Anytime Capacity of Stabilization of a Linear System over Noisy Channel

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

1. Introduction

2. A Counter Example

3. Necessity of Anytime Capacity

4. Conclusions
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”?
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”?
  - \textit{Shannon capacity} may not be adequate for stronger notions of stability. Need \textit{anytime capacity}.
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 \textit{unstable} even if the Shannon capacity of the channel is \textit{infinite}.
- A necessary condition for stability in terms of anytime capacity.
Main Reference

Outline

1. Introduction

2. A Counter Example

3. Necessity of Anytime Capacity

4. Conclusions
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:
- \textit{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: \( \eta \)-stability**

A closed-loop system is \( \eta \)-stable if there exists \( K < \infty \), such that

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

for all \( t \geq 0 \).

(More general notions of stability can be defined, but we will focus on \( \eta \)-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: $A = \mathbb{R}$.
- Output alphabet: $B = \mathbb{R}$.
- Transition probabilities

\[
\begin{align*}
p(x|x) &= 1 - \delta, \\
p(0|x) &= \delta.
\end{align*}
\]

I.e., a symbol is either received \textit{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.
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 $\eta$-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 $\eta$-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 $t \to \infty$. 
Lesson learned: notion of information depends on the strength of stability required (e.g., values of $\eta$).

- 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”).
Lesson learned: notion of information depends on the strength of stability required (e.g., values of $\eta$).

- 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”).
Lesson learned: notion of information depends on the strength of stability required (e.g., values of $\eta$).

- 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

1 Introduction

2 A Counter Example

3 Necessity of Anytime Capacity

4 Conclusions
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 $t$. 

---

**Communication System**

Anytime Reliability and Capacity
Anytime Reliability and Capacity

Anytime Reliability

A rate $R$ communication system achieves \textit{anytime reliability} $\alpha$ if there exists constant $K$ such that

$$\mathbb{P} \left( \hat{M}_1^i(t) \neq M_1^i \right) \leq K2^{-\alpha(t-i)}.$$  

The system is \textit{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

$\alpha$-anytime Capacity

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

More stringent than Shannon capacity, $C$:

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

for any $\alpha > 0$. 

Theorem: Necessity of Anytime Capacity

If there exists an observer / controller pair that achieves $\eta$-stability under bounded disturbance, then the channel’s feedback anytime capacity satisfies

$$C_{\text{any}}(\eta \log_2 \lambda) \geq \log_2 \lambda,$$
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 $\eta$-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 $\eta$-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 $\eta$-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 $\eta$-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 $Z_t$. 
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 $i$th bit in the system.

Write

$$Y_t = \lambda Y_{t-1} + W_{t-1}$$

$$= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j.$$
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 $i$th bit in the system.
- Write

\[
Y_t = \lambda Y_{t-1} + W_{t-1} = \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j.
\]
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 $i$th bit in the system.
- Write

\[
Y_t = \lambda Y_{t-1} + W_{t-1}
\]

\[
= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j.
\]
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 $i$th bit in the system.
- Write

$$Y_t = \lambda Y_{t-1} + W_{t-1}$$

$$= \lambda^{t-1} \sum_{j=0}^{t-1} \lambda^{-j} W_j.$$
Part 1: Encoding

- Encoding: choose $W_t$ to be the value of the fractional representation of $\{S_i\}$, $[Rt] + 1 \leq i \leq [R(t + 1)]$.

- In particular, set

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

- 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 $\{S_i\}$, $\lfloor R_t \rfloor + 1 \leq i \leq \lfloor R(t + 1) \rfloor$,

- In particular, set

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

- 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 $0 \leq j \leq t$,\[ \{ \omega | \exists i \leq j, \hat{S}_i(t) \neq \hat{S}_i(t) \} \subset \{ \omega | |X_t| \geq \lambda^{t-j} \left( \frac{\gamma \epsilon_1}{1 + \epsilon_1} \right) \} \]
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 $0 \leq j \leq t$,

$$\left\{ \omega | \exists i \leq j, \hat{S}_i(t) \neq \hat{S}_i(t) \right\} \subset \left\{ \omega | \left| X_t \right| \geq \lambda^{t-j} \left( \frac{\gamma \epsilon_1}{1 + \epsilon_1} \right) \right\}$$
Part 2: Decoding

Technical Lemma

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

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

- 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 messages differ in the first bit, $S_1$, how much will they differ on resulting $Y_t$?

\[
\inf_{\bar{S} : \bar{S}_1 \neq S_1} |Y_1(S) - Y_1(\bar{S})| \\
\geq \gamma \lambda^t \left( \left( 1 - \sum_{k=1}^{[Rt]} (2 + \epsilon_1)^{-k} \right) - \left( -1 - \sum_{k=1}^{[Rt]} (2 + \epsilon_1)^{-k} \right) \right) \\
> \gamma \lambda^t 2 \left( 1 - \sum_{k=1}^{[Rt]} (2 + \epsilon_1)^{-k} \right) \\
= \lambda^t \left( \frac{2\epsilon_1 \gamma}{1 + \epsilon_1} \right)
\]

- Note the exponential dependence on $\lambda$ (will force controller to report good estimates).
Part 2: Decoding

- More generally,

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

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

\[
|Z_t(S) - Z_t(\bar{S})| \geq |Y_t(S) - Y_t(\bar{S})| - |X_t(S) - X_t(\bar{S})|.
\]

- In other words, smallness of \( X_t \) guarantees the closeness of \(-Z_t\) and \( Y_t \).
Part 3: Probability of Error

\[ \mathbb{P} (|X_t| > m) = \mathbb{P} (|X_t|^\eta > m^\eta) \]
\[ \leq \mathbb{E} (|X_t|^\eta) m^{-\eta} \]
\[ < Km^{-\eta} \text{ (definition of } \eta\text{-stability).} \]

Combine this with the Technical Lemma

\[ \mathbb{P} \left( \hat{S}^i_1(t) \neq S^i_1(t) \right) \leq \mathbb{P} \left( |X_t| > \lambda^{t-\frac{j}{R}} \left( \frac{\gamma \epsilon_1}{1 + \epsilon_1} \right) \right) \]
\[ < \left( K \left( \frac{1}{\gamma} + \frac{1}{\gamma \epsilon_1} \right)^\eta \right) 2^{- (\eta \log_2 \lambda) (t - \frac{i}{R})}. \]

This proves the theorem.
Theorem: Necessity of Anytime Capacity

If there exists an observer / controller pair that achieves $\eta$-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

1 Introduction

2 A Counter Example

3 Necessity of Anytime Capacity

4 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?