Local Optimality Certificates for LP Decoding of Tanner Codes

Nissim Halabi
Guy Even

11th Haifa Workshop on Interdisciplinary Applications of Graph Theory, Combinatorics and Algorithms.  
May 2011
Error Correcting Codes – Worst Case Vs. Average Case Analysis

- An \([N,K]\) linear code \(C\) – \(K\)-dimensional subspace of the vector space \(\{0,1\}^N\)

- **Worst case analysis** – assuming adversarial channel.
  e.g., how many bit flips, in any pattern, can decoding recover?
  \[
  \Pr(\text{fail : worst case}) \sim p^{d_{\text{min}}/2}
  \]

- **Average case analysis** – probabilistic channel
  e.g., given that every bit is flipped with probability \(p\) independently, what is the probability that decoding succeeds?
  possibly, \(\Pr(\text{fail : avg. case}) \ll p^{d_{\text{min}}}\)
Error Correcting Codes for Memoryless Binary-Input Output-Symmetric Channels (1)

- Memoryless Binary-Input Output-Symmetric Channel
  - characterized by conditional probability function $P(y | c)$
  - Errors occur randomly and are independent from bit to bit (memoryless)
  - Assumes transmitted symbols are binary
  - Errors affect ‘0’s and ‘1’s with equal probability (i.e., symmetric)

- Example: Binary Symmetric Channel (BSC)
Error Correcting Codes for Memoryless Binary-Input Output-Symmetric Channels (2)

Channel Encoding $c \in \mathcal{C} \subseteq \{0,1\}^N$\hspace{1cm} Noisy Channel $\lambda(y) \in \mathbb{R}^N$\hspace{1cm} Channel Decoding $\hat{c} \in \{0,1\}^N$

- **Log-Likelihood Ratio (LLR) $\lambda_i$** for a received observation $y_i$:

$$\lambda_i(y_i) = \ln \left( \frac{\mathbb{P}_{Y_i|X_i}(y_i | x_i = 0)}{\mathbb{P}_{Y_i|X_i}(y_i | x_i = 1)} \right)$$

- $\lambda_i > 0 \implies y_i$ is more likely to be ‘0’
- $\lambda_i < 0 \implies y_i$ is more likely to be ‘1’

- $\lambda \leftrightarrow y \implies$ replace $y$ by $\lambda$
Maximum-Likelihood (ML) Decoding

- Maximum-likelihood (ML) decoding for any binary-input memory-less channel:

\[
ML(\lambda) = \arg \min_{x \in C} \langle \lambda, x \rangle
\]

- Maximum-likelihood (ML) decoding formulated as a linear program:

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

No Efficient Representation

\[C \in \{0, 1\}^N\]

\[\text{conv}(C) \in [0, 1]^N\]
Linear Programming (LP) Decoding

**Linear Programming (LP) decoding** \cite{Fel03, FWK05} – relaxation of the polytope $\text{conv}(C)$

$$LP(\lambda) = \arg\min_{x \in \mathcal{P}} \langle \lambda, x \rangle$$

\[ \mathcal{P} : \]
1. All codewords $x$ in $C$ are vertices
2. All new vertices are fractional (therefore, new vertices are not in $C$)
3. Has an efficient representation

Solve LP

- $LP(\lambda)$ integral $\Rightarrow$ $LP(\lambda) = ML(\lambda)$
- $LP(\lambda)$ fractional $\Rightarrow$ fail!

$C \in \{0, 1\}^N$
$\text{conv}(C) \in [0,1]^N$
$\mathcal{P}$
$\text{conv}(C) \subseteq \mathcal{P}$
Linear Programming (LP) Decoding

- Linear Programming (LP) decoding \([\text{Fel03, FWK05}]\) – relaxation of the polytope \(\text{conv}(C)\)

\[
LP(\lambda) = \arg \min_{x \in P} \langle \lambda, x \rangle
\]

\(P\) : (1) All codewords \(x\) in \(C\) are vertices
(2) All new vertices are fractional (therefore, new vertices are not in \(C\))
(3) Has an efficient representation

\(\text{LP decoder finds} \quad \text{ML codeword} \)

\(\text{LP} (\lambda) \) integral \(\Rightarrow\) \(\text{LP}(\lambda) = \text{ML}(\lambda)\)
\(\text{LP}(\lambda) \) fractional \(\Rightarrow\) fail!

\(\mathcal{C} \in \{0,1\}^N\)
\(\text{conv}(\mathcal{C}) \in [0,1]^N\)
\(\mathcal{P}\)
\(\text{conv}(\mathcal{C}) \subseteq \mathcal{P}\)
Linear Programming (LP) Decoding

- Linear Programming (LP) decoding \([\text{Fel}03, \text{FWK}05]\) – relaxation of the polytope \(\text{conv}(C)\)

\[
LP(\lambda) = \arg\min_{x \in P} \langle \lambda, x \rangle
\]

\(P\): (1) All codewords \(x\) in \(C\) are vertices
(2) All new vertices are fractional (therefore, new vertices are not in \(C\))
(3) Has an efficient representation

- Solve LP
- \(LP(\lambda)\) integral \(\Rightarrow\) \(LP(\lambda) = ML(\lambda)\)
- \(LP(\lambda)\) fractional \(\Rightarrow\) \text{fail!}

\(\mathcal{C} \subseteq \{0,1\}^N\)
\(\text{conv}(\mathcal{C}) \subseteq [0,1]^N\)
\(\mathcal{P}\)
\(\text{conv}(\mathcal{C}) \subseteq P\)
Tanner Codes [Tan81]

Factor graph representation of Tanner codes:

- Every Local-code node $C_j$ is associated with linear code of length $\deg_G(C_j)$

- Tanner code $C(G)$ and codewords $x$:

  $$x \in C(G) \iff \forall j. \ x_{\mathcal{N}(C_j)} \in \text{local-code } C_j$$

- \(d^* = \min_j \left[ d_{\min} \left( \text{local-code } C_j \right) \right]\)

- Extended local-code $C_j \subseteq \{0,1\}^N$: extend to bits outside the local-code

- Example: Expander codes [SS’96]
  Tanner graph is an expander; Simple bit flipping decoding algorithm.

\[ G = ( \mathcal{I} \cup \mathcal{J}, E ) \]
LP Decoding of Tanner Codes

- **Maximum-likelihood (ML) decoding:**

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

where \( \text{conv}(C) = \cap_{j} \text{conv}(\text{extended local-code } C_j) \)

- **Linear Programming (LP) decoding** [following Fel03, FWK05]:

\[ LP(\lambda) = \arg \min_{x \in \mathcal{P}} \langle \lambda, x \rangle \]

where \( \mathcal{P} = \cap_{j} \text{conv}(\text{extended local-code } C_j) \)
Criterions of Interest

- Let $\lambda \in \mathbb{R}^N$ denote an LLR vector received from the channel.
- Let $x \in \mathcal{C}(G)$ denote a codeword.
- Consider the following questions:

  - $x = \text{ML}(\lambda)$
    - NP-Hard
  - $x = \text{LP}(\lambda)$

  - $\text{ML}(\lambda)$ unique?
  - $\text{LP}(\lambda)$ unique?

- E.g., efficient test via local computations $\rightarrow$ “Local Optimality” criterion
Let \( x \in C(G) \subseteq \{0,1\}^N \)

\[ f \in [0,1]^N \subseteq \mathbb{R}^N \]

[Fel03] Define relative point \( x \oplus f \) by \( (x \oplus f)_i = |x_i - f_i| \)

Consider a finite set \( B \subseteq [0,1]^N \)

**Definition**: A codeword \( x \in C \) is **locally optimal** for \( \lambda \in \mathbb{R}^N \) if for all vectors \( \beta \in B \),

\[
\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle
\]

**Goal**: find a set \( B \) such that:

1. \( x \in LO(\lambda) \Rightarrow x \in ML(\lambda) \) and \( ML(\lambda) \) unique
2. \( x \in LO(\lambda) \Rightarrow x \in LP(\lambda) \) and \( LP(\lambda) \) unique

\[
\mathbb{P}\{\text{LP decoding fails}\} \leq \mathbb{P}\{\exists \beta \in B. \langle \lambda, \beta \rangle \leq 0 | c = 0^N\}
\]

\[
\mathbb{P}\left( \bigcap_{\beta \in B} \langle \lambda, \beta \rangle > 0 | c = 0^N \right) = 1 - o(1)
\]

All-Zeros Assumption
Combinatorial Characterization of Local Optimality (2)

- **Goal**: find a set $\mathcal{B}$ such that:
  1. $x \in \text{LO}(\lambda) \Rightarrow x \in \text{ML}(\lambda)$ and $\text{ML}(\lambda)$ unique
  2. $x \in \text{LO}(\lambda) \Rightarrow x \in \text{LP}(\lambda)$ and $\text{LP}(\lambda)$ unique
  3. $\mathbb{P}\left( \bigcap_{\beta \in \mathcal{B}} \langle \lambda, \beta \rangle > 0 \mid c = 0^N \right) = 1 - o(1)$

- Suppose we have properties (1), (2).
  - Large support($\beta$) $\Rightarrow$ property (3). (e.g., Chernoff-like bounds)

- If $\mathcal{B} = \mathcal{C}$, then: $x \in \text{LO}(\lambda) \iff x \in \text{ML}(\lambda)$ and $\text{ML}(\lambda)$ unique
  - However, analysis of property (3) ???

$\beta$ – GLOBAL Structure
Combinatorial Characterization of Local Optimality (2)

- **Goal:** Find a set $B$ such that:
  1. $x \in LO(\lambda) \Rightarrow x \in ML(\lambda)$ and $ML(\lambda)$ unique
  2. $x \in LO(\lambda) \Rightarrow x \in LP(\lambda)$ and $LP(\lambda)$ unique
  3. $\mathbb{P}\left( \bigcap_{\beta \in B} \langle \lambda, \beta \rangle > 0 \mid c = 0^N \right) = 1 - o(1)$

Suppose we have properties (1), (2).
Large support($\beta$) $\Rightarrow$ property (3). (e.g., Chernoff-like bounds)

- For analysis purposes, consider structures with a local nature
  $\Rightarrow B$ is a set of TREES [following KV’06]

- Strengthen analysis by introducing layer weights! [following ADS’09]
  $\Rightarrow$ better bounds on
  $\mathbb{P}\left( \bigcap_{\beta \in B} \langle \lambda, \beta \rangle > 0 \mid c = 0^N \right)$

- Finally, $\text{height}(\text{subtrees}(G)) < \frac{1}{2} \text{girth}(G) = O(\log N)
  \Rightarrow$ Take path prefix trees – not bounded by girth!
Consider a graph $G=(V,E)$ and a node $r \in V$:

- $\hat{V}$ – set of all backtrackless paths in $G$ emanating from node $r$ with length at most $h$.
- $\hat{E} \triangleq \left\{(p_1, p_2) \in \hat{V} \times \hat{V} \mid p_1 \text{ is a prefix of } p_2 \text{ and } |p_1| + 1 = |p_2|\right\}$.
- $T^h_r(G) \triangleq (\hat{V}, \hat{E})$ – path-prefix tree of $G$ rooted at node $r$ with height $h$. 

Path-Prefix Tree
Path-Prefix Tree

Consider a graph \( G=(V,E) \) and a node \( r \in V \):

- \( \hat{V} \) – set of all backtrackless paths in \( G \) emanating from node \( r \) with length at most \( h \).
- \( \hat{E} \triangleq \{(p_1, p_2) \in \hat{V} \times \hat{V} | p_1 \text{ is a prefix of } p_2 \text{ and } |p_1| + 1 = |p_2|\} \).
- \( T_r^h(G) \triangleq (\hat{V}, \hat{E}) \) – path-prefix tree of \( G \) rooted at node \( r \) with height \( h \).
Consider a path-prefix tree $T^h_r(G)$ of a Tanner graph $G = (\mathcal{I} \cup \mathcal{J}, E)$.

$d$-tree $\mathcal{T}[r,h,d]$ – subgraph of $T^h_r(G)$

- root $= r$
- $\forall v \in \mathcal{T} \cap \mathcal{I}: \deg_{\mathcal{T}}(v) = \deg_G(v)$.
- $\forall c \in \mathcal{T} \cap \mathcal{J}: \deg_{\mathcal{T}}(c) = d$.

2-tree = skinny tree / minimal deviation

3-tree

4-tree

Not necessarily a valid configuration!
Cost of a Projected Weighted Subtree

Consider layer weights $\omega : \{1, \ldots, h\} \rightarrow \mathbb{R}$, and a subtree $\mathcal{T}_{\hat{r}}$ of a path prefix tree $T_r^{2h}(G)$.

Define a weight function $\mathcal{T}_{\hat{r}}^{(\omega)} : \hat{\mathcal{V}} \rightarrow \mathbb{R}$ for the subtree $\mathcal{T}_{\hat{r}}$ induced by $\omega$:

$$\mathcal{T}_{\hat{r}}^{(\omega)}(\hat{u}) = \frac{\omega_t}{\deg_G(v)} \cdot \prod_{\hat{u} \in P_{\hat{r}, \hat{v}} \setminus \{\hat{r}, \hat{v}\}} \frac{1}{\deg_{\mathcal{T}_{\hat{r}}}(\hat{u}) - 1}$$

where $t = \lceil \frac{d(\hat{r}, \hat{v})}{2} \rceil$ and $v \sim \hat{v}$.

$\pi_G \left[ \mathcal{T}_{\hat{r}}^{(\omega)} \right] \in \mathbb{R}^N$ – projection to Tanner graph $G$.
Combinatorial Characterization of Local Optimality

- **Setting:**
  - $C(G) \subseteq \{0,1\}^N$ Tanner code with minimal local distance $d^*$
  - $2 \leq d \leq d^*$
  - $\omega \in [0,1]^h \backslash \{0^N\}$
  - $B_d^{(\omega)}$ – set of all vectors corresponding to projections to $G$ by $\omega$-weighted $d$-trees of height $2h$ rooted at variable nodes

$$B_d^{(\omega)} = \left\{ \pi_G\left( T^{(\omega)} \right) \mid T^{(\omega)} \text{ is } \omega\text{-weighted a } d\text{-tree of height } h \right\}$$

- **Definition:** A codeword $x$ is $(h, \omega, d)$-locally optimal for $\lambda \in \mathbb{R}^N$ if for all vectors $\beta \in B_d^{(\omega)}$,

$$\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle$$
**Thm: Local Opt ⇒ ML Opt / LP Opt**

- **Theorem:** If \( x \in \mathcal{C}(G) \) is \((h, \omega, d)\)-locally optimal for \( \lambda \), then:
  1. \( x \) is the unique ML codeword for \( \lambda \).
  2. \( x \) is the unique optimal LP solution given \( \lambda \).

**Goals achieved:**
\[
\begin{align*}
\text{(1)} & \ x \in \text{LO}(\lambda) \Rightarrow x \in \text{ML}(\lambda) \text{ and } \text{ML}(\lambda) \text{ unique} \\
\text{(2)} & \ x \in \text{LO}(\lambda) \Rightarrow x \in \text{LP}(\lambda) \text{ and } \text{LP}(\lambda) \text{ unique}
\end{align*}
\]

**Left to show:**
\[
\begin{align*}
\text{(3)} & \ \Pr\{x \in \text{LO}(\lambda)\} = 1 - o(1)
\end{align*}
\]
Thm: Local Opt ⇒ ML Opt / LP Opt

**Theorem:** If $x \in \mathcal{C}(G)$ is $(h, \omega, d)$-locally optimal for $\lambda$, then:

1. $x$ is the unique ML codeword for $\lambda$.
2. $x$ is the unique optimal LP solution given $\lambda$.

**Interesting outcomes:**

- Characterizes the event of LP decoding failure. For example:

  **Theorem:** Fix $\omega \in \mathbb{R}^h_+$. Then
  \[
  \mathbb{P}\{\text{LP decoding fails}\} \leq \mathbb{P}\left\{ \exists d\text{-tree } \tau. \langle \pi_G[\tau], \lambda \rangle \leq 0 \mid c = 0^n \right\}.
  \]

- Work in progress: Design an *iterative message passing decoding algorithm* that computes an $(h, \omega, d)$-locally-optimal codeword after $h$ iterations.
  - $\exists x$ locally optimal codeword for $\lambda \Rightarrow$ weighted message passing algorithm computes $x$ in $h$ iterations + guarantee that $x$ is the ML codeword.
Local Optimality Certificates for LP decoding

Previous results:

- Koetter and Vontobel ’06 – Characterized LP solutions and provide a criterion for certifying the optimality of a codeword for LP decoding of regular LDPC codes.
  - Characterization is based on combinatorial structure of skinny trees of height \( h \) in the factor graph. \( (h < \frac{1}{4}\text{girth}(G)) \)

- Arora, Daskalakis and Steurer ’09 – Certificate for LP decoding of regular LDPC codes over BSC. Extension to MBIOS channels in [H-Even ’10].
  - Certificate characterization is based on weighted skinny trees of height \( h \).
    \( (h < \frac{1}{4}\text{girth}(G)) \)

Note: Bounds on the word error probability of LP decoding are computed by analyzing the certificates. By introducing layer weights, ADS certificate occurs with high probability for much larger noise rates.

- Vontobel ’10 – Implies a certificate for LP decoding of Tanner codes based on weighted skinny subtrees of graph covers.
Local Optimality Certificates for LP decoding – Summary of Main Techniques

<table>
<thead>
<tr>
<th></th>
<th>[KV06, ADS09, HE10]</th>
<th>Current work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Girth</strong></td>
<td>$h &lt; \frac{1}{4}\text{girth}(G)$ No dependencies on the factor graph.</td>
<td>$h$ is unbounded Characterization using computation trees</td>
</tr>
<tr>
<td><strong>Regularity</strong></td>
<td>Local Isomorphism</td>
<td>Irregular factor graph – add normalization factors according to node degrees.</td>
</tr>
<tr>
<td><strong>Check Nodes / Constraints</strong></td>
<td>Parity code.</td>
<td>Linear Codes. Tighter relaxation for the generalized fundamental polytope.</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
    x_1 & \quad C_1 \\
    x_2 & \quad C_2 \\
    x_3 & \quad C_3 \\
    x_4 & \quad C_4 \\
    x_5 & \quad C_5 \\
    x_6 & \quad C_6 \\
    x_7 & \quad C_7 \\
    x_8 & \quad C_8 \\
    x_9 & \quad C_9 \\
    x_{10} & \quad C_{10}
\end{align*}
\]
## Local Optimality Certificates for LP decoding – Summary of Main Techniques

<table>
<thead>
<tr>
<th></th>
<th>[KV06, ADS09, HE10]</th>
<th>Current work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Girth</strong></td>
<td>$h &lt; \frac{1}{4}\text{girth}(G)$ No dependencies on the factor graph.</td>
<td>$h$ is unbounded Characterization using computation trees</td>
</tr>
<tr>
<td><strong>Regularity</strong></td>
<td>Local Isomorphism</td>
<td>Irregular factor graph – add normalization factors according to node degrees.</td>
</tr>
<tr>
<td><strong>Check Nodes / Constraints</strong></td>
<td>Parity code.</td>
<td>Linear Codes. Tighter relaxation for the generalized fundamental polytope.</td>
</tr>
<tr>
<td><strong>Deviations</strong></td>
<td>“skinny” Locally satisfies inner parity checks.</td>
<td>“fat” – Certificates based on “fat” structures likely to occur with high probability for larger noise rates. Not necessarily a valid configuration!</td>
</tr>
</tbody>
</table>
## Local Optimality Certificates for LP decoding – Summary of Main Techniques

<table>
<thead>
<tr>
<th></th>
<th>[KV06,ADS09,HE10]</th>
<th>Current work</th>
</tr>
</thead>
</table>
| **Girth**            | $h < \frac{1}{4}\text{girth}(G)$  
No dependencies on the factor graph. | $h$ is unbounded Characterization using computation trees |
| **Regularity**       | Local Isomorphism  
**Regular** factor graph. | Irregular factor graph – add normalization factors according to node degrees. |
| **Check Nodes / Constraints** | Parity code. | Linear Codes.  
**Tighter relaxation** for the generalized fundamental polytope. |
| **Deviations**       | “skinny”  
Locally satisfies inner parity checks. | “fat” – Certificates based on “fat” structures likely to occur with high probability for larger noise rates.  
Not necessarily a valid configuration! |
| **LP solution analysis / characterization** | Dual / Primal LP analysis. Polyhedral analysis. | Use reduction to ML via characterization of graph cover decoding. |
Conclusions

- Combinatorial, graph theoretic, and algorithmic methods for analyzing decoding of (modern) error correcting codes over any memoryless channel:
  - New local optimality combinatorial certificate for LP decoding of irregular Tanner codes, i.e., Local Opt. ⇒ LP Opt.
    - The certificate is based on weighted “fat” subtrees of computation trees of height $h$.
      - $h$ is not bounded by girth.
  - Proofs using combinatorial decompositions and graph covers arguments.
  - Efficient algorithm (dynamic programming), runs in $O(|E| \cdot h)$ time, for the computation of local optimality certificate of a codeword given an LLR vector.
Future Work

- Work in progress: Design an iterative message passing decoding algorithm that computes an \((h, \omega, d)\)-locally-optimal codeword after \(h\) iterations.
  - \(\exists x\) locally optimal codeword for \(\lambda \Rightarrow\) weighted message passing algorithm computes \(x\) in \(h\) iterations + guarantee that \(x\) is the ML codeword.

- Asymptotic analysis of LP decoding and weighted min-sum decoding for ensembles of irregular Tanner codes.
  - Based on density evolution techniques and the combinatorial characterization of local optimality.
Thank You!