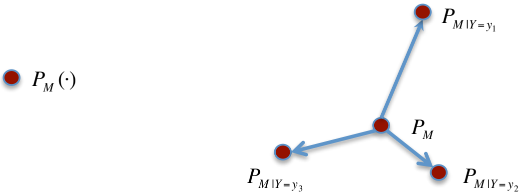
Instantaneously Efficiency of Dynamic Communications

Lizhong Zheng

ITMANET Program Review, Stanford, January 27, 2011

How to Describe Communication in One Time Instance



- Nothing is communicated reliably, so cannot say how many "bits" is transmitted.
- Mutual Information?
- Directions matters?

How is This Different From Achieving Capacity

- Capacity achieving random coding maximizes instantaneous mutual information.

$$I(M; Y^n) = \sum_t I(M; Y_t | Y^{t-1}) \to n \cdot \max_{P_X} I(X; Y)$$

How is This Different From Achieving Capacity

- Capacity achieving random coding maximizes instantaneous mutual information.

$$I(M;Y^n) = \sum_t I(M;Y_t|Y^{t-1}) \to n \cdot \max_{P_X} I(X;Y)$$

- Random coding: re-shaping input distribution at each time:

$$\begin{aligned} & \sum_m P_{M|\mathcal{H}}(m) \cdot D(P_{Y|M}(\cdot|m)||P_Y) \\ = & \sum_{x\in\mathcal{X}} \underbrace{\sum_{m:f(m)=x} P_{M|\mathcal{H}}(m)}_{P_X^*(x)} \cdot D(W_{Y|X}(\cdot|x)||P_Y) \end{aligned}$$

How is This Different From Achieving Capacity

- Capacity achieving random coding maximizes instantaneous mutual information.

$$I(M;Y^n) \quad = \quad \sum_t I(M;Y_t|Y^{t-1}) \rightarrow n \cdot \max_{P_X} I(X;Y)$$

- Random coding: re-shaping input distribution at each time:

$$\begin{aligned} & \sum_m P_{M|\mathcal{H}}(m) \cdot D(P_{Y|M}(\cdot|m)||P_Y) \\ = \quad & \sum_{x \in \mathcal{X}} \underbrace{\sum_{m:f(m)=x} P_{M|\mathcal{H}}(m)}_{P_X^*(x)} \cdot D(W_{Y|X}(\cdot|x)||P_Y) \end{aligned}$$

- Only an approximate solution, for a dominating fraction of the time.

Is Error Exponent a "True" Dynamic Metric

- Renyi entropy and divergence

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P^\alpha(x); \qquad D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P^\alpha(x) Q^{1-\alpha}(x)$$

- Mutual information and Decision making:

$$H_1 = \text{ Shannon entropy}; \qquad H_\infty(P) = -\log \max_x P(x)$$

- Renyi entropy and divergence

$$H_\alpha(P) = \frac{1}{1-\alpha}\log\sum_{x\in\mathcal{X}} P^\alpha(x); \qquad D_\alpha(P||Q) = \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} P^\alpha(x)Q^{1-\alpha}(x)$$

- Mutual information and Decision making:

$$H_1 = \text{ Shannon entropy; } \qquad H_\infty(P) = -\log\max_x P(x)$$

- Instantaneous optimization:

$$\max_{f:M\to\mathcal{X}}\sum_y P(y)\left(\sum_m P_M^{1-\alpha}(m)P_{M|Y}^\alpha(m|y)\right)^{\frac{1}{\alpha}}$$

$$\text{Random Coding} \quad \Rightarrow \quad \max_{P_X}\sum_y P(y)\left(\sum_x P_X^{1-\alpha}(x)P_{X|Y}^\alpha(x|y)\right)^{\frac{1}{\alpha}}$$

$$\text{Bayes Rule} \quad \Rightarrow \quad \max_{P_X}\sum_y \left(\sum_x P_X(x)W_{Y|X}^\alpha(y|x)\right)^{\frac{1}{\alpha}}$$

$$= \quad \max_{P_X} E_0(\rho, P_X) \quad \text{with } \rho = \frac{1}{\alpha} - 1$$

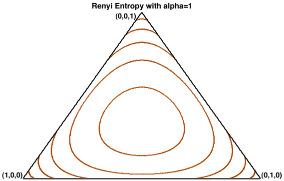
- In both capacity and error exponent optimization:
  - An instantaneous optimization is implicitly solved;
  - As $P_{M|\mathcal{H}}$ deviates from uniform, random coding deviates from the optimal solution;
  - With average-over-n performance metrics, the sub-optimality can be ignored.

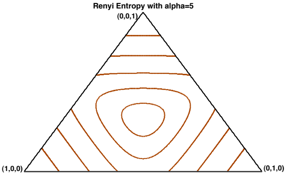
- In both capacity and error exponent optimization:
  - An instantaneous optimization is implicitly solved;
  - As $P_{M|\mathcal{H}}$ deviates from uniform, random coding deviates from the optimal solution;
  - With average-over-n performance metrics, the sub-optimality can be ignored.
- Communication optimized over every single time instance:
  - Only marginal gains in conventional metrics;
  - Better insights: finite time horizon, interference, soft information;
  - Easy implementation: greedy algorithms, approximate DP.

- There is no unique metric: the value of soft information depends on how it would be used.
  Example: $H([0.6, 0.2, 0.2]) \approx 1.38 > H([0.5, 0.5, 0]) = 1$

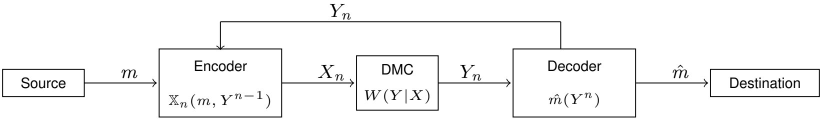
How to Measure Instantaneous Efficiency

- There is no unique metric: the value of soft information depends on how it would be used.
  Example: $H([0.6, 0.2, 0.2]) \approx 1.38 > H([0.5, 0.5, 0]) = 1$
- Renyi Divergence, α-divergence, etc..





Example 1: the Feedback Channel



- Encoder:

$$\mathbb{X}_t(m, Y^{t-1}) : \mathcal{M} \times \mathcal{Y}^{t-1} \rightarrow \mathcal{X} \qquad t \leq n$$

- Decoder:

$$\hat{m}(Y^n) : \qquad \mathcal{Y}^n \rightarrow \mathcal{M}$$

- Knowledge at time t, $\varphi_t(\cdot) = \mathbf{P}\left[m = \cdot \,|\, y^t\right]$
- Feedback does not increase channel capacity for DMC, but can improve error probability.

Instantaneous Formulation

- At each time t, the history is summarized in $\varphi_t(\cdot)$,

- At each time t, the history is summarized in $\varphi_t(\cdot)$,
- Define a score function

$$\zeta_n \triangleq \sum_k \varphi_n(k)^{1-\rho\eta}(\sum_{l \neq k} \varphi_n(l)^\eta)^\rho$$

  - Chosen from an exponential upper bound of P_e;
  - $\eta = \frac{1}{1+\rho}$ gives Renyi entropy;
  - In general, weighted sum of two Renyi entropies;

Instantaneous Formulation

- At each time t, the history is summarized in $\varphi_t(\cdot)$,
- Define a score function

$$\zeta_n \triangleq \sum_k \varphi_n(k)^{1-\rho\eta} (\sum_{l \neq k} \varphi_n(l)^{\eta})^{\rho}$$

  - Chosen from an exponential upper bound of P_e;
  - $\eta = \frac{1}{1+\rho}$ gives Renyi entropy;
  - In general, weighted sum of two Renyi entropies;
- Greedy optimization problem:

$$\max_{\mathbb{X}} \min_{t, y^t} E[\zeta_{t+1} | y^t]$$

- Solution: tilted posterior matching.

Extensions

- Posterior tilting: $P_{M|\mathcal{H}} \to P^{\eta}_{M|\mathcal{H}}$, slow down the process of decision making;
- Parameter η depends on the subject choice of the metric;
    - η can be optimized separately
    - but the assumption of same η for all t, y^t is artificial;

Extensions

- Posterior tilting: $P_{M|\mathcal{H}} \to P^{\eta}_{M|\mathcal{H}}$, slow down the process of decision making;
- Parameter η depends on the subject choice of the metric;
  - η can be optimized separately
  - but the assumption of same η for all t, y^t is artificial;
- Vary η based on t, y^t
  - Slow down if committed to a message too early, vice versa;
  - Balance progress for different outputs;

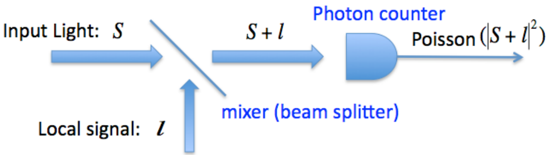
Extensions

- Posterior tilting: $P_{M|\mathcal{H}} \to P^{\eta}_{M|\mathcal{H}}$, slow down the process of decision making;
- Parameter η depends on the subject choice of the metric;
  - η can be optimized separately
  - but the assumption of same η for all t, y^t is artificial;
- Vary η based on t, y^t
  - Slow down if committed to a message too early, vice versa;
  - Balance progress for different outputs;
- General greedy instantaneous communication: at each time t
  - given a history: $\mathcal{H}_t$;
  - choose a metric: mutual information, Renyi divergence, etc.
  - design a few parameters: encoding, resource allocation, receiver designs, ...

Example: Quantum Detection

- Direct detection measure the intensity of light, resulting in Poisson channel, well studied;
- Theoretical optimization of coherent quantum detectors far more general than practical devices today;
- Realistic receivers can only use a few kinds of devices as building blocks

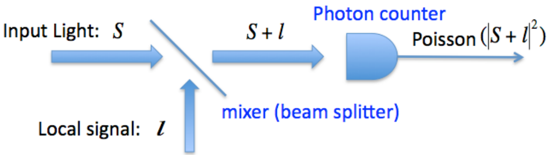
Example of Coherent Detector



- Design of local signal l changes the detection problem:

$$\begin{array}{lll} M = 0: & \text{Poisson}(|S_0|^2) & \multirow{2}{*}{$\longrightarrow$} \quad \text{Poisson}(\lambda_0 = |S_0 + l|^2) \\ M = 1: & \text{Poisson}(|S_1|^2) & \quad\quad\quad \text{Poisson}(\lambda_1 = |S_1 + l|^2) \end{array}$$

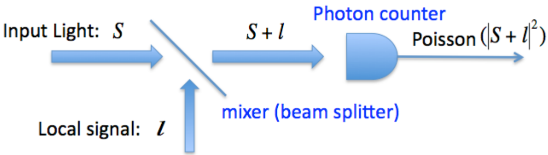
Example of Coherent Detector



- Design of local signal l changes the detection problem:

$$\begin{array}{lll} M = 0: & \text{Poisson}(|S_0|^2) & \\ & & \longrightarrow \\ M = 1: & \text{Poisson}(|S_1|^2) & \end{array} \quad \begin{array}{l} \text{Poisson}(\lambda_0 = |S_0 + l|^2) \\ \text{Poisson}(\lambda_1 = |S_1 + l|^2) \end{array}$$

- More general case requires a couple of more parameters;



- Design of local signal l changes the detection problem:

$$\begin{array}{lll} M=0: & \text{Poisson}(|S_0|^2) & \text{Poisson}(\lambda_0 = |S_0 + l|^2) \\ & & \longrightarrow \\ M=1: & \text{Poisson}(|S_1|^2) & \text{Poisson}(\lambda_1 = |S_1 + l|^2) \end{array}$$

- More general case requires a couple of more parameters;
- What is the sufficient statistic?

Instantaneous Design

- At time t, given the current knowledge $P_{M|\mathcal{H}_t}$, design $l(t)$ to be used for $[t, t+\Delta)$, to maximize mutual information;

Instantaneous Design

- At time t, given the current knowledge $P_{M|\mathcal{H}_t}$, design $l(t)$ to be used for $[t, t+\Delta)$, to maximize mutual information;
- Optimal performance for binary detection;
  - All metrics over 1-D space are equivalent;
  - Balanced progress;