*, W_t.
3. Controller must somehow know the disturbances, otherwise there is no way to stabilize the plant.
4. Decoder then reads off the *control actions* chosen by the controller to decode message.

Necessity of Anytime Capacity: Proof



Necessity of Anytime Capacity: Proof

- But what do you mean by "knowing the disturbances"?
- Alright, let's be more concrete here. Write

$$X_t = Y_t + Z_t,$$

such that

$$X_0 = Y_0 = Z_0 = 0.$$

Necessity of Anytime Capacity: Proof

- But what do you mean by “knowing the disturbances”?
- Alright, let’s be more concrete here. Write

$$X_t = Y_t + Z_t,$$

such that

$$X_0 = Y_0 = Z_0 = 0.$$

Necessity of Anytime Capacity: Proof

- Y_t is the **control branch**

$$Y_{t+1} = \lambda Y_t + W_t.$$

- Z_t is the **disturbance branch**

$$Z_{t+1} = \lambda Z_t + U_t.$$

- Can easily verify by recursion

$$X_t = Y_t + Z_t.$$

Necessity of Anytime Capacity: Proof

- Key idea: encoder can control W_t (hence Y_t), while decoder knows Z_t perfectly.
- The plant is η-*stable*, so $|X_t|$ must be *small at all times*.
- Therefore, we must have

$$Y_t \approx -Z_t.$$

- Voila! Decoder should be able to extract good information of W_t by looking at Z_t.

Necessity of Anytime Capacity: Proof

- Key idea: encoder can control W_t (hence Y_t), while decoder knows Z_t perfectly.
- The plant is η-*stable*, so $|X_t|$ must be *small at all times*.
- Therefore, we must have

$$Y_t \approx -Z_t.$$

- Voila! Decoder should be able to extract good information of W_t by looking at Z_t.

Necessity of Anytime Capacity: Proof

- Key idea: encoder can control W_t (hence Y_t), while decoder knows Z_t perfectly.
- The plant is η-*stable*, so $|X_t|$ must be *small at all times*.
- Therefore, we must have

$$Y_t \approx -Z_t.$$

- Voila! Decoder should be able to extract good information of W_t by looking at Z_t.

Necessity of Anytime Capacity: Proof

- Key idea: encoder can control W_t (hence Y_t), while decoder knows Z_t perfectly.
- The plant is η-*stable*, so $|X_t|$ must be *small at all times*.
- Therefore, we must have

$$Y_t \approx -Z_t.$$

- Voila! Decoder should be able to extract good information of W_t by looking at Z_t.

Part 1: Encoding

- Now, down to business: step 1, encoding.
- Let each message M_t be a collection of R bits.
- Let $S_i \in \{-1, 1\}$ be the ith bit in the system.
- Write

$$\begin{aligned} Y_t &= \lambda Y_{t-1} + W_{t-1} \\ &= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j. \end{aligned}$$

Part 1: Encoding

- Now, down to business: step 1, encoding.
- Let each message M_t be a collection of R bits.
- Let $S_i \in \{-1, 1\}$ be the ith bit in the system.
- Write

$$\begin{aligned} Y_t &= \lambda Y_{t-1} + W_{t-1} \\ &= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j. \end{aligned}$$

Part 1: Encoding

- Now, down to business: step 1, encoding.
- Let each message M_t be a collection of R bits.
- Let $S_i \in \{-1, 1\}$ be the ith bit in the system.
- Write

$$\begin{aligned} Y_t &= \lambda Y_{t-1} + W_{t-1} \\ &= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j. \end{aligned}$$

Part 1: Encoding

- Now, down to business: step 1, encoding.
- Let each message M_t be a collection of R bits.
- Let $S_i \in \{-1, 1\}$ be the ith bit in the system.
- Write

$$
\begin{aligned}
Y_t &= \lambda Y_{t-1} + W_{t-1} \\
&= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j.
\end{aligned}
$$

Part 1: Encoding

- Encoding: choose W_t to be the value of the fractional representation of $\{S_i\}$, $\lfloor Rt \rfloor + 1 \leq i \leq \lfloor R(t+1) \rfloor$,
- In particular, set

$$W_t = \gamma \lambda^{t+1} \sum_{k=\lfloor Rt \rfloor + 1}^{\lfloor R(t+1) \rfloor} (2 + \epsilon_1)^{-k} S_k.$$

- Need the right constants to make things work

$$\epsilon_1 = 2^{\frac{\log_2 \lambda}{R}} - 2,$$

$$\gamma = \frac{\Omega}{2\lambda^{1+\frac{1}{R}}}.$$

Part 1: Encoding

- Encoding: choose W_t to be the value of the fractional representation of $\{S_i\}$, $\lfloor Rt \rfloor + 1 \leq i \leq \lfloor R(t+1) \rfloor$,
- In particular, set

$$W_t = \gamma \lambda^{t+1} \sum_{k=\lfloor Rt \rfloor + 1}^{\lfloor R(t+1) \rfloor} (2 + \epsilon_1)^{-k} S_k.$$

- Need the right constants to make things work

$$\epsilon_1 = 2^{\frac{\log_2 \lambda}{R}} - 2,$$

$$\gamma = \frac{\Omega}{2\lambda^{1+\frac{1}{R}}}.$$

Part 1: Encoding

- Encoding: choose W_t to be the value of the fractional representation of $\{S_i\}$, $\lfloor Rt \rfloor + 1 \leq i \leq \lfloor R(t+1) \rfloor$,
- In particular, set

$$W_t = \gamma \lambda^{t+1} \sum_{k=\lfloor Rt \rfloor+1}^{\lfloor R(t+1) \rfloor} (2 + \epsilon_1)^{-k} S_k.$$

- Need the right constants to make things work

$$\epsilon_1 = 2^{\frac{\log_2 \lambda}{R}} - 2,$$

$$\gamma = \frac{\Omega}{2\lambda^{1+\frac{1}{R}}}.$$

Part 2: Decoding

- Sent...fingers crossed...how much separation did we get?
- Main technical lemma:

Technical Lemma

Let $\hat{S}_i(t)$ be the estimate of bit S_i at time t. For all $0 \leq j \leq t$,

$$\left\{\omega \middle| \exists i \leq j, \hat{S}_i(t) \neq \hat{S}_i(t)\right\} \subset \left\{\omega \middle| |X_t| \geq \lambda^{t-\frac{j}{R}} \left(\frac{\gamma \epsilon_1}{1+\epsilon_1}\right)\right\}$$

Part 2: Decoding

- Sent...fingers crossed...how much separation did we get?
- Main technical lemma:

Technical Lemma

Let $\hat{S}_i(t)$ be the estimate of bit S_i at time t. For all $0 \leq j \leq t$,

$$\left\{\omega \middle| \exists i \leq j, \hat{S}_i(t) \neq \hat{S}_i(t)\right\} \subset \left\{\omega \middle| |X_t| \geq \lambda^{t-\frac{j}{R}} \left(\frac{\gamma\epsilon_1}{1+\epsilon_1}\right)\right\}$$

Part 2: Decoding

Technical Lemma

Let $\hat{S}_i(t)$ be the estimate of bit S_i at time t. For all $0 \leq j \leq t$,

$$\left\{\omega \middle| \exists i \leq j, \hat{S}_i(t) \neq \hat{S}_i(t)\right\} \subset \left\{\omega \middle| |X_t| \geq \lambda^{t-\frac{j}{R}} \left(\frac{\gamma\epsilon_1}{1+\epsilon_1}\right)\right\}$$

- Intuition: if early disturbances (S_t) were guessed incorrectly, control will blow up exponentially fast!
- Hence small $|X_t|$ must imply good estimates of early disturbances.

Part 2: Decoding

Proof:

- If two message differ in the first bit, S_1, how much will they differ on resulting Y_t?

$$
\begin{aligned}
&\inf_{\bar{S}:\bar{S}_1\neq S_1} \left|Y_1(S) - Y_1(\bar{S})\right| \\
\geq\ & \gamma\lambda^t \left(\left(1 - \sum_{k=1}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k}\right) - \left(-1 - \sum_{k=1}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k}\right)\right) \\
>\ & \gamma\lambda^t\, 2 \left(1 - \sum_{k=1}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-k}\right) \\
=\ & \lambda^t \left(\frac{2\epsilon_1\gamma}{1+\epsilon_1}\right)
\end{aligned}
$$

- Note the exponential dependence on λ (will force controller to report good estimates).

Part 2: Decoding

- More generally,

$$\inf_{\bar{S}:\bar{S}_i \neq S_i} \left|Y_t(S) - Y_t(\bar{S})\right| > \lambda^{t-\frac{i}{R}} \left(\frac{2\epsilon_1\gamma}{1+\epsilon_1}\right),$$

  if $i \leq \lfloor Rt \rfloor$.
- We will decode to get the $\hat{S}_i$ by pretending that $-Z_t$ is Y_t.
- Complete the proof of Lemma by noting that

$$\left|Z_t(S) - Z_t(\bar{S})\right| \geq \left|Y_t(S) - Y_t(\bar{S})\right| - \left|X_t(S) - X_t(\bar{S})\right|.$$

- In other words, smallness of X_t guarantees the closeness of $-Z_t$ and Y_t.

Part 3: Probability of Error

$$
\begin{aligned}
\mathbb{P}\left(|X_t| > m\right) &= \mathbb{P}\left(|X_t|^\eta > m^\eta\right) \\
&\leq \mathbb{E}\left(|X_t|^\eta\right) m^{-\eta} \\
&< K m^{-\eta} \text{ (definition of } \eta\text{-stability).}
\end{aligned}
$$

Combine this with the Technical Lemma

$$
\begin{aligned}
\mathbb{P}\left(\hat{S}_1^i(t) \neq S_1^i(t)\right) &\leq \mathbb{P}\left(|X_t| > \lambda^{t-\frac{j}{R}}\left(\frac{\gamma\epsilon_1}{1+\epsilon_1}\right)\right) \\
&< \left(K\left(\frac{1}{\gamma} + \frac{1}{\gamma\epsilon_1}\right)^\eta\right) 2^{-(\eta \log_2 \lambda)\left(t-\frac{i}{R}\right)}.
\end{aligned}
$$

This proves the theorem.

Review: we just proved...

Theorem: Necessity of Anytime Capacity

If there exists an observer / controller pair that achieves η-stability under bounded disturbance, then the channel's *feedback anytime capacity* satisfies

$$C_{any}(\eta \log_2 \lambda) \geq \log_2 \lambda.$$

- Sufficient conditions for anytime reliability in stabilizing a plant is in the paper, but will not be covered here.
- Is the necessary condition tight? Are there simpler ways to interpret / proof this result?

Outline

Conclusions

Concluding Remarks

- Thinking about the required information rate for particular application: may need different (or stronger) notion of capacity / reliability.
- Exponentially unstable nature of linear control system underlies the higher information barrier.
- Put in an adversarial way, *instability* and *noise* are both our enemies in communications. Any other major adversaries that we should consider?