Towards a Convex-Analytic View of Impossibility Results in Stochastic Control and Information Theory

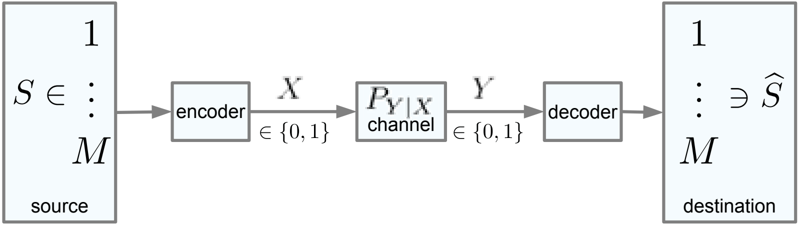
Ankur A. Kulkarni

Systems and Control Engineering
Indian Institute of Technology Bombay

November 5, 2016

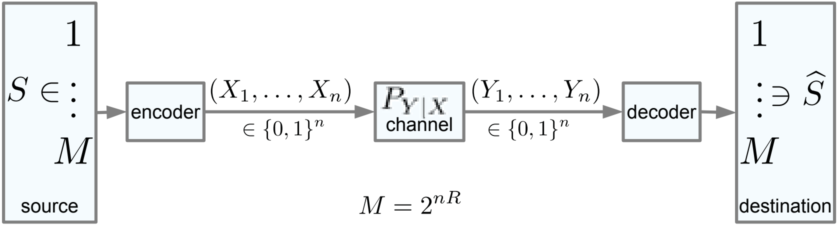
Shannon's noisy channel coding theorem (1944)

- Standard communication system



- Naive approach: naively encode and send each message multiple times, average out to decode
- Shannon's approach: block coding, encoding with redundancy and decoding so that chance of error becomes small

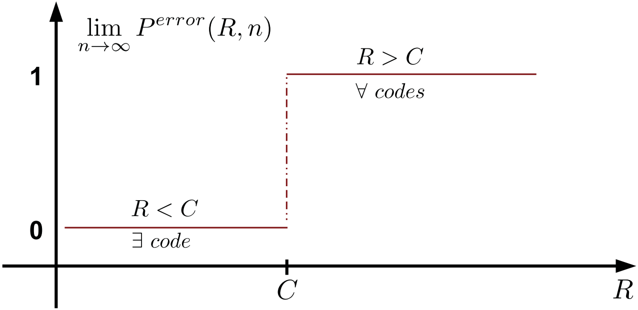
Shannon's noisy channel coding theorem (1944)



- Shannon considers the *block-coding* setup
- Shannon's channel coding theorem

$$P^{error}(code, R, n) = Prob(S \neq \widehat{S})$$

$$P^{error}(R, n) = \min_{codes} \; P^{error}(code, R, n)$$



-
- Channel capacity

$$C = \max_{p_X} I(X;Y)$$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

- Tremendous drop in complexity; “single letter characterization”

$$(2^n)^{2^{nR}} \times (2^{nR})^{2^n} \;\; codes$$

Open questions

- **Question:** What about finite blocklength performance?
- Discrete optimization. Extremely hard to compute as n gets large

$$P^{error}(R, n) = \min_{codes} \; P^{error}(code, R, n)$$

Open questions

- **Question:** What about finite blocklength performance?
- Discrete optimization. Extremely hard to compute as n gets large

$$P^{error}(R, n) = \min_{codes} \; P^{error}(code, R, n)$$

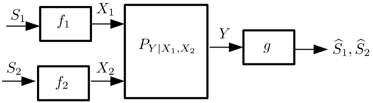
- **Question:** Why does this happen? And how does it transform to become tractable in the limit?

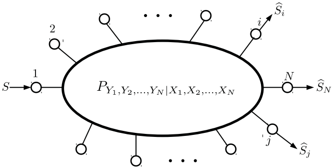
Open questions

- **Question:** What about finite blocklength performance?
- Discrete optimization. Extremely hard to compute as n gets large

$$P^{error}(R,n) = \min_{codes} \; P^{error}(code, R, n)$$

- **Question:** Why does this happen? And how does it transform to become tractable in the limit?
- **Question:** What about **networks**?



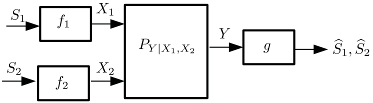


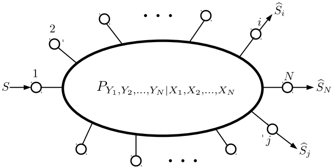
Open questions

- **Question:** What about finite blocklength performance?
- Discrete optimization. Extremely hard to compute as n gets large

$$P^{error}(R, n) = \min_{codes} \; P^{error}(code, R, n)$$

- **Question:** Why does this happen? And how does it transform to become tractable in the limit?
- **Question:** What about **networks**?





- Shannon invented a "calculus" of mutual information to prove theorems of information theory
- This calculus works well for information theoretic problems and fails miserably for other problems
- **Question:** Proofs and generalizations

Finite blocklength performance bounds

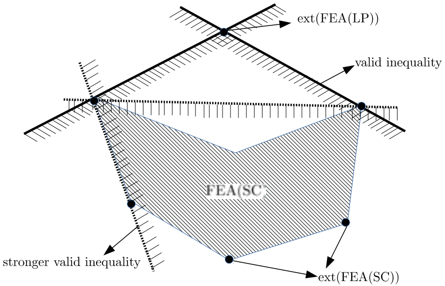
- Focus on lower bounds
- **Existing approach:** concentration inequalities, Gaussian approximations

Finite blocklength performance bounds

- Focus on lower bounds
- **Existing approach:** concentration inequalities, Gaussian approximations

Our approach

Convex analytic geometry of the finite blocklength problem.



[Jose and Kulkarni, 2015, IEEE CDC], [Jose and Kulkarni, 2016, IEEE Trans IT]

Original problem

$$\min_{codes} \; P^{error}(code, R, n)$$

$\Longleftrightarrow$

Optimization over distributions

(Nonconvex!)

$$\min_{distributions} \; P^{error}(distribution, R, n)$$

$\Downarrow$

Convex relaxation

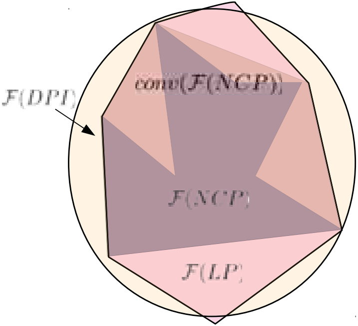
$$\min_{convex\ relaxation} \; P^{error}(dist, R, n)$$

$\Longleftarrow$

A particular LP relaxation

$$\min_{hyperplanes} \; P^{error}(dist, R, n)$$

Power of the LP relaxation



1. One can **systematically obtain** lower bounds

Infotheory problem	$\geq$	**value of LP relaxation**	$=$	**value of dual LP**	$\geq$	**value of any dual feasible point**

2. Gives a suite of lower bounds
3. Customizable, generalizable

Main results: point to point

Theorem (point to point)

1. *Asymptotically tight finite blocklength lower bounds for*
   - *Channel coding*
   - *Source coding*
   - *Joint source-channel coding*
2. *Improvement over previous finite blocklength bounds*

Main results: point to point

Theorem (point to point)

1. *Asymptotically tight finite blocklength lower bounds for*
   - *Channel coding*
   - *Source coding*
   - *Joint source-channel coding*
2. *Improvement over previous finite blocklength bounds*

Theorem

Consider a discrete memoryless binary symmetric channel with cross over prob ϵ*. For any code,*

$$Prob\{S \neq \widehat{S}\} \geq$$

$$\sup_{\epsilon > \delta > 0} \Big[1 + 2^{-n\left(H_2(\epsilon) - H_2(\epsilon-\delta) - \delta \log_2 \frac{1-\epsilon}{\epsilon} + \frac{1}{n} \log_2 l(n, \epsilon-\delta)\right)} - \frac{1}{M} 2^{n\left(1 - H_2(\epsilon) + \delta \log_2 \frac{1-\epsilon}{\epsilon}\right)}$$

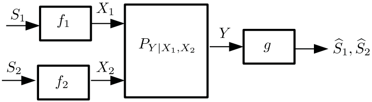
$$- (1-\epsilon)^{n - nH_{\frac{1}{1-\epsilon}}(\epsilon-\delta)} \Big],$$

$$\textit{where} \quad l(n, \alpha) = \frac{\exp(\lambda_1(n) - \lambda_2(n\alpha) - \lambda_2(n(1-\alpha)))}{\sqrt{2\pi\alpha(1-\alpha)n}}$$

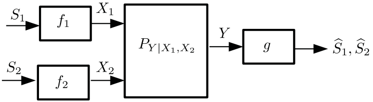
$$\textit{with} \quad \lambda_1(x) = \frac{1}{12x+1}, \quad \lambda_2(x) = \frac{1}{12x}, \quad \textit{and}$$

$$H_q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x).$$

Main results: networks



Main results: networks



Theorem (Converse for multiple access channel)

For any code,

$$\mathbb{P}[(S_1, S_2) \neq (\widehat{S}_1, \widehat{S}_2)] \geq \text{OPT(MAC)} \geq \text{OPT(DPMAC)}$$

$$\geq \sup_{\gamma>0, U, \bar{Y}} \max \Bigg(\sup_{a_1, a_2 : a_1 > a_2 \geq 0} \inf_{x_1, x_2} \mathbb{P}[a_1 i(x_1; \bar{Y}|x_2, U) + a_2 i(x_2; \bar{Y}|U) \leq a_1 \log M_1 + a_2 \log M_2 - \gamma] - \frac{\exp(-\gamma/a_1) \mid \mathcal{U} \mid}{\inf_{y, x_2, u : P_{\bar{Y}|X_2,U}(y|x_2,u)>0} P_{\bar{Y}|X_2,U}(y|x_2, u)^{(a_2/a_1)}},$$

$$\sup_{a_1, a_2 : a_2 > a_1 \geq 0} \inf_{x_1, x_2} \mathbb{P}[a_1 i(x_1; \bar{Y}|U) + a_2 i(x_2; \bar{Y}|x_1, U) \leq a_1 \log M_1 + a_2 \log M_2 - \gamma] - \frac{\exp(-\gamma/a_2) \mid \mathcal{U} \mid}{\inf_{y, x_1, u : P_{\bar{Y}|X_1,U}(y|x_1,u)>0} P_{\bar{Y}|X_1,U}(y|x_1, u)^{(a_1/a_2)}},$$

$$\sup_{a_1 : a_1 = a_2 > 0} \mathbb{P}[a_1 i(x_1, x_2; \bar{Y}|U) \leq a_1 \log M_1 M_2 - \gamma] - \mid \mathcal{U} \mid \exp(-\gamma/a_1) \Bigg),$$

Final remarks

- Classical stochastic control (Markov decision processes) has been written as optimization over distributions. Our results “unify” MDPs with stochastic control and information theory

Final remarks

- Classical stochastic control (Markov decision processes) has been written as optimization over distributions. Our results "unify" MDPs with stochastic control and information theory
- Resolve open issues about role of Shannon theory in stochastic decentralized control [**Kulkarni and Coleman, 2015, IEEE TAC**]

Final remarks

- Classical stochastic control (Markov decision processes) has been written as optimization over distributions. Our results "unify" MDPs with stochastic control and information theory
- Resolve open issues about role of Shannon theory in stochastic decentralized control [**Kulkarni and Coleman, 2015, IEEE TAC**]
- Nearly resolved a 60 year old open combinatorial problem posed by Levenshtein for the deletion channel [**Kulkarni and Kiyavash, 2013, IEEE Trans IT**], [**Kulkarni et. al, 2013, Discrete Mathematics**]

$$\frac{2^n}{n+1} \leq \textit{largest zero error code}$$

Final remarks

- Classical stochastic control (Markov decision processes) has been written as optimization over distributions. Our results "unify" MDPs with stochastic control and information theory
- Resolve open issues about role of Shannon theory in stochastic decentralized control [**Kulkarni and Coleman, 2015, IEEE TAC**]
- Nearly resolved a 60 year old open combinatorial problem posed by Levenshtein for the deletion channel [**Kulkarni and Kiyavash, 2013, IEEE Trans IT**], [**Kulkarni et. al, 2013, Discrete Mathematics**]

$$\frac{2^n}{n+1} \leq \textit{largest zero error code} \leq \frac{2^n}{n-1}$$

Final remarks

- Classical stochastic control (Markov decision processes) has been written as optimization over distributions. Our results "unify" MDPs with stochastic control and information theory
- Resolve open issues about role of Shannon theory in stochastic decentralized control **[Kulkarni and Coleman, 2015, IEEE TAC]**
- Nearly resolved a 60 year old open combinatorial problem posed by Levenshtein for the deletion channel **[Kulkarni and Kiyavash, 2013, IEEE Trans IT]**, **[Kulkarni et. al, 2013, Discrete Mathematics]**

$$\frac{2^n}{n+1} \leq \textit{largest zero error code} \leq \frac{2^n}{n-1}$$

Looking forward

- Convex analytic theory of fundamentally probabilistic results
- "Geometry of probability"

Acknowledgments

- DST
- IRCC, IIT Bombay

Acknowledgments

- DST
- IRCC, IIT Bombay

Thank you!