Local-Optimality Guarantees for Optimal Decoding Based on Paths

Nissim Halabi    Guy Even

School of Electrical Engineering, Tel-Aviv University

August 29, 2012
- **MBIOS channel**: memoryless, binary-input, output-symmetric
- Log-Likelihood-Ratio (LLR):
  \[
  \lambda_i(y_i) \triangleq \ln \left( \frac{\Pr(y_i \mid c_i = 0)}{\Pr(y_i \mid c_i = 1)} \right)
  \]
- Linear Code: \( C \subseteq \{0, 1\}^N \) is subspace of \( \mathbb{F}_2^N \) of dimension \( k \).
- **Optimal decoding**: Maximum Likelihood decoding. Input: \( y \). Output: \( \text{ML}(y) \).

\[
\text{ML}(y) \triangleq \arg \max_{x \in C} \Pr\{y \mid c = x\} = \arg \min_{x \in C} \langle \lambda(y), x \rangle
\]
Tanner Codes Defined by Tanner Graphs

\[ G = (\mathcal{V} \cup \mathcal{J}, E) \]

- Tanner code \( C(G, C^\mathcal{J}) \) represented by bipartite graph
- \( x \in C(G, C^\mathcal{J}) \) iff \( x \in C^j \) for every \( j \in \{1, \ldots, J\} \)

In general:
- **degrees:** can be regular, irregular, bounded, or arbitrary
- can allow arbitrary linear local codes

Examples: LDPC codes [Gallager’63], Expander codes [Sipser-Spielman’96]
Linear Programming (LP) Decoding

- \( \text{conv} (X) \subseteq \mathbb{R}^N \) - the convex hull a set of points \( X \subseteq \mathbb{R}^N \).
- **ML-decoding** can be rephrased:

  \[
  \text{ML}(y) \triangleq \arg \min_{x \in \text{conv}(C)} \langle \lambda(y), x \rangle
  \]

- **Generalized fundamental polytope** of a Tanner code \( C(G, C^J) \)
  - relaxation of \( \text{conv}(C) \) [following Feldman-Wainwright-Karger’05]

  \[
  \mathcal{P}(G, C^J) \triangleq \bigcap_{C^j \in C^J} \text{conv}(C^j)
  \]

- **LP-decoding**:

  \[
  \text{LP}(y) \triangleq \arg \min_{x \in \mathcal{P}(G, C^J)} \langle \lambda(y), x \rangle
  \]
LP Decoding with ML Certificate

```
LP-decode(\lambda)

solve LP: \hat{x}^{LP} \leftarrow \arg \min_{x \in \mathcal{P}(G,CJ)} \langle \lambda, x \rangle.

if \hat{x}^{LP} \in \{0, 1\}^N then
    return \hat{x}^{LP} is an ML codeword
else
    return fail
end if
```

- Polynomial time algorithm
- Applies to any MBIOS channel!
- Integral solution $\Rightarrow$ ML-certificate
Goal: Analysis of Finite Length Codes

Problem (Finite Length Analysis)

Design: Constant rate code $C(G, C^J)$ and an efficient decoding algorithm DEC.

Analyze: If $SNR > t$, then

$$Pr(DEC(\lambda) \neq x | c = x) \leq \exp(-N^{\alpha})$$

for some $0 < \alpha$.

Goal: Minimize $t$ (lower bound on $SNR$).

Remarks:

- Not an asymptotic problem
- Code is not chosen randomly from an ensemble
- Successful decoding $\neq$ ML decoding
Advances in analysis of finite-length codes via local-optimality: [Koetter-Vontobel’06], [Arora-Daskalakis-Steurer’09], [H-Even’10], [Vontobel’10], [H-Even’11]

Today

- Based on complicated combinatorial structures embedded in the Tanner graph of the codes and non-trivial analyses of random processes
- Demonstrate the proof technique - use simple characterization of local-optimality based on paths
- Simpler proofs obtained via local-optimality based on paths for the case of repeat-accumulate codes [Feldman-Karger’02], [Goldenberg-Burshtein’11]
**Problem (Optimality Certificate)**

**Input:** Channel observation $\lambda$ and a codeword $x \in C$

**Question 1:** Is $x$ ML-optimal with respect to $\lambda$? is it unique? *(NP-Hard)*

**Question 2:** Is $x$ LP-optimal with respect to $\lambda$? is it unique?

Relax: one-sided error test

- A positive answer = certificate for the optimality of $x$ w.r.t. $\lambda$
- A negative answer = don’t know if optimal or not (allow one sided error)

- “Local-Optimality” criterion: efficient test via local computations
Definition of Local-Optimality

- [Feldman’03] For $x \in \{0, 1\}^N$ and $f \in [0, 1]^N \subseteq \mathbb{R}^N$, define the relative point $x \oplus f$ by $(x \oplus f)_i \triangleq |x_i - f_i|$
- Consider a finite set of “deviations” $\triangleq B \subset [0, 1]^N$

Definition (following [Arora-Daskalakis-Steurer’09])

A codeword $x \in C$ is **locally-optimal** w.r.t. $\lambda \in \mathbb{R}^N$ if for all vectors $\beta \in B$,

$$\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle$$

Goal

Find a set $B$ of locally-structured deviations such that:

1. $x \in \text{LO}(\lambda) \Rightarrow x = \text{ML}(\lambda)$ & unique
2. $x \in \text{LO}(\lambda) \Rightarrow x = \text{LP}(\lambda)$ & unique
3. $\Pr\lambda\{x \in \text{LO}(\lambda) \mid c = x\} = 1 - o(1)$
Even Tanner Codes

Definition (Even Tanner codes)
- Variables nodes have even degree
- All local codewords have even weight

Example
- LDPC codes with even left degrees
- Irregular repeat accumulate codes where the repetition factors are even
- Expander codes with even variable node degrees and even weighted local codes
Deviations Based on Paths

- $p$ is a path of length $h$: $h$ can be greater than girth, $p$ may be not simple
- Each path $p$ defines a “characteristic” vector $\chi_G(p) \in \mathbb{R}^N$
  
  \[
  [\chi_G(p)]_v \triangleq \frac{1}{\deg_G(v)} \cdot |\{v \mid v \in p\}|.
  \]
- $\mathcal{B}^{(h)} \subset [0, 1]^N$ is the set of deviations
  
  \[
  \mathcal{B}^{(h)} \triangleq \left\{ \frac{\chi_G(p)}{h+1} \mid p \text{ is a backtrackless path of length } h \right\}
  \]

Example

- $p = (a, X, b, Z, a, Y, c, Z, b)$
- $h = 8$
- $\chi_G(p) = \{\frac{2}{3}, \frac{2}{3}, \frac{1}{3}, 0\}$
- $\frac{\chi_G(p)}{h+1} = \{\frac{2}{27}, \frac{2}{27}, \frac{1}{27}, 0\}$
Set of deviations $\mathcal{B}^{(h)} = \text{normalized characteristic vectors of } h\text{-paths.}$

$$\mathcal{B}^{(h)} \triangleq \left\{ \frac{\chi_G(p)}{h + 1} \mid p \text{ is a backtrackless path of length } h \right\}$$

**Definition**

A codeword $x \in \mathcal{C}$ is $h$-locally optimal w.r.t. $\lambda \in \mathbb{R}^N$ if for all vectors $\beta \in \mathcal{B}^{(h)}$,

$$\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle$$
Theorem

If $x$ is $h$-locally optimal w.r.t. $\lambda$, then $x$ is the unique ML-codeword w.r.t. $\lambda$.

Proof method:

**Lemma (Decomposition Lemma)**

Every codeword is a conical combination of $h$-paths in $G$

$$x = \alpha \cdot \mathbb{E}_{\beta \in \rho \mathcal{B}(h)} [\beta]$$

Proof of decomposition lemma

1. Every codeword is a conical combination of simple cycles in $G$
2. Every cycle is a conical combination $h$-paths in $G$

Following [ADS’09]: decomposition lemma $\Rightarrow$ unique ML
Verifying Local Optimality

- **Hard:** Is $x$ the unique ML-codeword?
- **Easy:** Is $x$ locally optimal?
  - Codeword can be efficiently verified to be locally-optimal w.r.t. $\lambda$ (dynamic programming / $\sim$Floyd's algorithm)
Theorem

If \( x \) is a \( h \)-locally optimal codeword w.r.t. \( \lambda \), then \( x \) is also the unique optimal LP solution given \( \lambda \).

Proof method: reduction to “ML” using graph covers.

\[ \tilde{z}^* = ML(\lambda^M) \]

In graph covers, realization of LP-Opt and ML codeword are the same

[15/23]
Local Optimality $\Rightarrow$ Unique LP optimality

**Theorem**

If $x$ is a $h$-locally optimal codeword w.r.t. $\lambda$, then $x$ is also the unique optimal LP solution given $\lambda$.

**Proof method:** reduction to “ML” using graph covers.

- **$M$-Covering Graph**
  - $\tilde{x} \triangleq x^\uparrow_M$ is locally-optimal w.r.t. $\lambda^\uparrow_M$

- **Base Graph**
  - $x$ is locally-optimal w.r.t. $\lambda$

Lemma: Local-optimality is *invariant* w.r.t. lifting to covering graphs
Local Optimality $\Rightarrow$ Unique LP optimality

Theorem

If $x$ is a $h$-locally optimal codeword w.r.t. $\lambda$, then $x$ is also the unique optimal LP solution given $\lambda$.

Proof method: reduction to “ML” using graph covers.

Thm: Local-Opt $\Rightarrow$ ML Opt.

M-Covering Graph $\tilde{z}^* = ML(\lambda^M)$ $\iff \tilde{x} \triangleq x^M$ is locally-optimal w.r.t. $\lambda^M$

Base Graph
Local Optimality $\Rightarrow$ Unique LP optimality

**Theorem**

*If $x$ is a $h$-locally optimal codeword w.r.t. $\lambda$, then $x$ is also the unique optimal LP solution given $\lambda$.*

**Proof method: reduction to “ML” using graph covers.**

- **Thm: Local-Opt $\Rightarrow$ ML Opt.**
  - $M$-Covering Graph: $\tilde{z}^* = ML(\lambda^M)$
  - $\tilde{x} \triangleq x^M$ is locally-optimal w.r.t. $\lambda^M$

- **Lemma: Local-optimality is invariant w.r.t. lifting to covering graphs**
  - [Vontobel-Koetter’05]

- **Base Graph**
  - $z^* = LP$ Opt.
  - $x$ is locally-optimal w.r.t. $\lambda$
Symmetry of local-optimality implies:

\[ \Pr\{\text{LP decoding fails}\} \leq \Pr\{\exists \beta \in \mathcal{B}^{(h)} \text{ s.t. } \langle \lambda, \beta \rangle \leq 0| c = 0^N \}. \]

- Let \( D \triangleq d_{\text{max}}^L \cdot d_{\text{max}}^R \)
- Bounds rely on: \( \text{girth}(G) > \log_D(N) \).
  
  [Gallager’63] gives an explicit construction of such graphs.

**Theorem**

**Consider BSC with crossover probability \( p \).**

**For every \( \epsilon > 0 \), if \( p < D^{\frac{1}{2}} \cdot (1 + \frac{d_{\text{min}}^L}{d_{\text{max}}^L}) \cdot (\epsilon + \frac{3}{2} + \frac{1}{2} \log_D(2)) \), then**

\[ \Pr\{\text{LP}(\lambda) \neq x | c = x\} \leq N^{-\epsilon} \]

- Analogous theorem derived for the BI-AWGN channel
- Obtain same results as in [Feldman-Karger’02], [Goldenberg-Burshtein’11] for RA(2) and RA(2q).
The proof technique in [KV’06], [ADS’09], [HE’10], [HE’11] is based on the following steps:

1. Define a set of deviations. A deviation is induced by combinatorial structures in the Tanner graph or the computation tree.

2. Define local-optimality. Loosely speaking, a codeword $x$ is locally-optimal if its cost is smaller than the cost of every relative point.

3. Local-optimality $\Rightarrow$ Unique ML-codeword.
   - *Decomposition Lemma*: Every codeword is a conical sum of deviations.

4. Local-optimality $\Rightarrow$ Unique LP-codeword.
   - *Lifting Lemma*: Local-optimality is invariant under liftings of codewords to covering graphs.

5. Analyze the probability that there does not exist a locally-optimal codeword.
Summary

Conclusions

- Simple application of the proof technique of "local-optimality" for bounds on the word error probability with LP-decoding
  - Even Tanner codes (both regular and irregular)
  - Local-optimality: deviations induced by paths in the Tanner graph
  - Inverse polynomial error bounds for the BSC and AWGNC

- Unified analysis framework that captures recent advances by [KV’06] [ADS’09] [HE’10] [HE’11] which present inverse exponential bounds on the decoding error probability for regular LDPC codes and Tanner codes.

Open questions

- Extend analysis of inverse exponential error bounds also to irregular Tanner codes
- Probabilistic analysis beyond the girth
Form of finite length bounds: \( \exists c > 1. \exists t. \forall \text{noise} < t \).

\[
\Pr\{\text{LP decoder fails}\} \leq e^{-c^{girth}}
\]

If girth = \( \theta(\log N) \), then

\[
\Pr\{\text{LP decoder fails}\} \leq e^{-N^\alpha}, \text{ for } 0 < \alpha < 1
\]

\( N \to \infty \): \( t \) is a lower bound on the threshold of LP-decoding with LO-certificate

<table>
<thead>
<tr>
<th>Decoder</th>
<th>[Koetter-Vontobel’06]</th>
<th>[Arora-Daskalakis-Steurer’09]</th>
<th>[H-Even’10][HE’11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>LP</td>
<td>LP</td>
<td>LP, Message-Passing</td>
</tr>
<tr>
<td></td>
<td>Dual LP witness, union bound</td>
<td>Primal LP, local optimality, rand. min-sum process</td>
<td>[ADS’09] + graph covers [VK’05]</td>
</tr>
<tr>
<td>channels</td>
<td>MBIOS</td>
<td>BSC</td>
<td>MBIOS</td>
</tr>
<tr>
<td>Example:</td>
<td>BSC((p)) threshold:</td>
<td>BSC((p)) threshold:</td>
<td>AWGN threshold:</td>
</tr>
<tr>
<td>(3, 6)-reg</td>
<td>( p^{\text{LP}} &gt; 0.01 )</td>
<td>( p^{\text{LP}} &gt; 0.05 )</td>
<td>( \frac{E_b}{N_0} ) &lt; 2.67dB</td>
</tr>
<tr>
<td>LDPC code</td>
<td>( \frac{E_b}{N_0}^{\text{LP}} &lt; 5.07\text{dB} )</td>
<td>( p^{\text{BP}} = 0.084 )</td>
<td>( \frac{E_b}{N_0} ) max–prod ( \approx 1.7\text{dB} )</td>
</tr>
</tbody>
</table>
Advances: Analysis of Finite-Length Codes - Tanner Codes

- **Form of finite length bounds:** \( \exists c > 1. \exists t. \forall \text{noise} < t. \)
  \[ \Pr\{\text{LP decoder fails}\} \leq \exp(-c^{\text{girth}}) \]
- If girth = \( \theta(\log N) \), then
  \[ \Pr\{\text{LP decoder fails}\} \leq \exp(-N^\alpha), \text{ for } 0 < \alpha < 1 \]
- \( N \to \infty \): \( t \) is a lower bound on the threshold of LP-decoding with LO-certificate

<table>
<thead>
<tr>
<th>Decoder</th>
<th>[Skachek-Roth'03]</th>
<th>[Feldman-Stein'05]</th>
<th>[H-Even'11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channels</td>
<td>Iterative</td>
<td>LP</td>
<td>LP</td>
</tr>
<tr>
<td>Technique</td>
<td>Expansion</td>
<td>Expansion</td>
<td>local-optimality + [Von'10] + sum-min-sum rand. process</td>
</tr>
<tr>
<td>Example: BSC((p)) threshold ((2, d_R))-reg Tanner code, Rate=0.375</td>
<td>( d_R &gt;&gt; 2 )</td>
<td>( d_R &gt;&gt; 2 )</td>
<td>( d_R = 16 )</td>
</tr>
<tr>
<td></td>
<td>( d^* &gt;&gt; 2 )</td>
<td>( d^* &gt;&gt; 2 )</td>
<td>( d^* = 4 )</td>
</tr>
<tr>
<td></td>
<td>( p^{\text{iterat.}} &gt; 0.0016 )</td>
<td>( p^{\text{LP}} &gt; 0.0008 )</td>
<td>( p^{\text{LP}} &gt; 0.044 )</td>
</tr>
</tbody>
</table>