Information Geometry and Capacity Limits

Lizhong Zheng

lizhong@mit.edu

Poster for ITMANET meeting

January, 2007

joint work with Shashi Borade, Emmanuel Abbe, Muriel Medard, Sean Meyn

Challenges in the Design of Structured Codes

– Generalize capacity region and scaling laws, parameterized by various system assumptions and constraints, characterized as a function of energy, delay, and outage

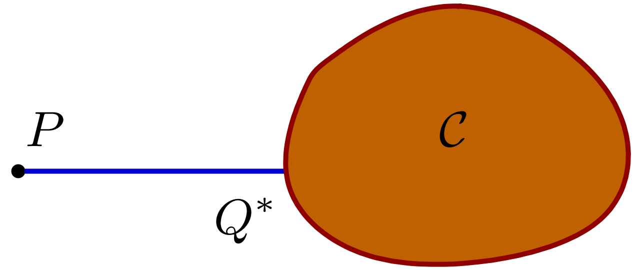
- Why is short and structured codes important in ITMANET?
  - Real-time applications, simple protocols, and wireless environments give sharp delay and reliability constraints;
  - Cooperative transmission requires early decoding (partial or soft information);
  - Network coding results suggest transmission of small pieces of information over the network;
  - Conjecture: joint source-channel-network codes require new performance metrics (eg. MDS) ;
- Why is this hard?
  - Shannon's results based on long block codes;
  - Randomization as the key step of information proofs;
  - Error exponents not well-understood:

$$E_{sp}(R) = \max_{P_x} \max_{\rho \geq 0} \{-\log(\sum_y (\sum_x P_{\mathbf{x}}(x) W_{\mathbf{y}|\mathbf{x}}(y|x)^{(1/1+\rho)})^{1+\rho}) - \rho R\}$$

Advantage of Geometric Approach

- i-Projection and Sanov's theorem:

  Exponential family characterizing typical error event



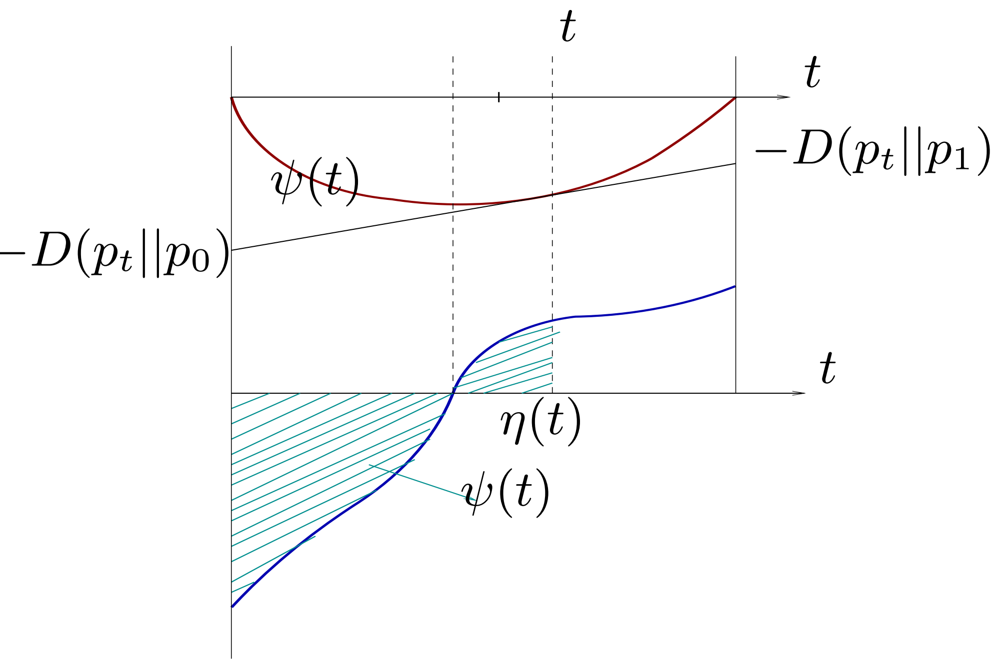
- Differential analysis allows layered coding structure with good reliability; eg. unequal error protection:

$$(M1, M2, ..., M_k) \Rightarrow X \rightarrow Y_1 \rightarrow Y_2 \rightarrow \ldots \rightarrow Y_k$$

- Relation to estimation related quantities allowing new encoder/decoder designs;

$$D(P_t||P_0) = \int\int \text{Fisher Information}$$

Useful Properties of Information Geometry

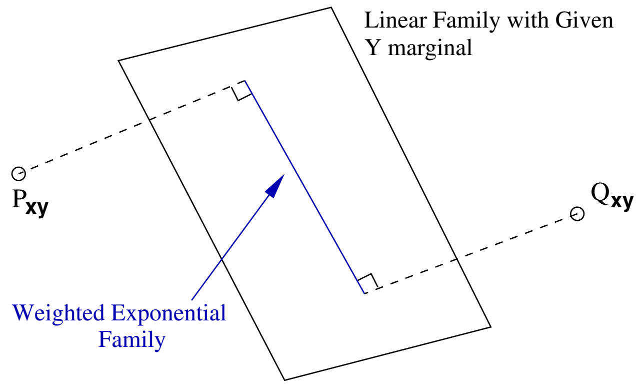


- Differential (local) Property:

$$\frac{\partial \psi}{\partial t} = \eta, \frac{\partial \phi}{\partial \eta} = t, \frac{\partial \eta}{\partial t} = g$$

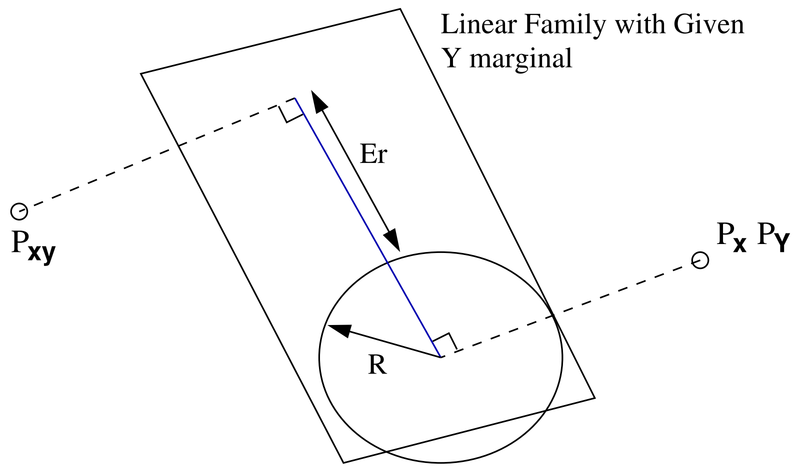
- Moment Generating Property.

- Marginalization:



Capacity and Error Exponents

$$E_r(R) = \min_t D(p_t||p_1) + [D(p_t||p_0) - R)]^+$$



- For $R > R_{crit}$, $+$ active,

$$E_r(R) = D(p_t||p_1), \quad \text{where } D(p_t||p_0) = R$$

- For $R < R_{crit}$, $+$ not active,

$$E_r(R) \quad = \quad \left[\min_t D(p_t||p_0) + D(p_t||p_1)\right] - R$$

Capacity with Mismatched Decoders

- Mismatched Receiver with decoding metric $F : \mathcal{X} \times \mathcal{Y} \rightarrow \Re$: pick the codeword with the largest value of

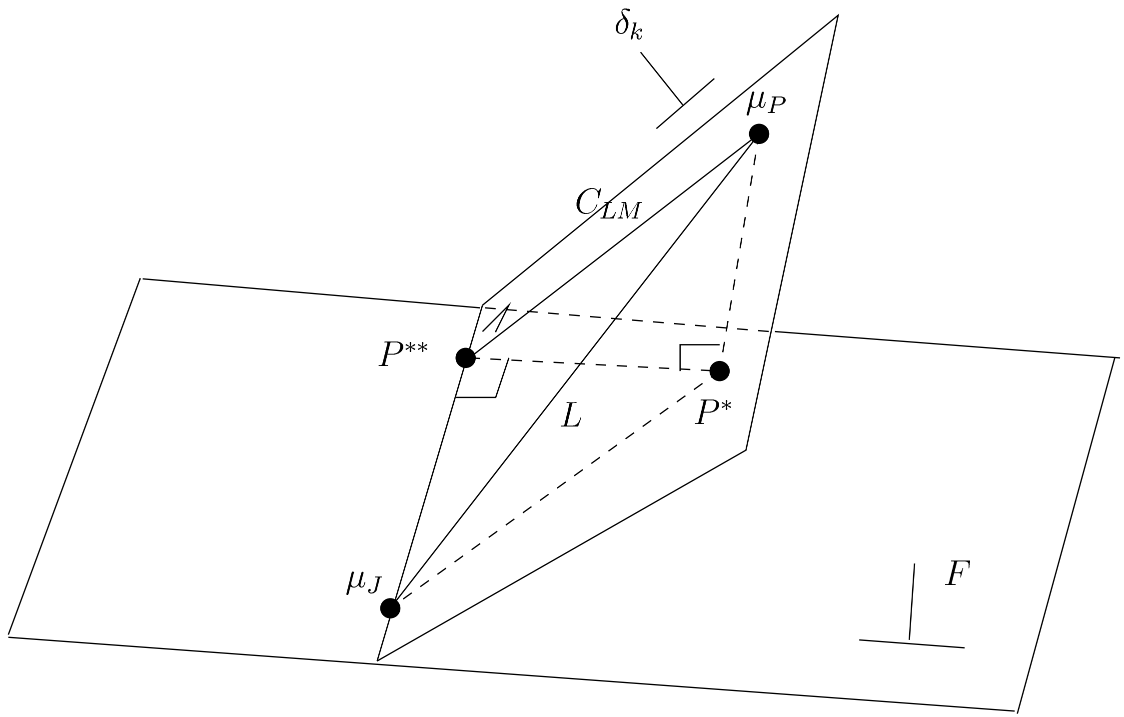
$$\sum_{i=1}^{n} F(x_i, y_i)$$

- ML decoder corresponds to $F(x,y) = \log \frac{P_{xy}(x,y)}{P_x(x)P_y(y)}$.
- Motivations
  - Wireless channel with non-perfect CSIR;
  - F maybe much simpler to evaluate than likelihood.
- Known results (Kaplan et.al.'94, Csiszar et.al.'95): optimization over input distributions difficult.
- Slightly different question: best mismatch out of a linear family of decoding metrics.

Geometric Solutions

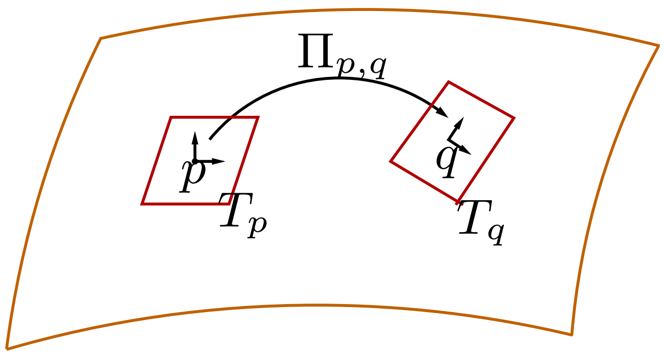
For given input distribution P_x, channel W,

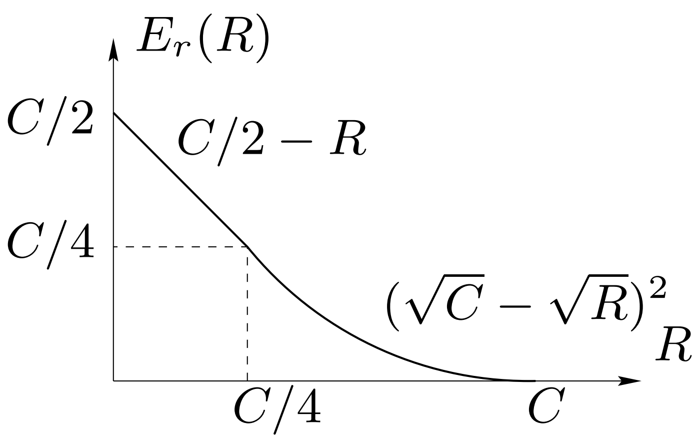
$$R = \min_{Q_{xy}:E_Q[F]=E_{\mu_J}[F],Q_y=P_y} D(Q||\mu_P)$$



.

Very Noisy Sitation



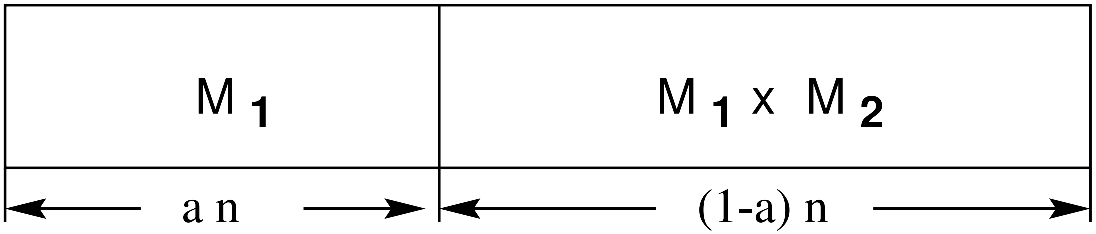


- Euclidean approximation of probability model
- Quadratic approximation of divergence, o^{th} order approximation of Fisher information

Mismatched capacity in very noisy channels: $C = \frac{1}{2}\frac{\langle \tilde{F}, L\rangle}{\|\tilde{F}\|^2}$

Layered Codes

- Two message tree code:



In very noisy case, at high rates,

$$
\begin{aligned}
\Pr(E_1) &= \exp\left[-n\left(\sqrt{C}-\sqrt{aR_1}-\sqrt{(1-a)R_2}\right)^2\right] \\
\Pr(E_2) &= \exp\left[-(1-a)n\left(\sqrt{C}-\sqrt{R_2/(1-a)}\right)^2\right]
\end{aligned}
$$

- Degraded broadcasting network (ISIT 07)