Decoding Error-Correcting Codes via Linear Programming

Ph.D. Thesis Defense

Jon Feldman
jonfeld@theory.lcs.mit.edu

Advisor: David Karger

Joint work with David Karger, Martin Wainwright

MIT Laboratory for Computer Science

June 3, 2003
Binary Error-Correcting Code

Transmitter with encoder

Information: "lg. pepperoni"  codeword("lg. pepperoni")

\[ x \]
\[
\begin{array}{c}
010011 \\
\leftarrow k \rightarrow
\end{array}
\]

\[ y \]
\[
110011101001 \\
\leftarrow n \rightarrow
\]

Binary Symmetric Channel:

Flip each bit w/ probability \( p < 1/2 \)

"lg. pepperoni"  codeword("lg. pepperoni")  corrupt codeword

\[ x \]
\[
010011 \\
\leftarrow k \rightarrow
\]

\[ y \]
\[
110011101001 \\
\leftarrow n \rightarrow
\]

\[ \hat{y} \]
\[
110010101011 \\
\leftarrow n \rightarrow
\]

Receiver with decoder
Repetition Code Example

- **Encoder:** Repeat each information bit 5 times.
  - Information word: 1011
  - Codeword: 11111 00000 11111 11111
  - Corrupt codeword: 10110 01000 01001 10111

- **Decoder:** Take majority within every group of 5.
  - Decoded codeword: 11111 00000 00000 11111
  - Decoded info word: 1001

- Information transmitted successfully
  \[\leq\] at most 2 bits flipped per group of 5.
Outline

- Coding background
- Our contributions:
  - LP decoding: general method.
  - LP decoders for turbo, LDPC codes.
  - Performance bounds for turbo, LDPC codes.
  - Connections to message-passing decoders.
  - Experiments (supporting theory).
  - Methods for tightening the relaxation.
  - New dual-based message-passing algorithms.
- Future work
Basic Coding Terminology

- A code is a subset $C \subseteq \{0, 1\}^n$, where $|C| = 2^k$.
- Block length = length = $n$. Affects latency, encoder/decoder complexity, performance.
- Minimum distance = distance = $d = \min_{y,y' \in C} \Delta(y, y')$. “Worst case” measure of performance.
- Word error rate (WER) = probability of decoding failure = $\Pr_{\text{noise}}[\text{transmitted } \overline{y} \neq \text{decoded } y]$. Practical measure of performance.
Maximum-Likelihood (ML) Decoding

- ML decoders minimize WER.
  - BSC: Finds $y \in C$ s.t. $\Delta(y, \hat{y})$ is minimum.
  - Corrects errors up to half the minimum distance.

\[ \hat{y} = \text{rec. cw} \]

\[ \text{cw(“Red Sox win”)} \]

\[ \text{cw(“Yankees win”)} \]

\[ d/2 \]

- Cost function $\gamma_i$: negative log-likelihood ratio of $y_i$.

\[ [\gamma_i > 0 \implies y_i \text{ likely 0}] \quad [\gamma_i < 0 \implies y_i \text{ likely 1}] \]

**ML Decoding**:

Given corrupt codeword $\hat{y}$, find $y \in C$ such that $\sum_i \gamma_i y_i$ is minimized.
LP Decoding

Polytope $P \subseteq [0, 1]^n$

$\min \sum_i \gamma_i y_i$

$P \cap \{0, 1\}^n = C$

- LP variables $y_i$ for each code bit, relaxed $0 \leq y_i \leq 1$.
- Alg: Solve LP. If $y^*$ integral, output $y^*$, else "error."
- ML certificate property
LP Decoding Success Conditions

Objective function cases

(a) No noise
(b) Both succeed
(c) ML succeed, LP fail
(d) Both fail, detected
(e) Both fail, undetected

trans. cw(“The Eagle has landed”)

cw(“The beagle was branded”)
Fractional Distance

- Another way to define (classical) distance $d$:
  - $d = \min l_1$ dist. between two integral vertices of $P$.

- Fractional distance:
  - $d_{frac} = \min l_1$ distance between an integral vertex and any other vertex of $P$.
  - Lower bound on classical distance: $d_{frac} \leq d$.

Theorem: In the binary symmetric channel, LP decoders can correct up to $\left\lfloor d_{frac}/2 \right\rfloor - 1$ errors.

- Given facets of $P$, fractional distance can be computed efficiently.
Turbo Codes + LDPC Codes

- Low-Density Parity-Check (LDPC) codes [Gal ’62].
- Turbo Codes introduced [BGT ’93], unprecedented error-correcting performance.
- Ensuing LDPC “Renaissance:”
  - Expander codes [SS ’94]
  - Message-passing algorithms [Wib ’96]
  - Connection to belief-propagation [MMC ’98]
  - Message-passing capacity [RU, LMSS, RSU, BRU, CFDRU, ’99-’01]
  - Designing irregular codes [LMSS ’01]
  - Connection to “Bethe free energy” [Yed ’02]
**Factor Graph**

- *Factor (Tanner) Graph* of a linear code: bipartite graph modeling the *parity check matrix* of the code.

- "Variable nodes" \( y_1, \ldots, y_n \).
- "Check Nodes" \( c_1, \ldots, c_m \).
- \( N(j) \): neighborhood of check \( c_j \).

- Codewords: \( y \in \{0, 1\}^n \) s.t.:

  \[
  \forall c_j, \sum_{i \in N(j)} y_i = 0 \pmod{2}
  \]

- Codewords: 0000000, 1110000, 1011001, etc.
LP Relaxation on the Factor Graph

For all var. nodes $i$:
- $0 \leq f_i \leq 1$

For all check nodes $j$:
- $\{f_i : i \in N(j)\} \in P_j$

$P_j$: Parity Polytope
- [Yan ’99, Jer ’75]

$(f_2, f_4, f_6) \in P_j$
Fractional Solutions

- Suppose: \( \gamma_1 = -2.8 \)
  \( \gamma_2 = +0.8 \)
  \( \gamma_3 \ldots 7 = +1 \)

- ML codeword: \([1, 1, 1, 0, 0, 0, 0]\)
- ML codeword cost: \(-1\).

- Frac. sol: \( f = [1, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, \frac{1}{2}] \).
- Satisfies LP constraints?
  A: \([1, \frac{1}{2}, \frac{1}{2}, 0] = \frac{1}{2}[1, 1, 0, 0] + \frac{1}{2}[1, 0, 1, 0]\)
  B, C: similar.
- Frac. sol cost: \(-1.4\).
LP Decoding Success Conditions

- \( \Pr[ \text{Decoding Success} ] = \Pr[ \overline{y} \text{ is the unique OPT} ] \).
- Can we assume \( \overline{y} = 0^n \)? (This is a common assumption for linear codes.)

**Thm:** For LP decoding of binary linear codes, the WER is independent of the transmitted codeword.

- \( \Pr[ \overline{y} \text{ is the unique OPT} ] = \Pr[ \text{All pcw’s cost} > 0 ] \).
- “Combinatorial” characterization of pseudocodewords (scale the LP vertices).

**Thm:** The LP decoder succeeds iff all non-zero pseudocodewords have positive cost.
Performance Bounds

- **Turbo Codes:** For rate-1/2 RA (cycle) codes: If $G$ has large girth, negative-cost points in $P$ are rare.
  - Erdös (or [BMMS ’02]): Hamiltonian 3-regular graph with girth $\log n$.

  **Thm:** For any $\alpha > 0$, if $p < 2^{f(\alpha)}$, then $\text{WER} \leq n^{-\alpha}$.

- **LDPC Codes:** All var. nodes in $G$ have degree $\geq d_\ell$:

  **Thm:** If $G$ has girth $g$, then $d_{frac} \geq (d_\ell - 1)^{\lceil g/4 \rceil - 1}$

  - Can achieve $d_{frac} = \Omega(n^{1-\epsilon})$. Stronger graph properties (expansion?) are needed for stronger results.
Growth of Fractional Distance

- Random (3,4) LDPC Code
Message-Passing Decoders

(a) Var-to-check messages

(b) Check-to-var messages

(c) Hard Decision

\( m_{ij} \)

\( \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \cdots \quad \gamma_n \)

\( m_{ji} \)

\( \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \cdots \quad \gamma_n \)

0 1 1 0 1

J. Feldman, Ph.D. thesis defense – p.17/26
Min-Sum Update Rules

\[ m_{ij} = \gamma_i + \sum m_{ji} = -8 \]

\[ m_{ji} = \min(S : i = 1) - \min(S : i = 0) \]

\[
\begin{array}{c}
[1, 0, 1] : -1 \\
[0, 1, 1] : +2 \\
[0, 0, 0] : 0 \\
[1, 1, 0] : +1
\end{array}
\]

\[-1 - 0 = -1\]

- Let \( x = \sum m_{ji} + \gamma_i \).
  - if \( x > 0 \), output 0
  - if \( x < 0 \), output 1
Analyzing Message-Passing Decoders

- Sum-product, min-sum, Gallager, Sipser/Spielman, tree-reweighted max-product [WJW ’02].
- Message cycles: dependencies difficult to analyze.
- Density Evolution [RU ’01, LMSS ’01, ...]:
  - Assume “tree-like” message neighborhood, random graph from ensemble.
  - If err < threshold, any WER achievable (with high probability), for sufficiently large $n$.
- Finite-length analysis: combinatorial error conditions known for the binary erasure channel [DPRTU ’02].
- LP Decoding: well-characterized error conditions for general channels, any block length, even with cycles.
Unifying other “pseudocodewords”

- **BEC**: Sum-prod. fails $\iff$ *stopping set* [DPTRU ’02].
  - Thm: LP pseudocodewords $=$ stopping sets.

- **Tail-Biting trellisses**: Min-sum fails $\iff$ neg-cost dominant pseudocodeword [FKMT ’98].
  - Thm: LP pcws. $=$ dominant pseudocodewords

- **Cycle Codes**: Min-sum fails $\iff$ neg-cost irreducible closed walk [Wib ’96].
  - Thm: LP pcws. $=$ irreducible closed walks

- **LDPC codes**: Min-sum fails $\iff$ neg-cost deviation set in computation tree [Wib ’96].
  - LP pseudocodewords: natural “closed” analog of deviation sets.
Performance Comparison

WER Comparison: Random Rate-1/2 (3,6) LDPC Code

- Length 200, left degree 3, right degree 6.
Tightening the Relaxation

- If constraints are added to the polytope, the decoder can only improve. **Example: redundant parity checks.**

- Generic tightening techniques [LS ’91] [SA ’90].
Using Lift-And-Project

WER Comparison: Random Rate-1/4 (3,4) LDPC Code

- Length 36, left degree 3, right degree 4.
New Message-Passing Algorithms

Original LP relaxation

- Dual variables: messages.
- Enforce dual constraints.
- Convergence to codeword $\implies$ primal optimum.
- ML certificate.
New Message-Passing Algorithms

- Tree-reweighted max-product uses LP dual variables
  \[\implies\] TRMP has ML certificate property.

- Using complimentary slackness, conventional message-passing algorithms gain ability to show an ML certificate.

- Use subgradient algorithm to solve dual directly.
  - Gives message passing algorithm with ML certificate property, combinatorial success characterizations.
Future Work

- New WER, fractional distance bounds:
  - Lower rate turbo codes (rate-1/3 RA).
  - Other LDPC codes, including
    * Expander codes,
    * Irregular LDPC codes,
    * Other constructible families.
  - Random linear/LDPC codes?
- ML Decoding using IP, branch-and-bound?
- Using “lifting” procedures to tighten relaxation?
- Deeper connections to “sum-product” (belief-prop)?
- LP decoding of other code families, channel models?