Linear Universal Decoder for Compound Channels

E. Abbe, L. Zheng

FLoWS

STATUS QUO

- *Maximum mutual information receiver is universal for compound channels, but complexity does not reduce when structured codes are used*
- *Lack of theoretical tools to analyze misinformed receivers*

NEW INSIGHTS

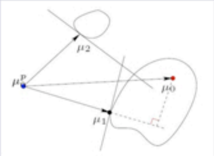
Linear receivers, with additive decoding metrics, naturally reuse existing decoders; and develop notion of projection in the space of prob. distributions

ACHIEVEMENT DESCRIPTION

MAIN RESULT:

- **Generalized MAP receivers achieve compound channel capacity; GLRT doesn't;**
- **Complete tradeoff between performance and complexity MMI;**

HOW IT WORKS:



- **Local approximation greatly simplifies the analysis, and serves as a new canonical example;**
- **Lifting to general cases;**
- **Universal receiver as decoding to all directions, and tradeoff possible when complexity limited.**

END-OF-PHASE GOAL

- Generalized geometric view of multi-terminal problems
- Practical designs of polytope receivers, with good tradeoff between performance and complexity

COMMUNITY CHALLENGE

The geometric view of multi-terminal problems gives new techniques in proving converse and providing insights to problems with high dimensionality and interactions among users

Generalized linear universal receiver achieves compound channel capacity

Why Do We Care About Geometry?

- Statistical coupling: controlling signals at one end, changing distributions at another;
- Typicality argument: static and point-to-point, relatively few distributions involved;
- Lessons from error exponents: variation in the space of distributions;
- Variation of distributions in multi-terminal problems;
- What tool do we have? Kullback-Leibler Divergence.

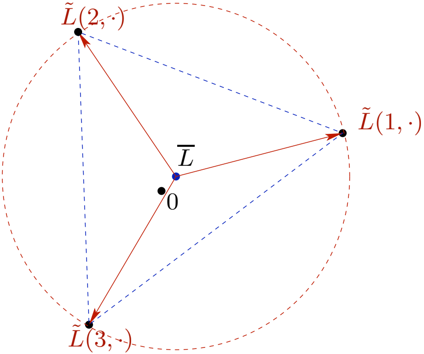
This CAN'T BE SUFFICIENT

We need notions of angle, inner product, projection, on the space of distributions.

- Why don't we look at distributions as points on the simplex?
- Amari: distributions as points on a manifold
  - local chart (tangent plane) $\longleftrightarrow$ Fisher metric;
  - length of geodesic $\longleftrightarrow$ K-L divergence;
- Over simplification: how about just look at the local picture?
  - How does it work?
  - When does this apply?
  - What new things can we say?
  - How about more general cases?
- Suppose $Q(x) = P(x)(1 + \epsilon \cdot L(x))$, with $\sum_x P(x)L(x) = 0$,

$$\begin{aligned} D(P||Q) &= -\sum_x P(x) \log \frac{Q(x)}{P(x)} = -\sum_x P(x) \log(1 + \epsilon \cdot L(x)) \\ &= -\epsilon \sum_x P(x)L(x) + o(\epsilon) \\ &= \underbrace{\frac{\epsilon^2}{2} \sum_x P(x)L^2(x)}_{\|L\|_P^2 \approx \|L\|_Q^2} + o(\epsilon^2) = D(Q||P) + o(\epsilon^2) \end{aligned}$$

For a given channel W, from $\mathcal{X} \to \mathcal{Y}$.



- Very Noisy Assumption

$$W_\epsilon(y|x) = P_0(y)(1 + \epsilon L(x,y))$$

- Given P_X, $P_Y = P_0(1 + \epsilon\overline{L})$

$$\overline{L}(y) = \sum_x P_X(x) L(x,y), \quad \forall y$$

- Definition
$\tilde{L}(x,y) = L(x,y) - \overline{L}(y), \quad \forall y$
- Mutual information

$$\begin{aligned} & I(P_X, W) \\ = \; & E_X\left[D(W(\cdot|x)||P_Y(\cdot))\right] \\ = \; & \frac{\epsilon^2}{2} E_X\left[\|\tilde{L}(x,\cdot)\|^2_{P_0}\right] + o(\epsilon^2) \end{aligned}$$

Compound Channel and Universal Receivers

- Compound channel: $W \in S$, S is known and compact.
- Goal: find one pair of encoder/decoder with reliable communication for any W in S.
- Capacity

$$C(S) = \max_{P_X} \inf_{W \in S} I(P_X, W)$$

  fixed composition random code works.
- For this talk, we fix P_X, and find efficient universal decoder.
- MMI decoder: maximize empirical mutual information, practical difficulty to implement.
- Linear Receivers:
  - Decoding metric: $d : \mathcal{X} \times \mathcal{Y} \to \mathcal{R}$
  - For each codeword $\underline{x}_m$, compute symbol-by-symbol sum

$$d^n(\underline{x}_m, \underline{y}) = \frac{1}{n} \sum_{i=1}^{n} d(x_m(i), y(i))$$

  and pick the largest.

Linear Receiver and Simplified Geometry

- Example: maximum likelihood decoder, $d(x, y) = \log W(y|x)$
- Linear in empirical distribution $\hat{P}_{(\underline{x}_m, \underline{y})}$

$$d^n(\underline{x}_m, \underline{y}) = E_{\hat{P}_{(\underline{x}_m, \underline{y})}}[d(X, Y)]$$

- Linear complexity? Think of LDPC and convolutional codes.
- Kaplan, Lapidoth, Shamai, Merhav'94, Csiszar, Narayan'95

$$R(P_X, W_0, d = \log W_1) = \inf_{\mu: E_\mu[\log W_1] = E_{\mu_0}[\log W_1]} D(\mu || \mu_0^p)$$

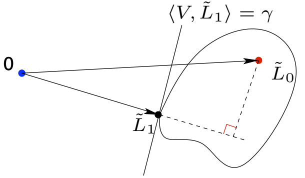
- Very noisy approximation: for $i = 0$(correct), 1(mismatched),

$$W_{i,\epsilon}(b|a) = P_Y(b)(1 + \epsilon L_i(a, b)), \quad \sum_{b \in \mathcal{Y}} L_i(a, b) P_Y(b) = 0, \forall a \in \mathcal{X}$$

- Scaled rates:

$$\lim_{\epsilon \to 0} \frac{2}{\epsilon^2} R(P_X, W_{0,\epsilon}, \log W_{1,\epsilon}) = \frac{\langle \tilde{L}_1, \tilde{L}_0 \rangle^2}{\|\tilde{L}_1\|^2}$$

When is Linear Receiver Sufficient?



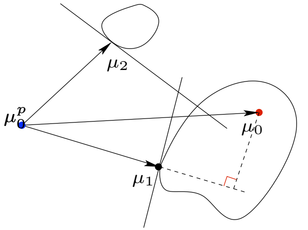
- Let L_1 correspond to the worst channel in S, If S is one-sided, i.e.

$$\frac{\langle \tilde{L}_1, \tilde{L}_0 \rangle}{\|\tilde{L}_1\|} \geq \|\tilde{L}_1\|, \quad \forall L_0$$

  decoding using $d = \log W_1$ achieves compound capacity;
- No requirement of convexity;
- Definition of General One-sided sets:

$$D(\mu_0||\mu_1^p) \geq D(\mu_0||\mu_1) - D(\mu_1||\mu_1^p), \quad \forall W_0 \in S, \mu_0 = P_X \cdot W_0$$

Generalized Linear Test: When the Compound Set is not One-Sided



- Generalized linear test: allow multiple metrics $d_1, \ldots, d_K$
- Decode to the largest of them all

$$\hat{m} = \arg\max_m \vee_{k=1}^K \sum_{i=1}^n d_k(x_m(i), y(i))$$

- Well known example: GLRT
- Conjecture: GLRT with the worst channel from each one-sided component achieve capacity?

Results from Geometric Analysis

- If S is a finite union K of one-sided sets, then GMAP test over the K worst channels

$$d_i = \log P^i_{X|Y}, \quad , i = 1, \ldots, K$$

  achieves capacity;
- GLRT over the worst channels is not universal: geometric intuition and counter example;
- MMI receiver can be viewed as GMAP over all channels;
- Polytope receiver, tradeoff between performance and complexity.