Universal Anytime Codes: An approach to uncertain channels in control

paper by Stark Draper and Anant Sahai
presented by Sekhar Tatikonda

Wireless Foundations
Department of Electrical Engineering and Computer Sciences
UC Berkeley

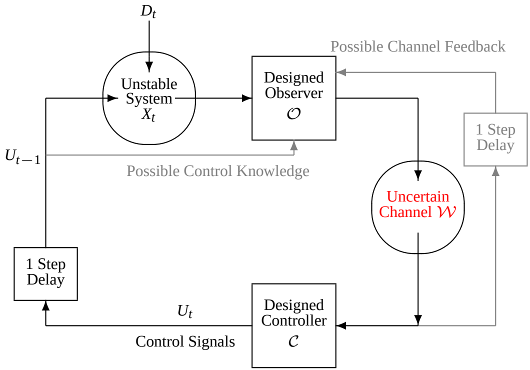
Mitsubishi Electric Research Labs
Cambridge, MA, USA

April 20, 2007

Outline

1. Problem setup
   - Channel uncertainty and stabilization
   - Review of past results
2. New result: universal anytime codes
3. Sufficient condition for stabilization
4. Conclusion

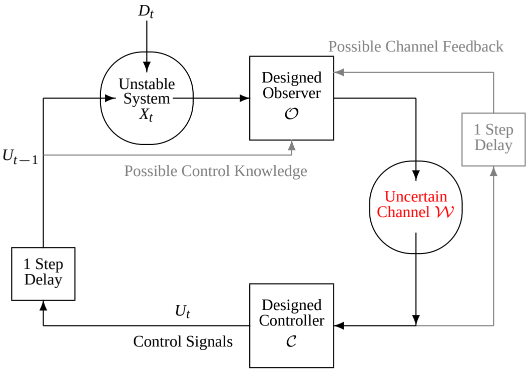
Our simple distributed control problem



$$X_{t+1} = \lambda X_t + U_t + D_t$$

- Unstable $\lambda > 1$, bounded initial condition and disturbance D.

Our simple distributed control problem



$$X_{t+1} = \lambda X_t + U_t + D_t$$

- Unstable $\lambda > 1$, bounded initial condition and disturbance D.
- Goal: η-stabilization $\sup_{t>0} E[\|X_t\|^\eta] \le K$ for some $K < \infty$.

Model of Channel Uncertainty: “Compound Channels”

- Channel W is known to be memoryless across time.
- Input and output alphabets are known $(\mathcal{Y}, \mathcal{Z})$ and finite.
- **Set-valued uncertainty** $\mathcal{W}$

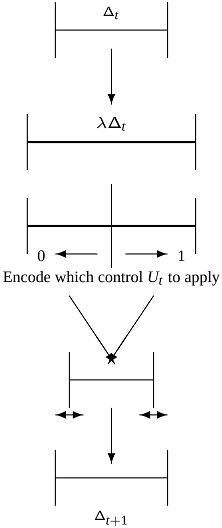
Model of Channel Uncertainty: "Compound Channels"

- Channel W is known to be memoryless across time.
- Input and output alphabets are known $(\mathcal{Y}, \mathcal{Z})$ and finite.
- **Set-valued uncertainty** $\mathcal{W}$
  - Nature can choose any particular $W \in \mathcal{W}$
  - Choice remains fixed for all time.

Model of Channel Uncertainty: "Compound Channels"

- Channel W is known to be memoryless across time.
- Input and output alphabets are known $(\mathcal{Y}, \mathcal{Z})$ and finite.
- **Set-valued uncertainty** $\mathcal{W}$
  - Nature can choose any particular $W \in \mathcal{W}$
  - Choice remains fixed for all time.
- Capacity well understood: $C = \sup_Q \inf_{W \in \mathcal{W}} I(Q, W)$

Review: Entirely noiseless channel



Window known to contain X_t

will grow by factor of $\lambda > 1$

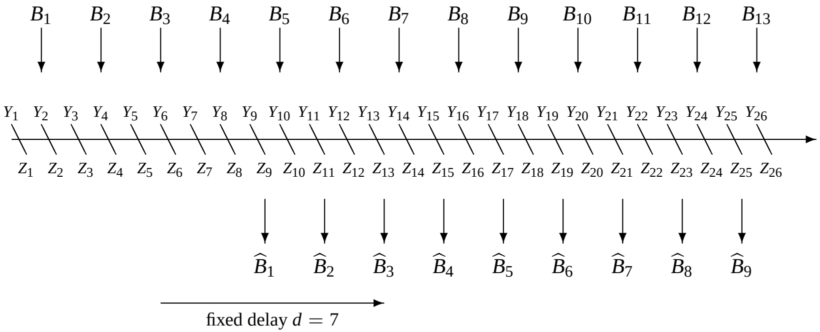
Sending R bits, cut window by a factor of 2^{-R}

grows by $\frac{\Omega}{2}$ on each side

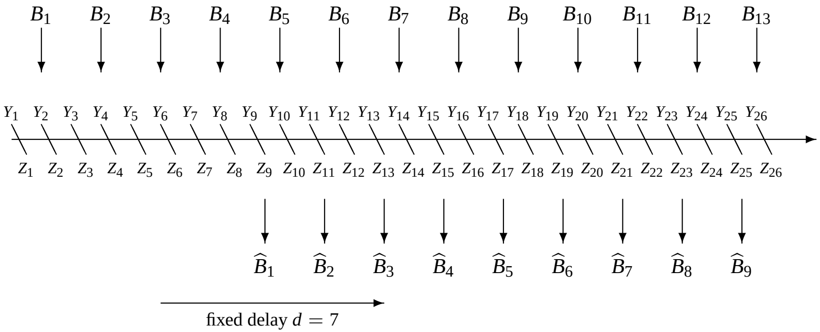
giving a new window for X_{t+1}

As long as $R > \log_2 \lambda$, we can have Δ stay bounded forever.

Review: Delay-universal (*anytime*) communication

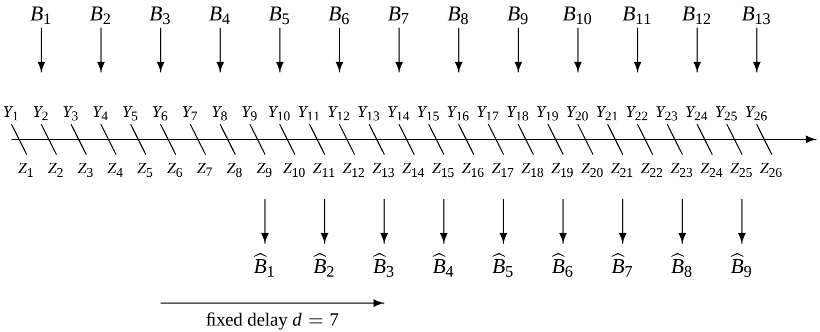


Review: Delay-universal (*anytime*) communication



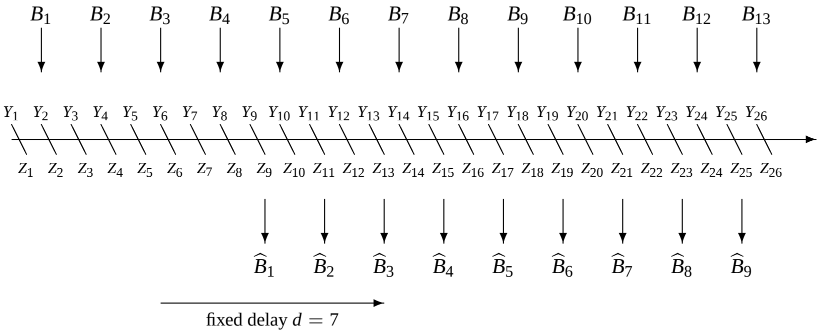
- Fixed-delay reliability α is achievable if there exists a sequence of encoder/decoder pairs with increasing end-to-end delays d_j such that $\lim_{j\to\infty} \frac{-1}{d_j} \ln P(B_i \neq \hat{B}_i^j) = \alpha$.

Review: Delay-universal (*anytime*) communication



- Reliability α is achievable *delay-universally* or in an *anytime fashion* if a single encoder works for all sufficiently large delays d.

Review: Delay-universal (*anytime*) communication



- The anytime capacity $C_{\text{any}}(\alpha)$ is the supremal rate at which reliability α is achievable in a delay-universal way.

Review: Separation theorem for scalar control

Necessity: If a scalar system with parameter $\lambda > 1$ can be stabilized with finite η-moment across a noisy channel, then the **channel with noiseless feedback** must have

$$C_{\text{any}}(\eta \ln \lambda) \geq \ln \lambda$$

In general: If $P(|X| > m) < f(m)$, then $\exists K : P_{\text{error}}(d) < f(K\lambda^d)$

Review: Separation theorem for scalar control

Necessity: If a scalar system with parameter $\lambda > 1$ can be stabilized with finite η-moment across a noisy channel, then the **channel with noiseless feedback** must have

$$C_{\text{any}}(\eta \ln \lambda) \geq \ln \lambda$$

In general: If $P(|X| > m) < f(m)$, then $\exists K : P_{\text{error}}(d) < f(K\lambda^d)$

Sufficiency: If there is an $\alpha > \eta \ln \lambda$ for which the **channel with noiseless feedback** has

$$C_{\text{any}}(\alpha) > \ln \lambda$$

then the scalar system with parameter $\lambda \geq 1$ with a bounded disturbance can be stabilized across the noisy channel with finite η-moment.

Review: Separation theorem for scalar control

Necessity: If a scalar system with parameter $\lambda > 1$ can be stabilized with finite η-moment across a noisy channel, then the **channel with noiseless feedback** must have

$$C_{\text{any}}(\eta \ln \lambda) \geq \ln \lambda$$

In general: If $P(|X| > m) < f(m)$, then $\exists K : P_{\text{error}}(d) < f(K\lambda^d)$

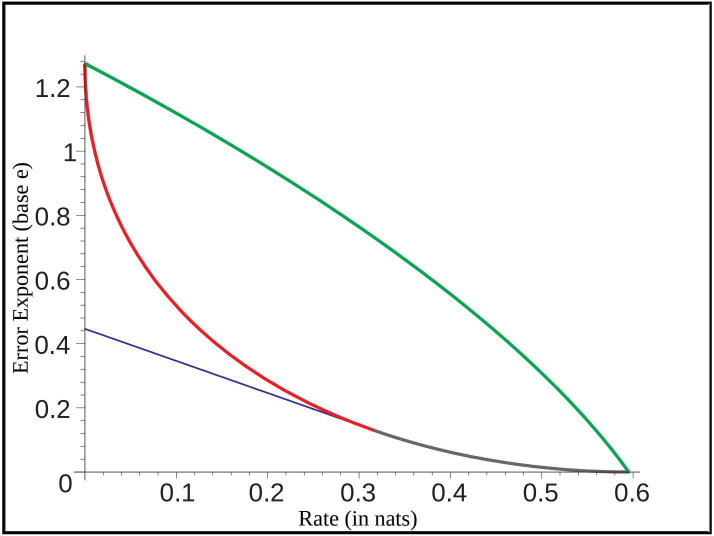
Sufficiency: If there is an $\alpha > \eta \ln \lambda$ for which the **channel with noiseless feedback** has

$$C_{\text{any}}(\alpha) > \ln \lambda$$

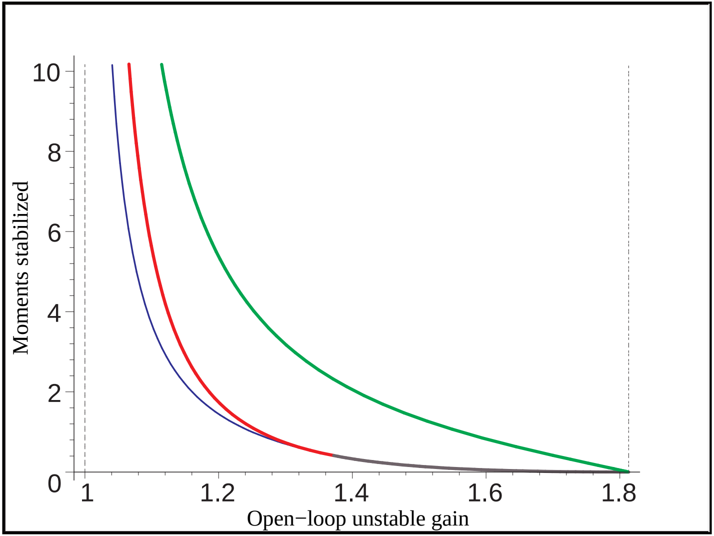
then the scalar system with parameter $\lambda \geq 1$ with a bounded disturbance can be stabilized across the noisy channel with finite η-moment.

Proved using a direct equivalence so it also holds for compound channels.

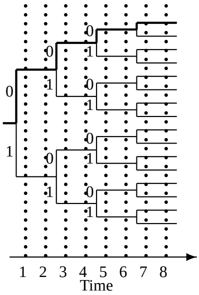
From Rate/Reliability to Gain/Moments



From Rate/Reliability to Gain/Moments



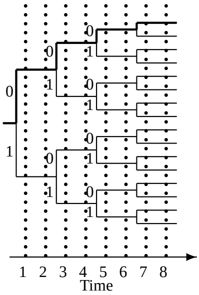
Why the random-coding bound works: tree codes



Tree with iid random labels:

- Data chooses a path through the tree
- Transmit the path labels through the channel
- Feedback is not used

Why the random-coding bound works: tree codes

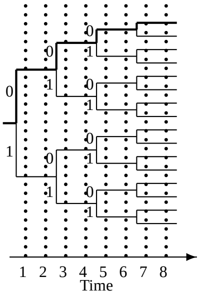


Tree with iid random labels:

- Data chooses a path through the tree
- Transmit the path labels through the channel
- Feedback is not used

- ML path decoding
  - **Log likelihoods add along path.**
  - Disjoint segments are pairwise independent of the true path.
  - $E_r(R)$ analysis applies at the suffix

Why the random-coding bound works: tree codes



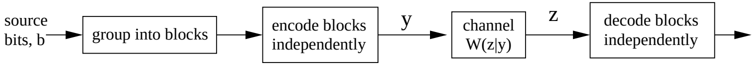
Tree with iid random labels:

- Data chooses a path through the tree
- Transmit the path labels through the channel
- Feedback is not used

- ML path decoding
  - **Log likelihoods add along path.**
  - Disjoint segments are pairwise independent of the true path.
  - $E_r(R)$ analysis applies at the suffix
  - Shortest suffix dominates
- **Achieves** $P_e(d) \leq K \exp(-E_r(R)d)$ **for every** d **for all** $R < C$

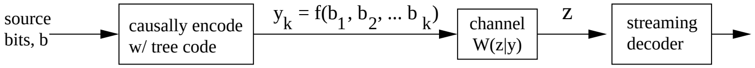
Relevant aspects of block and anytime codes

Block coding:



- Rare blocks in error
- Erroneous blocks never recovered
- Eventually results in exponential instability

Anytime coding:



- Can revisit earlier bit estimates
- Estimate reliabilities increase with delay
- Eventually detect and compensate for any incorrect controls

Outline

1. Problem setup
   - Channel uncertainty and stabilization
   - Review of past results without uncertainty
2. New result: universal anytime codes
3. Sufficient condition for stabilization
4. Conclusion

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?
  - The true codeword gives rise to an empirical channel that is like the true channel.

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?
  - The true codeword gives rise to an empirical channel that is like the true channel.
  - There are only a polynomial $(n+1)^{|\mathcal{Y}||\mathcal{Z}|}$ set of joint types.
  - The random-coding error exponent $E_r(R, Q, W)$ is achieved.

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?
  - The true codeword gives rise to an empirical channel that is like the true channel.
  - There are only a polynomial $(n+1)^{|\mathcal{Y}||\mathcal{Z}|}$ set of joint types.
  - The random-coding error exponent $E_r(R, Q, W)$ is achieved.
- What are the difficulties in generalizing to trees?

Review: universal block codes for compound channels

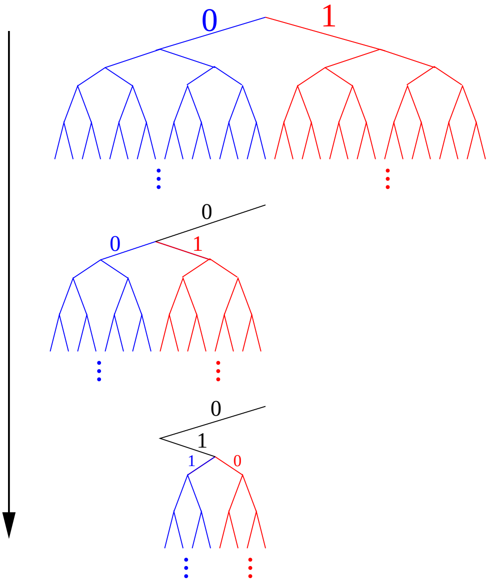
- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?
    - The true codeword gives rise to an empirical channel that is like the true channel.
    - There are only a polynomial $(n+1)^{|\mathcal{Y}||\mathcal{Z}|}$ set of joint types.
    - The random-coding error exponent $E_r(R, Q, W)$ is achieved.
- What are the difficulties in generalizing to trees?
    - The empirical mutual information (EMI) is **not additive**.
    - Ex: (0101,0101) + (1010,0101) gives $4 + 4 \neq 0$

Review: universal block codes for compound channels

- A single input distribution Q must be chosen for all $W \in \mathcal{W}$
- Look at the empirical mutual information (EMI) between $\vec{Z}$ and candidate codewords $\vec{Y}_m$
- Choose the one with the highest mutual information.
- Why does this work?
  - The true codeword gives rise to an empirical channel that is like the true channel.
  - There are only a polynomial $(n+1)^{|\mathcal{Y}||\mathcal{Z}|}$ set of joint types.
  - The random-coding error exponent $E_r(R, Q, W)$ is achieved.
- What are the difficulties in generalizing to trees?
  - The empirical mutual information (EMI) is **not additive**.
  - Ex: (0101,0101) + (1010,0101) gives $4 + 4 \neq 0$
  - The polynomial term grows with n not with delay d
  - Delays must be longer as time goes on.

Sequential Suffix-EMI Decoding

Decode sequentially, using max empirical mutual inform. (EMI) comparisons at each stage



If a blue codeword has the max EMI, decode first bit to 0, else to 1

If a blue codeword suffix has the max EMI, decode second bit to 0, else to 1

If a blue codeword suffix has the max EMI, decode third bit to 0, else to 1

The universal anytime coding theorem

Given a rate $R > 0$ and a compound discrete memoryless channel $\mathcal{W}$, there exists a random anytime code such that for all $E < E_{\text{any,univ}}(R)$ there is a constant $K > 0$ such that $\Pr[\widehat{B}_{n-d} \neq B_{n-d}] \leq K2^{-dE}$ for all $n, d \geq 0$ where

$$\begin{aligned} E_{\text{any,univ}}(R) &= \sup_{Q} \inf_{W \in \mathcal{W}} \inf_{P,V} D(P \times V \| Q \times W) + \max\{0, I(P,V) - R\} \\ &= \sup_{Q} \inf_{W \in \mathcal{W}} E_r(R, Q, W) \end{aligned}$$

- No feedback is needed
- Does as well as could be hoped for — essentially hits $E_r(R)$

Outline

Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$

Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$
  - Picking Q is unavoidable — determines codebook and capacity

Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$
  - Picking Q is unavoidable — determines codebook and capacity
  - But evaluating this bound can be difficult.

Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$
  - Picking Q is unavoidable — determines codebook and capacity
  - But evaluating this bound can be difficult.
- Is there a simpler condition that only depends on capacity $C(\mathcal{W})$?

Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$
  - Picking Q is unavoidable — determines codebook and capacity
  - But evaluating this bound can be difficult.
- Is there a simpler condition that only depends on capacity $C(\mathcal{W})$?
  - Gallager Exercise 5.23 tells us:
    $E_r(R, Q, W) \geq (I(Q, W) - R)^2 / (\frac{8}{e^2} + 4(\ln |\mathcal{Z}|)^2)$

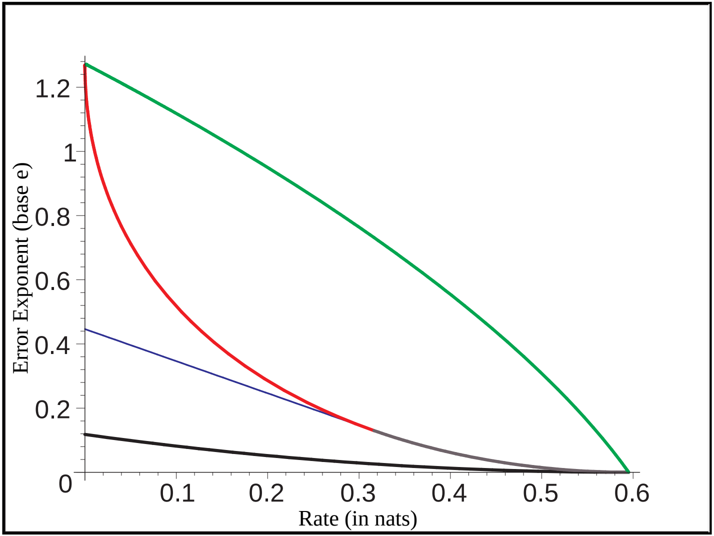
Back to the stabilization problem

- Can use $\sup_Q \inf_{W \in \mathcal{W}} E_r(R, Q, W)$ to evaluate compound channels $\mathcal{W}$
  - Picking Q is unavoidable — determines codebook and capacity
  - But evaluating this bound can be difficult.
- Is there a simpler condition that only depends on capacity $C(\mathcal{W})$?
  - Gallager Exercise 5.23 tells us:
    $E_r(R, Q, W) \geq (I(Q, W) - R)^2 / (\frac{8}{e^2} + 4(\ln |\mathcal{Z}|)^2)$
  - Translate to η-stabilization:

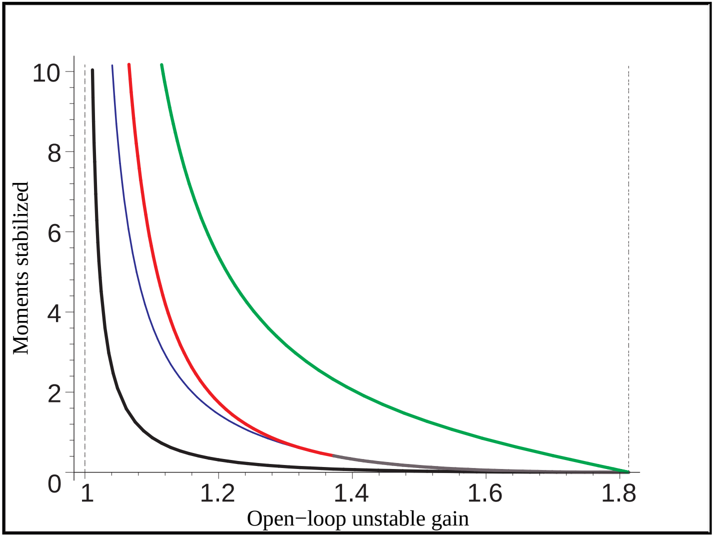
$$C(\mathcal{W}) - \log \lambda > (2 \log |\mathcal{Z}|) \sqrt{\frac{\eta \log \lambda}{\log e} (1 + \frac{2(\log e)^2}{e^2 (\log |\mathcal{Z}|)^2})}$$

  - Says that $O(\sqrt{\eta \log \lambda})$ extra capacity suffices to get η-th moment stabilized.

Visualizing the new universal sufficient condition



Visualizing the new universal sufficient condition



Conclusion

- Random **anytime codes** exist for compound channels.
- Can thus stabilize systems over uncertain channels.

Conclusion

- Random **anytime codes** exist for compound channels.
- Can thus stabilize systems over uncertain channels.
- Can operate with coarse description of channel uncertainty
  - Channel input distribution Q
  - Resulting capacity C
  - Set-size proxy: output alphabet size $|\mathcal{Z}|$.

Conclusion

- Random **anytime codes** exist for compound channels.
- Can thus stabilize systems over uncertain channels.
- Can operate with coarse description of channel uncertainty
  - Channel input distribution Q
  - Resulting capacity C
  - Set-size proxy: output alphabet size $|\mathcal{Z}|$.
- Performance loss for coarse channel uncertainty.
- Bounds should be improved.

Conclusion

- Random **anytime codes** exist for compound channels.
- Can thus stabilize systems over uncertain channels.
- Can operate with coarse description of channel uncertainty
  - Channel input distribution Q
  - Resulting capacity C
  - Set-size proxy: output alphabet size $|\mathcal{Z}|$.
- Performance loss for coarse channel uncertainty.
- Bounds should be improved.
- Feedback should be used.