Simplicity and Complexity of Belief-Propagation

Elchanan Mossel\textsuperscript{1

\textsuperscript{1}MIT

July 2020
A *Double* phase transition for large $q$

**Theorem (Count Reconstruction, Robust Reconstruction (Mossel-Peres, Janson-Peres))**

For all $q$ and $d$-ary tree, $d\theta^2 = 1$ is the threshold for: census and robust reconstruction.

**Theorem (Reconstruction for large $q$ (Mossel 00))**

If $d \theta > 1$ then for $q > q_\theta$ can distinguish the root better than random:

$$\lim_{h \to \infty} \text{Var}[\mathbb{E}[X_0|X_{L_h}]] > 0$$

$\implies$ Non-linear estimators are superior.

**Pf:** Shows fractal nature of information.
Proof sketch

- For $q = \infty$, clearly threshold is $d\theta = 1$.
- For finite $q, d = 2$, fix $\theta$ such that $d\theta > 1$.
- Inference: Infer root color to be $c$ if there is an $\ell$-diluted binary subtree $T' \subset T$ with root at 0 and where all leaves have color $c$.
- **Exercise 1**: There exists an $\ell, \varepsilon > 0$ such that if the root is $c$, the probability that such a tree exists is at least $\varepsilon$.
- **Exercise 2**: For all $\varepsilon > 0$, if $q$ is sufficiently large, and if the root is not $c$, the probability that there is an $\ell$-diluted $2^\ell - 1$ tree with all the leaves of color $\neq c$ is at least $1 - \varepsilon/10$.
- **Exercise 3**: Prove that if $d\lambda \leq 1$, then the root and leaves are asymptotically independent.
Sly 11: Defined magnetization $m_n = E[M_n]$ such that if $m_n$ is small then:

$$m_{n+1} = d\theta^2 m_n + (1 + o(1))\frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \theta^4 m_n^2.$$ 

$\implies$ if $q \geq 5$, the KS bound is not tight.

Also proved that if $q = 3$ and $d \geq d_{\text{min}}$ is large then KS bound is tight.

M-01: For general Markov chains, can have $\lambda_2(M) = 0$, yet root and leaves are not independent.

Exercise: Prove this for following chain on $F_2^2$.
$M(x, y) = (r, r \oplus x)$ or $(r, r \oplus y)$ with probability $1/2$ each.

More sophisticated examples in Mossel-Peres.
Two conjectures about inference

- Consider a model where different edges have different $\theta$'s.
- Let $q$ so that for $\theta \in (\theta_R, \theta_{KS})$, $\text{Var}[\mathbb{E}[X_0|X_h]] \to \alpha > 0$.
- Conj 1: There is no estimator $f$ such that $f(X_h)$ and $X_0$ have no negligible correlation for all models with $\theta(e) \in (\theta_R, \theta_{KS})$ for all edges.
- Conj 2: It is “impossible” to recover phylogenetic trees using $O(h)$ samples under the conditions above.
- Strong version of impossible would mean information theoretically. Weak version would mean computationally.
Part 3: Complexity of BP
What is the complexity of BP?

Low: Runs in linear time.
But: Uses real numbers - is this necessary?
But: Uses depth - is this necessary?
Fractal picture suggests maybe depth is needed.
Understanding the Omnipresence

What is everywhere and understand everything?
“Omnipresence”.
A: The deep-net on your smartphone that understands you.
Mathematically, it is natural to ask if there are data generative process satisfying 3 natural criteria:

1. **Realism**: Reasonable data models.

2. **Reconstruction**: Provable efficient algorithms to reverse engineer the generative process.

3. **Depth**: Proof that depth is needed.

4. Also: why does BP use real numbers, when the generating process is discrete?
**Q:** What are the memory requirements for BP?

**Conjecture (EKPS-00):** For $q = 2$, any recursive algorithm on the tree which uses at most $B$ bits of memory per node can only distinguish the root value better than random if

\[ \theta < \theta(B) \text{ where } d\theta(B)^2 > 1. \]

**Thm: (Jain-Koehler-Liu-M-19):** Conjecture is true:

\[ \theta(B) - \theta = B^{-O(1)}. \]
Problem Setup

Generation tree (broadcast model)

Reconstruction (message passing)
Problem Setup (cont.)

- Broadcast process on $d$-regular tree of height $h$.
- Each reconstruction $Y_i = f_i(Y_{2i}, Y_{2i+1})$ is an arbitrary log $L$-bit string (memory constraint).
\( \text{AC}^0 \) := class of bounded depth circuits with AND/OR (unbounded fan) and NOT gates.

**Thm**: Moitra-M-Sandon-20:
\( \text{AC}^0(X_h) \) cannot classify \( X_0 \) better than random.

Is this trivial?

Maybe not: **Thm** MMS-20: \( \text{AC}^0 \) generates leaf distributions.
\( \textbf{TC}^0 \)

- \( \textbf{TC}^0 := \text{like AC}^0 \) but with Majority gates.
- “Bounded depth deep nets”.
- Thm (MMS-20): When \( q = 2 \) and \( 0.9999 < \theta < 1 \), there exists an algorithm \( A \) in \( \textbf{TC}^0 \) such that 
  \[
  \lim_h P[A(X_h) = X_0] = \lim_h P[BP(X_h) = X_0].
  \]
- Conj: This is true for all \( \theta \) when \( q = 2 \).
- So maybe we can classify optimally in \( \textbf{TC}^0 \)?
- Maybe bounded depth nets suffice?
\[ \text{NC}^1 \]

- \( \text{NC}^1 \) := class of \( O(\log n) \) depth circuits with AND/OR (fan 2) and NOT gates.
- Known that \( \text{TC}^0 \subset \text{NC}^1 \). Open if they are the same.
- Thm (MMS-20): One can classify as well as BP in \( \text{NC}^1 \).
- Thm (MMS-20): There is a broadcast process for which classifying better than random is \( \text{NC}^1 \)-complete.
- So, unless \( \text{TC}^0 = \text{NC}^1 \), \( \log n \) depth is needed.
The KS bound and Circuit Complexity

- The threshold $2\theta^2 = 1$ is called the Kesten-Stigum threshold.
- Above this threshold it is known that one neuron can classify the root better than random (Kesten-Stigum-66).
- Below this threshold, one neuron cannot (M-Peres-04).
- Below this threshold, with enough i.i.d. noise on the leaves, BP becomes trivial (Janson-M-05).
- Related to “Replica Symmetry Breaking” in statistical physics models (Mezard-Montanari-06).
- Conjecture (MMS-20): For any broadcast process, below the KS bound and where BP classifies better than random, classification is $\text{NC}^1$-complete.
Conclusion

BP is simple:
- Runs in linear time.
- Above KS bound behaves like a Linear Algorithm.

BP is complex:
- Below KS bound, tend to be fractal.
- Statistical/computation gaps.
- Requires depth / precision.