FROM MAXIMUM LIKELIHOOD TO ITERATIVE DECODING

F. Alberge, Z. Naja, P. Duhamel

Laboratoire des Signaux et Systèmes (L2S)

2011 International Conference on Acoustic, Speech and Signal processing

Introduction

Purpose:

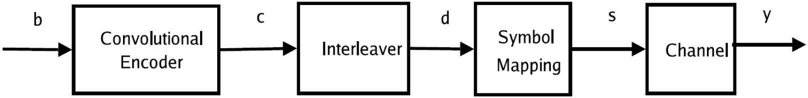
- Clarify the relation between Maximum Likelihood Sequence Detection and Iterative Decoding (BICM, Turbo,...)
- Derive Iterative Decoding as an optimization problem
- Obtain an evaluation of the reliability of the result

State of the art:

- Analysis of iterative decoding: EXIT charts, Density evolution [TenBrink2001], [Gamal2001].
  Useful for design but limited to large block length
- Convergence analysis: Factor Graphs, Belief Propagation [Kshishgang2001], [Pearl88].
  Useful if the corresponding graph is a tree
- Information geometry [Richardson2000], [Ikeda2004].
  Very important analysis but difficult to use for design or improvement of the iterative decoding
- First steps using optimization: [Walsh2006], [Alberge2008] ......

System model and Notations

BICM transmission scheme (results also apply to serially concatenated turbo-codes)



- **b**: binary message (vector of n_b bits)
- **c**: encoded bits (vector of n bits)
- **d**: interleaved encoded sequence (vector of n bits)
- **s**: complex transmitted sequence of symbols (vector of $\frac{n}{m}$ symbols)
- **y**: sequence of received symbols (vector of $\frac{n}{m}$ symbols) - Noisy memoryless channel

Maximum Likelihood Decoding

Maximum Likelihood Sequence Detection (MLSD):

$$\widehat{\mathbf{b}}_{MLD} = \arg \max_{\mathbf{b} \in \{0,1\}^{n_b}} p(\mathbf{y} \mid \mathbf{b})$$

One to one mapping between binary message $\mathbf{b}$ and interleaved coded sequence $\mathbf{d}$ $\Rightarrow$ MLSD reads:

$$\widehat{\mathbf{d}}_{MLD} = \arg \max_{\mathbf{d} \in \{0,1\}^{n}} \underbrace{p_{ch}(\mathbf{y} \mid \mathbf{d})}_{\substack{\text{channel} \\ \text{probability}}} \underbrace{I_{co}(\mathbf{d})}_{\substack{\text{indicator} \\ \text{function} \\ \text{of the code}}}$$

Equivalent to seeking optimal weighting for maximizing :

$$\textbf{(MLSD)} \quad \widehat{\mathbf{p}}_{MLD}(\mathbf{d}) = \arg \max_{\mathbf{p} \in \mathcal{E}_s} \sum_{\mathbf{d}} \mathbf{I}_{co}(\mathbf{d}) \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \mathbf{p}(\mathbf{d})$$

Two benefits : (i) $\mathcal{E}_s$: fully-factorized PMFs $\Rightarrow$ $\mathbf{p}(\mathbf{d}) = \prod_i p(d_i)$, (ii) $p(d_i)$ is continuous

Towards a suboptimal process (1/2)

(MLSD) is untractable: interleaver + numerical value of n

(MLSD) can be modified in the following manner:

- Consider separately channel/mapping and coding : $\mathbf{p}(\mathbf{d}) = \mathbf{l}(\mathbf{d})\mathbf{q}(\mathbf{d})$
- Compute bit-marginals (n variables instead of 2^n)

Bit-marginals computation can be introduced as (without any approx.):

$$\left(\widehat{\mathbf{l}}_{MLD}(\mathbf{d}), \widehat{\mathbf{q}}_{MLD}(\mathbf{d})\right) = \arg\max_{\mathbf{l},\mathbf{q}\in\mathcal{E}_s} \sum_{d_k} \sum_{\mathbf{d}:d_k} \mathbf{l}_{co}(\mathbf{d})\mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d})\mathbf{l}(\mathbf{d})\mathbf{q}(\mathbf{d})$$

Approximation: the bit-marginals of the product are replaced by the product of the bit-marginals.

Towards a suboptimal process (2/2)

Suboptimal **(MLSD)**: maximize $\mathcal{C}_k$ defined as:

$$\tilde{\mathcal{C}}_k(\mathbf{l},\mathbf{q}) = \sum_{d_k} \left(\sum_{\mathbf{d}:d_k} \mathbf{l}_{co}(\mathbf{d}) \prod_i q_i(d_i) \right) \left(\sum_{\mathbf{d}:d_k} \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_i l_i(d_i) \right)$$

Some comments on the suboptimal problem **(**$MLSD_{approx}$**)**:

- Computation of the bit-marginals is tractable: $\sum_{\mathbf{d}:d_k} \mathbf{l}_{co}(\mathbf{d}) \prod_i q_i(d_i)$ is the output of a BCJR
- $\mathcal{C}_k$ is a function of k (a bit position), but also depends on the other bits.

The original problem **(MLSD)** is replaced by a **distributed optimization strategy** based on the n cost functions $\mathcal{C}_k$.

$\mathcal{C}_k$ is relevant for a maximization over $l_k(d_k)q_k(d_k)$ the marginal of bit in position k. Nothing in this new formulation ensures consistency of the estimates (useful later)

Global maximum of $\mathbf{MLSD}_{approx}$: some results

Proposition

The maximum of $\tilde{\mathcal{C}}_k$, $1 \leq k \leq n$ is obtained for $\mathbf{q} = \widehat{\mathbf{q}}$ and $\mathbf{l} = \widehat{\mathbf{l}}$ such that

$$\widehat{\mathbf{l}}(\mathbf{d}')\widehat{\mathbf{q}}(\mathbf{d}) = \begin{cases} 1, & (\mathbf{d}, \mathbf{d}') = (\widehat{\mathbf{d}}_{co}, \widehat{\mathbf{d}}_{ch}) \\ 0, & \textit{otherwise} \end{cases} \quad (1)$$

where $(\widehat{\mathbf{d}}_{co}, \widehat{\mathbf{d}}_{ch}) = \arg\max_{(\mathbf{d},\mathbf{d}') \in \mathcal{S}_k} \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}')\mathbf{l}_{co}(\mathbf{d})$ and $\mathcal{S}_k$ denotes the set of pairs $(\mathbf{d}, \mathbf{d}')$ of binary words such that $d_k = d'_k$.

A separate maximization of $\tilde{\mathcal{C}}_k$ $\Rightarrow$ agreement between coder and mapping/channel for bit in position k.

Global maximum of MLSD$_{approx}$: some results

Now define the global criterion

$$\tilde{\mathcal{C}} = \sum_{k=1}^{n} \tilde{\mathcal{C}}_k$$

The value of the maximum of $\tilde{\mathcal{C}}$ $\Rightarrow$ an indication of the agreement between coder and mapping/channel for the whole sequence.

Proposition

Assume that $\tilde{\mathcal{C}}$ has a global maximum at $(\widehat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \widehat{\mathbf{q}}_{\tilde{\mathcal{C}}})$. If $(\hat{\mathbf{l}}_{\tilde{\mathcal{C}}}, \widehat{\mathbf{q}}_{\tilde{\mathcal{C}}})$ is such that $\widehat{\mathbf{l}_{\tilde{\mathcal{C}}}\hat{\mathbf{q}}_{\tilde{\mathcal{C}}}}(\mathbf{d}) = \delta_{\mathbf{d}_0}$ at $\mathbf{d} = \mathbf{d}_0$ then $\mathbf{d}_0 = \widehat{\mathbf{d}}_{MLD}$.

If the global maximum of $\widetilde{\mathcal{C}}$ is a Delta-Kronecker PMF $\Rightarrow$ MLSD (high SNR)

Local maximization process

A distributed maximization strategy:

$$\left(\widehat{l}_k, \widehat{q}_k\right) = \arg \max_{l_k, q_k \in \mathcal{F}} \tilde{\mathcal{C}}_k \quad 1 \leq k \leq n$$

$$\left(\widehat{l}_k, \widehat{q}_k\right) = \arg \max_{l_k, q_k \in \mathcal{F}} \sum_{d_k} l_k(d_k) q_k(d_k) \underbrace{\left(\sum_{\mathbf{d}:d_k} \mathbf{I}_{co}(\mathbf{d}) \prod_{i \neq k} q_i(d_i)\right)\left(\sum_{\mathbf{d}:d_k} \mathbf{p}_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_{i \neq k} l_i(d_i)\right)}_{x_{-k}(d_k)}$$

$\mathcal{F}$: set of all possible PMFs on d_k.

- hard solution (local minima?):

$$\begin{aligned} \widehat{l}_k(d_k)\hat{q}_k(d_k) &= 1 \quad \textit{if} \;\; x_{-k}(d_k) > x_{-k}(1-d_k) \\ &= 0 \quad \textit{otherwise} \end{aligned}$$

- soft solution (preferred):

$$\widehat{l}_k(d_k)\widehat{q}_k(d_k) \propto x_{-k}(d_k)$$

Corresponding Iterative Maximization

- Initialization:

$$l_k^{(0)}(dk) = q_k^{(0)}(d_k) = 1/2 \quad 1 \leq k \leq n \quad d = k \in \{0, 1\}$$

- Repeat
  - Set $l_k(d_k) = l_k^{(it-1)}(d_k)$, $1 \leq k \leq n$ and $q_i(d_i) = q_i^{(it-1)}(d_i)$ for $i \neq k$ (Jacobi implementation)
  - Compute $q_k^{(it)}$ based on soft solution:

$$q_k^{(it)}(d_k) \propto \sum_{\mathbf{d}:d_k} l_{co}(\mathbf{d}) \sum_{\mathbf{d}:d_k} p_{ch}(\mathbf{y} \mid \mathbf{d}) \prod_{j \neq k} l_j^{(it-1)}(d_j)$$

  - Set $l_i(d_i) = l_i^{(it-1)}(d_i)$ for $i \neq k$ and $q_k(d_k) = q_k^{(it)}(d_k)$ for $1 \leq k \leq n$ (Jacobi/Gauss-Seidel implementation)
  - Compute $l_k^{(it)}$ based on soft solution:

$$l_k^{(it)}(d_k) \propto \frac{\sum_{\mathbf{d}:d_k} l_{co}(\mathbf{d}) \prod_{j \neq k} q_j^{(it)}(d_j)}{\sum_{\mathbf{d}:d_k} l_{co}(\mathbf{d})}$$

l_k, q_k are the EXTRINSICS propagated in BICM-ID

From Maximum Likelihood to iterative decoding: summary

An optimal optimization problem: *MLSD*

$\downarrow$

Approximation: fully-factorized PMFs

A sub-optimal (global) optimization problem: $MLSD_{approx}$

global maximum in MLSD $\stackrel{?}{=}$ *global maximum in* $MLSD_{approx}$

$\downarrow$

Distributed optimization strategy: the actual BICM-ID algorithm

n sub-optimal (local) optimization problems ($\mathcal{C}_k$)

Efficiency of the joint optimization problem? $\Rightarrow$ *value of* $\widetilde{\mathcal{C}} = \sum_{k=1}^{n} \widetilde{\mathcal{C}}_k$

$\downarrow$

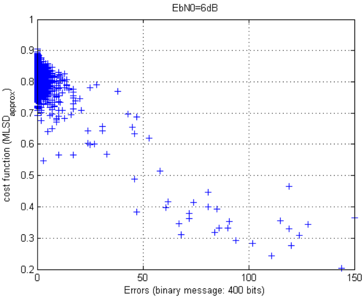
Evaluation of the quality of the obtained solution

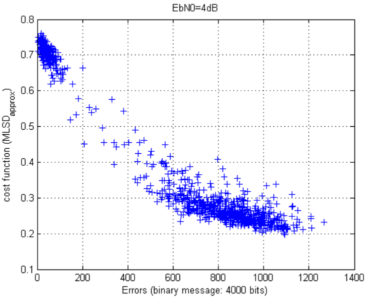
BICM iterative decoding

Convergence? (nonlinear Gauss-Seidel/Jacobi)

Simulation (1/2)

- $n_b = 400$ (left) $n_b = 4000$ (right)
- $EbN0 = 6dB$ (left) $EbN0 = 4dB$ (right)
- Modulation: 16QAM
- Mapping: SP
- Convolutive Code: [5 7]





$\Rightarrow$ Correlation between value of $\widetilde{C}$ and number of errors

Simulation (2/2)

- Modulation: 16QAM
- Mapping: SP
- Convolutive Code: [5 7]
- $Eb/N0 \in \{4dB, 5dB, ..., 11dB, 12dB\}$ (uniform distribution)

Threshold on $\widetilde{\mathcal{C}}$ (log)	-20	-10	-5
BER_a *(frames above treshold)*	$8,78.10^{-4}$	$4,68.10^{-4}$	$2,08.10^{-4}$
BER_s *(frames under treshold)*	$0,205$	$0,13$	$9,28.10^{-2}$
p_s *(rejected frames) (%)*	$6,4\%$	$10,8\%$	$14,8\%$
$p_{false,s}$ *(false alarm) (%)*	$2,5\%$	$36,4\%$	$53,83\%$

$\Rightarrow$ **A target BER can be guaranteed even in an unsteady noisy environment**

Conclusion

- Iterative Decoding derived from Maximum Likelihood
- No specific assumption (block length, tree, ...)
- Extrinsics proceed from an hybrid Jacobi/Gauss-Seidel scheduling
- Convergence study connected with nonlinear Jacobi/Gauss-Seidel (submitted to EUSIPCO 2011)
- Efficiency of the distributed optimization process: checkable at the receiver side through evaluation of the criterion