11. Density evolution
Probabilistic analysis of message passing algorithms

Consider factor graph model $G = (V, F, E)$ and

$$\mu(x) = \frac{1}{Z} \prod_{a \in F} \psi_a(x_{\partial a}) \prod_{i \in V} \psi_i(x_i)$$

Sum-product algorithm and max-product algorithms are instances of message-passing algorithms

- Discrete $x_i \in \mathcal{X}$
- Two sets of messages $\{\nu_{i \rightarrow a}(x_i)\}$ and $\{\tilde{\nu}_{a \rightarrow i}(x_i)\}$
- Update:

$$\nu_{i \rightarrow a}^{(t+1)} = F_{i \rightarrow a}(\{\tilde{\nu}_{b \rightarrow i}^{(t)} : b \in \partial i \setminus a\})$$

$$\tilde{\nu}_{a \rightarrow i}^{(t)} = G_{a \rightarrow i}(\{\nu_{j \rightarrow a}^{(t)} : j \in \partial a \setminus i\})$$
• assumptions for probabilistic analysis
  ▶ a random graph is a graph $G = (V, F, E)$ where $E$ is drawn randomly
    from a set of possible graphs
    e.g., Erdös-Renyi graph, random regular graph
  ▶ asymptotic analysis: in the limit $n \to \infty$
• density evolution is used in
  ▶ analyzing channel codes
  ▶ analyzing solution space of XORSAT
  ▶ analyzing a message-passing algorithm for crowdsourcing
  ▶ etc.
Example: channel coding

- sending messages through a noisy channel

![Channel Diagram](image)

- channel is defined by $\mathbb{P}_{Y|X}(y|x)$
- Binary Erasure Channel (BEC)
  - $x \in \{0, 1\}$, $y \in \{0, 1, *\}$

![BEC Diagram](image)

- goal: estimate $\hat{x}_1, \ldots, \hat{x}_n$ given $y_1, \ldots, y_n$
- performance metric: average bit error probability

$$P_{\text{error}} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(x_i \neq \hat{x}_i)$$
Binary erasure channel

![Binary erasure channel diagram]

- no coding: $(01001) \Rightarrow (01 \ast 0\ast)$
  - block length $n = 5$
  - $P_{\text{error}} = \epsilon/2$
- repetition code: $(0001110000000111) \Rightarrow (0 \ast \ast 1 \ast 10 \ast 0 \ast \ast \ast 111)$
  - $n = 15$
  - $P_{\text{error}} = \epsilon^3/2$
  - rate $= 1/3$
- information theory
  - capacity of a BEC is $1 - \epsilon$
  - there exists a code such that $\lim_{n \to \infty} P_{\text{error}} = 0$ with rate $r = 1 - \epsilon$
  - using the BEC $n$ times, one can reliably send $(1 - \epsilon)n$ bits of messages
Modern coding theory

- modern codes = iterative decoding (belief propagation)
  - Turbo code
  - Low-Density Parity Check (LDPC) code
  - Polar code
  - etc.

- LDPC code is defined by a factor graph model

\[ \psi_a(x_i, x_j, x_k) = I(x_i \oplus x_j \oplus x_k = 0) \]

- block length \( n = 4 \)
- number of factors \( m = 2 \)
- allowed messages = \{0000, 0111, 1010, 1101\}
- message size \( k = n - m = 2 \)
- rate \( r = k/n = 1/2 \)
- received \( y = (0 \star 1\star) \), then \( \hat{x} = (0111) \)
- received \( y = (0 \star **) \), then ?
Modern coding theory

- decoding using belief propagation

\[ \mu_y(x) = \frac{1}{Z} \prod_{i \in V} \mathbb{P}_{Y|X}(y_i | x_i) \prod_{a \in F} \mathbb{I}(\oplus x_a = 0) \]

- use (parallel) sum-product algorithm to find \( \mu(x_i) \) and let

\[ \hat{x}_i = \arg \max \mu(x_i) \]

- minimizes bit error rate

Parallel sum-product for BEC

- \( \nu_{i \rightarrow a}^{(t)} \in \{0, 1, *\} \) our belief about \( x_i \)
- \( \tilde{\nu}_{a \rightarrow i}^{(t)} \in \{0, 1, *\} \) our belief about \( x_i \)

at iteration 0:

\[ \nu_{i \rightarrow a}^{(0)} = y_i \]

at iteration \( t \):

\[ \tilde{\nu}_{a \rightarrow i}^{(t)} = \begin{cases} * & \text{if any of the incoming messages is a *} \\ \oplus x_{\partial a \setminus i} & \text{otherwise} \end{cases} \]
Peeling decoder
Peeling decoder
Peeling decoder

Density evolution
Peeling decoder

Density evolution
Peeling decoder

Density evolution
Peeling decoder

Density evolution
Peeling decoder
Peeling decoder

Density evolution
Peeling decoder

Density evolution
Probabilistic analysis: density evolution

- a few assumptions
  - sparse random graph construction
e.g. random \((\ell, r)\)-regular graph from the configuration model
  - asymptotic analysis:
in the limit \(n \to \infty\) but finite number of iterations \(t\)
- locally-tree like structure ensures that the incoming messages are i.i.d.
- formally, as \(n \to \infty\) local neighborhood of a node converges in probability to a random tree
- density evolution

- \(z_t \in [0, 1]\) be the probability a randomly chosen message from \(\{\nu_{i \to a}^{(t)}\}\) is an erasure
- \(w_t \in [0, 1]\) be the probability a randomly chosen message from \(\{\nu_{a \to i}^{(t)}\}\) is an erasure
- in the limit \(n \to \infty\), they satisfy the density evolution equations

\[
\begin{align*}
  w_t &= 1 - (1 - z_{t-1})^{r-1} \\
  z_t &= \epsilon w_{t-1}^{\ell-1}
\end{align*}
\]
\[ z_t = \epsilon (1 - (1 - z_{t-1})^{r-1})^{\ell-1} \]

with initial condition \( z_0 = \epsilon \)

- density evolution for (3,6) code with \( \epsilon = 0.4 \) (left) and 0.45 (right)

- rate of this code = 0.5, threshold \( \approx 0.4 \)...

- \( P_{\text{error}}(t) = \lim_{n \to \infty} P_{\text{error}}(n, t) \)

- analyze \( \lim_{t \to \infty} \lim_{n \to \infty} P_{\text{error}}(n, t) \), is this what we want?
bit error rate of $(3, 6)$-codes

\[
\left( \frac{z_t}{\epsilon} \right)^{1/(\ell-1)} = 1 - (1 - z_{t-1})^{r-1}
\]

\[
\begin{align*}
\epsilon &= 0.4 \\
\epsilon &= 0.44
\end{align*}
\]

extend this analysis to construct capacity achieving *tornado codes*

\[
\int_0^1 \epsilon y^{\ell-1} dy = \frac{\epsilon}{\ell}, \quad \int_0^1 (1 - (1 - x)^{r-1}) dx = 1 - \frac{1}{r}
\]

Density evolution 11-11
density evolution for general message passing algorithms

-variable nodes
- factor nodes

\[
\psi_a(x_i, x_j, x_k)
\]

consider factor graph model \( G = (V, F, E) \) and

\[
\mu(x) = \frac{1}{Z} \prod_{a \in F} \psi_a(x_{\partial a}) \prod_{i \in V} \psi_i(x_i)
\]

\( \nu_{i \to a}^{(t+1)} = F_{i \to a}(\{\tilde{\nu}_{b \to i}^{(t)} : b \in \partial i \setminus a\}) \)

\( \nu_{a \to i}^{(t)} = G_{a \to i}(\{\nu_{j \to a}^{(t)} : j \in \partial a \setminus i\}) \)

density evolution equation

\[
\nu^{(t+1)} = F(\nu_1^{(t)}, \ldots, \nu_{\ell-1}^{(t)})
\]

\[
\nu^{(t)} = G(\nu_1^{(t)}, \ldots, \nu_{k-1}^{(t)})
\]
formally, as $n \to \infty$ a randomly chosen message from $\{\nu_i^{(t)} \}$ converge in probability to $z^{(t)}$

who cares about random graphs?

who cares about asymptotics?

<table>
<thead>
<tr>
<th>alphabet $x_i \in \mathcal{X}$</th>
<th>messages $\nu_{i \to a} \in \mathcal{Y}$</th>
<th>density $\mathcal{Z}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>discrete ${0, 1}$</td>
<td>discrete ${0, 1, *}$</td>
<td>continuous $\mathbb{R}^2$</td>
</tr>
<tr>
<td>discrete</td>
<td>continuous $\mathbb{R}^{</td>
<td>\mathcal{X}</td>
</tr>
<tr>
<td>continuous $\mathbb{R}$</td>
<td>distribution over $\mathbb{R}$</td>
<td>dist. over dist. over $\mathbb{R}$</td>
</tr>
</tbody>
</table>

how do we compute evolution of distributions?

- quantization
- Gaussian approximation
- *population dynamics*: represent the density using ‘samples’