Source Coding with Low Density Nonlinear Nodes

Marc Mézard

Santa Fe, January 10-12, 2005
Collaboration

M. Mézard (Université Paris Sud and ’Stipco*’ EC network)

S. Ciliberti (Université Paris Sud)

R. Zecchina (ICTP Trieste and ’Stipco’ EC network)

† Thanks to David Saad and Jonathan Yedidia

(*) “Statistical Physics of Information Processing and Combinatorial Optimization”, EC supported Research Training Network
General idea

• Success of LDPC codes in channel coding

• Seeking low density graphical codes in source coding
  – **Parity Source Coder**: theoretically close to optimum, no good encoding algorithm
  – **Nonlinear Source Coder**: theoretically close to optimum, good encoding with survey propagation
General idea

- Success of LDPC codes in channel coding

- Seeking low density graphical codes in source coding
  - Parity Source Coder: theoretically close to optimum, no good encoding algorithm
  - Nonlinear Source Coder: theoretically close to optimum, good encoding with survey propagation

- Two main issues:
  - Theoretical capacity: approaching the Shannon limit
  - Algorithmic performance: polynomial time encoding and decoding
Lossy Data Compression

\[ \{y_1, y_2, \ldots, y_M\} \xrightarrow{\text{encoding}} \{x_1, x_2, \ldots, x_N\} \xrightarrow{\text{decoding}} \{y_1^*, y_2^*, \ldots, y_M^*\} \]

\( y_a, y_a^* \in \text{alphabet } S; \ x_i \in \{0, 1\} . \)

Here \( S = \{0, 1\} \), and \( y_a \) are iid, \( = 0, 1 \) with probability \( 1/2 \).

Rate \( R = N/M \). Distortion \( D = (1/M) \sum_{i=1}^{M} (1 - \delta(y_a, y_a^*)) \)
Lossy Data Compression

\[ \{y_1, y_2, \ldots, y_M\} \xrightarrow{\text{encoding}} \{x_1, x_2, \ldots, x_N\} \xrightarrow{\text{decoding}} \{y_1^*, y_2^*, \ldots, y_M^*\} \]

\( y_a, y_a^* \in \text{alphabet } S; \ x_i \in \{0, 1\}. \)

Here \( S = \{0, 1\}, \) and \( y_a \) are iid, \( = 0, 1 \) with probability 1/2.

Rate \( R = N/M. \) Distortion \( D = (1/M) \sum_{i=1}^{M} (1 - \delta(y_a, y_a^*)) \)

**Shannon:** the minimum achievable rate \( R \) given the distortion \( D \) is

\[ R(D) = 1 - h_2(D) \]
Uncorrelated unbiased binary source

\[ R(D) = 1 - h_2(D) \]

\( h_2(\cdot) \) being the binary entropy

Shannon’s bound
General strategy

Compressed message: \( \{x_1, x_2 \ldots x_N\} = N \) bits

Initial message: \( M \) symbols. \( \rightarrow M \) function nodes

Example: ‘Parity Source coder’: a LDPC code for compression
Distortion and energy

\[ y = (0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1) \]

\[ x = (x_1 \ x_2 \ \ldots \ x_N) \]

Encoding: \( y \rightarrow x = \) Find a configuration of \( x \) which violates the smallest number of checks. = Find a ground state.
Distortion and energy

\[ y = (0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1) \]

\[ x = (x_1 \ x_2 \ \ldots \ldots \ x_N) \]

Encoding: \( y \rightarrow x = \) Find a configuration of \( x \) which violates the smallest number of checks. \( = \) Find a ground state.

Decoding: \( x \rightarrow y^* = \) trivial. Example: \( y_1^* = x_1 \oplus x_2 \oplus x_3 \)
Distortion and energy

Encoding: $y \rightarrow x =$ Find a configuration of $x$ which violates the smallest number of checks. $\Rightarrow$ Find a ground state.

Decoding: $x \rightarrow y^* =$ trivial. Example: $y_1^* = x_1 \oplus x_2 \oplus x_3$

Distortion $= \text{number of violated checks in the encoding} = \text{ground state energy of the encoding}$. 

\[ y = (0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1) \]

\[ x = (x_1 \quad x_2 \quad \ldots \quad x_N) \]
'XORSAT' problem in combinatorial optimization, or 'p-spin' model in statistical physics.

Instance: \{ Geometry, y_1, \ldots, y_M \}. \quad E(x) = \sum_a \left[ 1 - \delta \left( \bigoplus_{i \in V(a)} x_i, y_a \right) \right]

SAT configuration: \( x \) with \( E(x) = 0 \). \quad UNSAT instance: \( \forall x : E(x) > 0 \).

MAX-XORSAT: ground state energy \( E_0 \) (\( = \) distortion)
Theoretical performance of the Parity Source Coder 

'XORSAT’ problem in combinatorial optimization, or 'p-spin’ model in statistical physics.

Instance= \{ Geometry, y_1, \ldots, y_M \}. \ E(x) = \sum_a [1 - \delta (\bigoplus_{i \in V(a)} x_i, y_a)]

SAT configuration: \ x \ with \ E(x) = 0. \ UNSAT instance: \ \forall x : E(x) > 0.

MAX-XORSAT: ground state energy \ E_0 \ (=\text{distortion})

Analytic study: “1-step RSB” cavity method (MM,Ricci-Tersenghi,Zecchina 2002). Random $K-$XORSAT problem with $M, N \to \infty$ and $M/N = \alpha$ ($= 1/R$). (Function nodes: degree $K$. Variable nodes: degree Poisson($K\alpha$)). \ \to \ \text{Phase diagram, } E_0.
Theoretical performance of the Parity Source Coder II

Ground state energy $E_0$  
$K = 3, 4, 5, 6$ and Shannon

Rapidly approaching Shannon’s bound when $K$ increases. Optimal data compression...
Theoretical performance of the Parity Source Coder II

Ground state energy $E_0$

$K = 3, 4, 5, 6$ and Shannon

Rapidly approaching Shannon’s bound when $K$ increases. Optimal data compression... in theory!
Practical performance of the Parity Source Coder

Encoding: \( y \rightarrow x = \) Find a configuration of \( x \) which violates the smallest number of checks. = Find a ground state. DIFFICULT!!

Belief propagation: does not converge (NB: random initial condition, very different from the case of LDPC codes).
Practical performance of the Parity Source Coder

Encoding: \( y \rightarrow x = \text{Find a configuration of } x \text{ which violates the smallest number of checks.} = \text{Find a ground state. DIFFICULT!!!} \)

Belief propagation: does not converge (NB: random initial condition, very different from the case of LDPC codes).

Clusters of solutions
Clusters in the XORSAT problem

Random parity checks (graph, $y_a$); $N, M \rightarrow \infty$, $\alpha = M/N$ fixed.

With probability one:
$\alpha < \alpha_c$: SAT
$\alpha > \alpha_c$: UNSAT

But three phases:
Easy SAT, Hard SAT, UNSAT.

<table>
<thead>
<tr>
<th>K</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_c$</td>
<td>.92</td>
<td>.97</td>
<td>.99</td>
<td>1.00</td>
</tr>
<tr>
<td>$\alpha_d$</td>
<td>.82</td>
<td>.77</td>
<td>.70</td>
<td>.63</td>
</tr>
</tbody>
</table>
Why clusters ‘kill’ belief propagation

Belief propagation:

\[ P_{a \rightarrow 1}(x_1) = \sum_{x_2, x_3} C_a(x_1, x_2, x_3) P^{(a)}(x_2) P^{(a)}(x_3) \]
\[ P^{(b)}(x_1) \propto \prod_{a \in V(1) \setminus b} P_{a \rightarrow 1}(x_1) \]
Why clusters ‘kill’ belief propagation

Belief propagation:

\[ P_{a\rightarrow 1}(x_1) = \sum_{x_2, x_3} C_a(x_1, x_2, x_3) P^{(a)}(x_2) P^{(a)}(x_3) \]

\[ P^{(b)}(x_1) \propto \prod_{a \in V(1) \setminus b} P_{a\rightarrow 1}(x_1) \]

Basic underlying idea: \( P^{(a)}(x_2, x_3) \sim P^{(a)}(x_2) P^{(a)}(x_3) \). Correct if

1) \( x_2, x_3 \) distant (OK)

2) Measure restricted to one cluster (Wrong in the hard SAT phase).
From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief → Warning → Survey

Belief $P_{a \rightarrow i}(x_i), P^{(a)}(x_i)$ are probabilities, in $[0, 1]$.

For each belief, e.g. $P_{a \rightarrow i}(x_i)$, construct the warning $\rho_{a \rightarrow i}(x_i) \in \{0, \ast\}$.

$$\rho = I(P) = \begin{cases} 
0 & \text{if } P_{a \rightarrow i}(x_i) = 0 \\
\ast & \text{if } P_{a \rightarrow i}(x_i) > 0
\end{cases}$$

Warning propagation: focuses on forced variables.
From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief \( \rightarrow \) Warning \( \rightarrow \) Survey

Belief \( P_{a \rightarrow i}(x_i), P^{(a)}(x_i) \) are probabilities, in \([0, 1]\).

For each belief, e.g. \( P_{a \rightarrow i}(x_i) \), construct the warning \( \rho_{a \rightarrow i}(x_i) \in \{0, *\} \).

\[
\rho = I(P) = \begin{cases} 
0 & \text{if } P_{a \rightarrow i}(x_i) = 0 \\
* & \text{if } P_{a \rightarrow i}(x_i) > 0
\end{cases}
\]

Warning propagation: focuses on forced variables.

Survey: \( \eta_{a \rightarrow i} = \) probability of a warning being sent from constraint \( a \) to variable \( i \), when a cluster is picked up at random. Propagate the surveys along the graph, then use them to decimate. OK in satisfiability/colouring problem.
From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief $\rightarrow$ Warning $\rightarrow$ Survey

Belief $P_{a\rightarrow i}(x_i), P^{(a)}(x_i)$ are probabilities, in $[0, 1]$.

For each belief, e.g. $P_{a\rightarrow i}(x_i)$, construct the warning $\rho_{a\rightarrow i}(x_i) \in \{0, *\}$.

$$\rho = I(P) = \begin{cases} 0 & \text{if } P_{a\rightarrow i}(x_i) = 0 \\ * & \text{if } P_{a\rightarrow i}(x_i) > 0 \end{cases}$$

Warning propagation: focuses on forced variables.

Survey: $\eta_{a\rightarrow i} =$ probability of a warning being sent from constraint $a$ to variable $i$, when a cluster is picked up at random. Propagate the surveys along the graph, then use them to decimate. OK in satisfiability/colouring problem. Not in XORSAT (totally symmetric messages)...
Nonlinear nodes

Introduce generalized function nodes, different from parity checks

- Keep theoretical performances nearly as good as in parity checks
- Break the symmetry $\rightarrow$ SP converges and allows to encode a message

$$y= (0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1)$$

$$x= (x_1 \ x_2 \ \ldots \ \ldots \ x_N)$$
Random Nonlinear nodes

Parity check

Random check

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

SCRAMBLE

Symmetry
Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation.
Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation. Survey Propagation (SP): On each edge $a \rightarrow i$: survey = Proba(warning), when a cluster of ground states is chosen at random.

Penalty when conflicting warnings → works also in the UNSAT phase

\[ \eta_{a \rightarrow 1}: \text{known exactly from} \]
\[ \eta_{b \rightarrow 2} \text{ and } \eta_{c \rightarrow 3}. \]

\[ \eta_{a \rightarrow 1} = \text{Prob(warning)} \]
Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation. Survey Propagation (SP): On each edge $a \rightarrow i$: survey = Proba(warning), when a cluster of ground states is chosen at random.

\[ \eta_{a \rightarrow 1}: \text{known exactly from} \]
\[ \eta_{b \rightarrow 2} \text{ and } \eta_{c \rightarrow 3}. \]

Penalty when conflicting warnings → works also in the UNSAT phase

Statistical analysis → phase diagram.

Single sample → SP algorithm
Theoretical capacity (20 types of random nodes)
Phase diagram

- static
- dynamic
- Gardner

\( \alpha \)
\( \alpha_{\text{d}} \)
\( \alpha_{\text{s}} \)
\( \alpha_{\text{G}} \)

energy

0.8 1 1.2 1.4 1.6 1.8 2

\( \alpha \)
Performance

\[ \text{distortion rate} = \frac{1}{\alpha} \]

Theoretical capacity
Shannon’s bound
Algorithm
Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems
- Theoretical capacity \( \approx \) Shannon’s bound
- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy \( \Leftrightarrow \) approaching the Shannon’s bound
Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems

- Theoretical capacity \( \approx \) Shannon’s bound

- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy \( \Leftrightarrow \) approaching the Shannon’s bound

- To be improved: Encoding works, but still slow (limited to \( N < 10000 \), using general purpose SP software...)
Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems

- Theoretical capacity $\approx$ Shannon’s bound

- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy $\Leftrightarrow$ approaching the Shannon’s bound

- To be improved: Encoding works, but still slow (limited to $N < 10000$, using general purpose SP software...)

- Perspectives: Generalize this algorithm in order to compress sequences of real numbers. Revisit nonlinear function nodes in channel coding.