Uniqueness of BP fixed point for Ising models

Yury Polyanskiy

EECS
Massachusetts Institute of Technology

Feb. 25, 2022 Harvard, Cambridge, MA

with: Qian Yu (Princeton) – *on the job market!*

What is “BP” in the title?

Belief propagation (BP) operator $\mathcal{Q}$ is a map of probability measures:

- Fix a probability measure μ on $\mathbb{R}$ and let $\mathcal{Q}\mu$ be the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,,$$

where $R_i \stackrel{iid}{\sim} \mu$ independent of $X_i \stackrel{iid}{\sim} \mathrm{Bern}(\delta)$ and

$$F_\delta(x) = 2\mathrm{atanh}((1-2\delta)\tanh(\frac{x}{2}))$$

What is "BP" in the title?

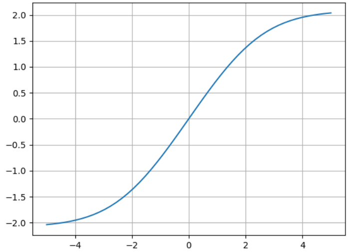
Belief propagation (BP) operator $\mathcal{Q}$ is a map of probability measures:

- Fix a probability measure μ on $\mathbb{R}$ and let $\mathcal{Q}\mu$ be the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,,$$

where $R_i \overset{iid}{\sim} \mu$ independent of $X_i \overset{iid}{\sim} \mathrm{Bern}(\delta)$ and

$$F_\delta(x) = 2\mathrm{atanh}((1-2\delta)\tanh(\frac{x}{2}))$$



What is "BP" in the title?

Belief propagation (BP) operator $\mathcal{Q}$ is a map of probability measures:

- Fix a probability measure μ on $\mathbb{R}$ and let $\mathcal{Q}\mu$ be the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,,$$

  where $R_i \stackrel{iid}{\sim} \mu$ independent of $X_i \stackrel{iid}{\sim} \mathrm{Bern}(\delta)$ and

$$F_\delta(x) = 2\mathrm{atanh}((1-2\delta)\tanh(\frac{x}{2}))$$

- Variations:
  - d itself could be random (most often: $d \sim \mathrm{Poi}(\bar{d})$)
  - we can have "side information" or "survey" $S_0 \sim \mu_0$:

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0$$

BP-operator: $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \mathrm{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$\mathcal{Q}\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

BP-operator: $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \stackrel{iid}{\sim} \mu, X_i \stackrel{iid}{\sim} \mathrm{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$\mathcal{Q}\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* and $\mathcal{Q}^k\mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.

BP-operator: $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$\mathcal{Q}\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* and $\mathcal{Q}^k\mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.

Resolves multiple conjectures:

- [Kanade-Mossel-Schramm'2014]: labeled 2-SBM
- [Mossel-Xu'2015]: optimality of local algorithms for 2-SBM
- [Mossel-Neeman-Sly'2016]*: independence to leaf noise in BOT
- [Abbe-Cornacchia-Gu-P.'2021]*: entropy characterization in 2-SBM

BP-operator: $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \stackrel{iid}{\sim} \mu, X_i \stackrel{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$\mathcal{Q}\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* and $\mathcal{Q}^k\mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.

Resolves multiple conjectures:

- [Kanade-Mossel-Schramm'2014]: labeled 2-SBM
- [Mossel-Xu'2015]: optimality of local algorithms for 2-SBM
- [Mossel-Neeman-Sly'2016]*: independence to leaf noise in BOT
- [Abbe-Cornacchia-Gu-P.'2021]*: entropy characterization in 2-SBM

The last two contain partial result: uniqueness for $(1-2\delta)^2 d \gg 1$

BP-operator: $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

Main Question: Characterize fixed points

$$\mathcal{Q}\mu = \mu$$

Note: When $S_0 = 0$ choice $\mu = \delta_0$ is always a (trivial) fixed point

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* and $\mathcal{Q}^k\mu \to \mu^*$ as $k \to \infty$ for any $\mu \neq \delta_0$.

This talk: What is this relevant for?
How information theory helped us prove it?

Resc

-
- BM
- in BOT
- [Abbe-Cornacchia-Gu-P.'2021]*: entropy characterization in 2-SBM

The last two contain partial result: uniqueness for $(1-2\delta)^2 d \gg 1$

Three application domains

- Application 1: Statistical Physics
- Application 2: Machine Learning
- Application 3: Information propagation

Application 1: Ising Model on Sparse Graphs

- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}$$

with Hamiltonian

$$H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

and $x_u \in \{\pm 1\}$ – binary spin variables.

- Many graphs are "complex" but "sparse". So model them as a random d-regular graph.

Application 1: Ising Model on Sparse Graphs

- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}$$

  with Hamiltonian

$$H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

  and $x_u \in \{\pm 1\}$ – binary spin variables.
- Many graphs are "complex" but "sparse". So model them as a random d-regular graph.
- Locally such graphs are just trees.

Application 1: Ising Model on Sparse Graphs

- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}$$

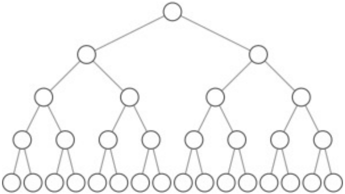
with Hamiltonian

$$H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

and $x_u \in \{\pm 1\}$ – binary spin variables.

- Many graphs are "complex" but "sparse". So model them as a random d-regular graph.
- Locally such graphs are just trees.
- So let us understand Ising model on trees.

Application 1: Ising Model on trees

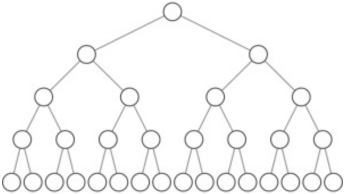


- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}, \qquad H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

  and $x_u \in \{\pm 1\}$ – binary spin variables.
- Problem: How to extend the above to infinite trees?

Application 1: Ising Model on trees



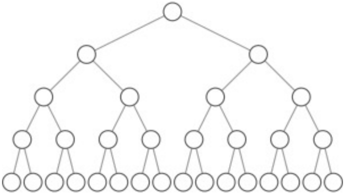
- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}, \qquad H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

  and $x_u \in \{\pm 1\}$ – binary spin variables.

- Problem: How to extend the above to infinite trees?
- How to understand Gibbs measure on infinite tree?

Application 1: Ising Model on trees

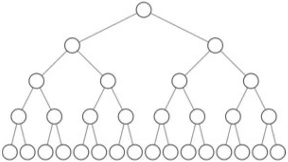


- Model of correlated phenomena: Ising model on a graph $G = (V, E)$

$$\mathbb{P}[\mathbf{X} = \mathbf{x}] = \frac{1}{Z} e^{\beta H(\mathbf{x})}, \qquad H(\mathbf{x}) = \sum_{u \sim v} x_u x_v$$

  and $x_u \in \{\pm 1\}$ – binary spin variables.
- Problem: How to extend the above to infinite trees?
- How to understand Gibbs measure on infinite tree?
- Similar issue on finite locally-tree-like graphs with correlated boundary.

Application 1: Ising Model on infinite trees



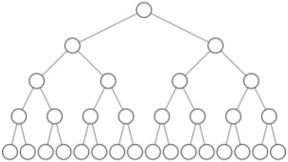
- For every finite subtree T with boundary $L = \partial T$ we have:

$$\mathbb{P}[\mathsf{X}_T = \mathsf{x}_T | \mathsf{X}_L = \mathsf{y}_L] = \frac{1}{Z} e^{\beta H_T(\mathsf{x}_T)} 1\{\mathsf{x}_L = \mathsf{y}_L\}$$

where H_T is the restriction of Hamiltonian to a subtree T:

$$H_T(\mathsf{x}_T) = \sum_{u \sim v: u \in T, v \in T} x_u x_v$$

Application 1: Ising Model on infinite trees



- For every finite subtree T with boundary $L = \partial T$ we have:

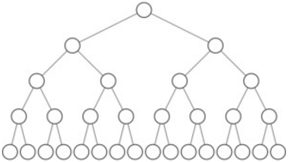
$$\mathbb{P}[\mathbf{X}_T = \mathbf{x}_T | \mathbf{X}_L = \mathbf{y}_L] = \frac{1}{Z} e^{\beta H_T(\mathbf{x}_T)} 1\{\mathbf{x}_L = \mathbf{y}_L\}$$

where H_T is the restriction of Hamiltonian to a subtree T:

$$H_T(\mathbf{x}_T) = \sum_{u \sim v: u \in T, v \in T} x_u x_v$$

- As usual there is a phase transition:
  - high temp $\beta \le \frac{1}{\operatorname{atanh}(d)}$: there is a unique Gibbs measure on infinite tree. The choice of boundary condition $\mathbf{y}_L$ is irrelevant.
  - medium temp $\beta > \frac{1}{\operatorname{atanh}(d)}$: depending on $\mathbf{y}_L$ we can get (uncountably many) Gibbs measures on infinite tree.

Application 1: Ising Model on infinite trees



- For every finite subtree T with boundary $L = \partial T$ we have:

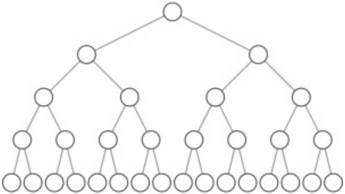
$$\mathbb{P}[\mathbf{X}_T = \mathbf{x}_T | \mathbf{X}_L = \mathbf{y}_L] = \frac{1}{Z} e^{\beta H_T(\mathbf{x}_T)} 1\{\mathbf{x}_L = \mathbf{y}_L\}$$

where H_T is the restriction of Hamiltonian to a subtree T:

> Problem: How to classify these Gibbs measures?

-
  - high temp $\beta \leq \frac{1}{\mathrm{atanh}(d)}$: there is a unique Gibbs measure on infinite tree. The choice of boundary condition $\mathbf{y}_L$ is irrelevant.
  - medium temp $\beta > \frac{1}{\mathrm{atanh}(d)}$: depending on $\mathbf{y}_L$ we can get (uncountably many) Gibbs measures on infinite tree.

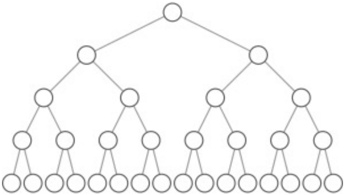
Application 1: free boundary Gibbs measure



- For every finite subtree T we define

$$\mathbb{P}_T[\mathsf{X}_T = \mathsf{x}_T] = \frac{1}{Z} e^{\beta H_T(\mathsf{x}_T)}$$

Application 1: free boundary Gibbs measure

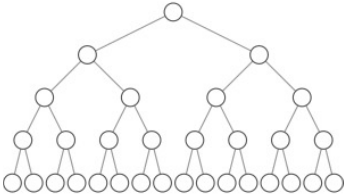


- For every finite subtree T we define

$$\mathbb{P}_T[\mathsf{X}_T = \mathsf{x}_T] = \frac{1}{Z} e^{\beta H_T(\mathsf{x}_T)}$$

- As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure

Application 1: free boundary Gibbs measure



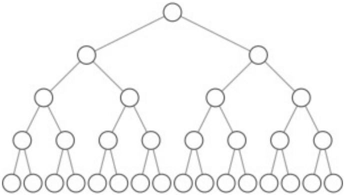
- For every finite subtree T we define

$$\mathbb{P}_T[\mathbf{X}_T = \mathbf{x}_T] = \frac{1}{Z} e^{\beta H_T(\mathbf{x}_T)}$$

- As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure
- For low temp $\beta > \frac{1}{\operatorname{atanh}(\sqrt{d})}$ measure $\mathbb{P}_\infty$ is not extremal and decomposes as

$$\mathbb{P}_\infty = \sum_\alpha w_\alpha \mathbb{P}_\alpha$$

Application 1: free boundary Gibbs measure



- For every finite subtree T we define

$$\mathbb{P}_T[\mathsf{X}_T = \mathsf{x}_T] = \frac{1}{Z} e^{\beta H_T(\mathsf{x}_T)}$$

- As $T \to \infty$ we have $\mathbb{P}_T \to \mathbb{P}_\infty$ – the free boundary Gibbs measure
- For low temp $\beta > \frac{1}{\operatorname{atanh}(\sqrt{d})}$ measure $\mathbb{P}_\infty$ is not extremal and decomposes as

$$\mathbb{P}_\infty = \sum_\alpha w_\alpha \mathbb{P}_\alpha$$

- Question: How to "count" $\mathbb{P}_\alpha$'s?

Mezard-Parisi: Replica-Symmetry Breaking (1RSB)

- To understand $\mathbb{P}_\infty$ we can sample boundary condition $\mathbf{y}_L$ from the distribution $\mathbb{P}_T$

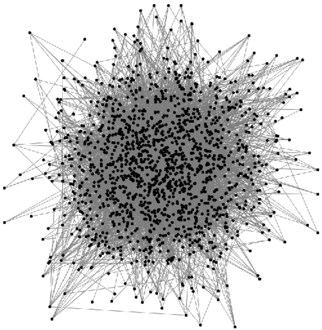
Mezard-Parisi: Replica-Symmetry Breaking (1RSB)

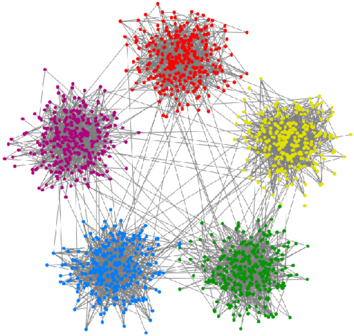

- To understand $\mathbb{P}_\infty$ we can sample boundary condition $\mathbf{y}_L$ from the distribution $\mathbb{P}_T$
- ... and then compute the finite distribution $\mathbb{P}[\mathbf{X}_T|\mathbf{X}_L = \mathbf{y}_L]$. Note: it is a random distribution.
- Mezard and Parisi: this random distribution is described by a fixed point $\mathcal{Q}\mu = \mu$

Mezard-Parisi: Replica-Symmetry Breaking (1RSB)

- To understand $\mathbb{P}_\infty$ we can sample boundary condition $\mathbf{y}_L$ from the distribution $\mathbb{P}_T$
- ... and then compute the finite distribution $\mathbb{P}[\mathbf{X}_T|\mathbf{X}_L = \mathbf{y}_L]$. Note: it is a random distribution.
- Mezard and Parisi: this random distribution is described by a fixed point $\mathcal{Q}\mu = \mu$
- Our contribution. We prove this ansatz: Iterations $\mathcal{Q} \circ \mathcal{Q} \cdots \mathcal{Q}\mu_0$ converge to a unique fixed point regardless of μ_0.





- Unsupervised clustering problem
- Input: graph
- Want: Label clusters

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color $X_v \sim \mathrm{Unif}[k]$.

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color $X_v \sim \mathrm{Unif}[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

$$\mathbb{P}[(u,v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } X_u = X_v \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color $X_v \sim \mathrm{Unif}[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

$$\mathbb{P}[(u, v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } X_u = X_v \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

- **Goal 1:** Design algorithms that recover **X** from $\mathbb{G}$.
- **Goal 2:** Characterize fundamental uncertainty $H(\mathbf{X}|\mathbb{G})$.

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a,b), $a,b>0$
- n vertices, each assigned a uniformly random color $X_v \sim \mathrm{Unif}[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

$$\mathbb{P}[(u,v)\in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } X_u = X_v \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

- **Goal 1:** Design algorithms that recover **X** from $\mathbb{G}$.
  Our work: for $k=2$ poly-time algo recovering optimal fraction of **X**
- **Goal 2:** Characterize fundamental uncertainty $H(\mathbf{X}|\mathbb{G})$.
  Our work: closes this as well.

Application 2: stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color $X_v \sim \mathrm{Unif}[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

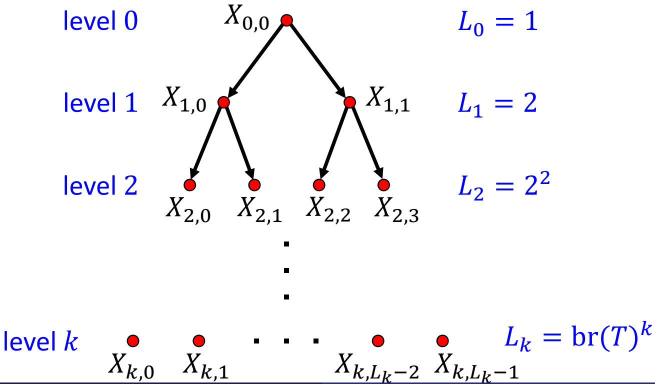
$$\mathbb{P}[(u,v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } X_u = X_v \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

- **Goal 1:** Design algorithms that recover $\mathbf{X}$ from $\mathbb{G}$.
  Our work: for $k = 2$ poly-time algo recovering optimal fraction of $\mathbf{X}$
- **Goal 2:** Characterize fundamental uncertainty $H(\mathbf{X}|\mathbb{G})$.
  Our work: closes this as well.
- Key fact: locally the recovery problem looks like BOT (next):

$$d \triangleq \deg(u) \sim \mathrm{Poi}(a+b), \qquad \delta \triangleq \mathbb{P}[X_u \neq X_v | u \sim v] = \frac{b}{a+b}$$

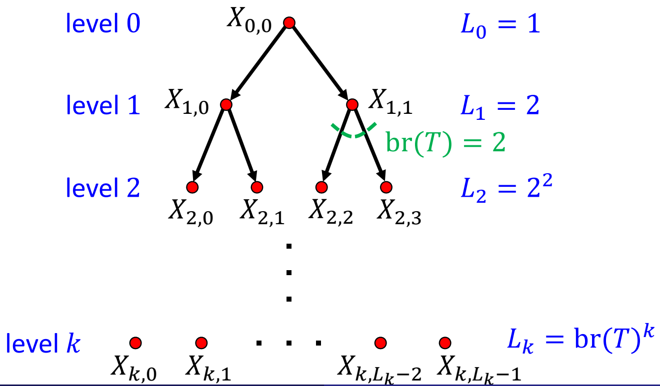
Application 3: Broadcasting on Trees

- Fix infinite oriented tree T with branching number $\mathrm{br}(T) := d$.



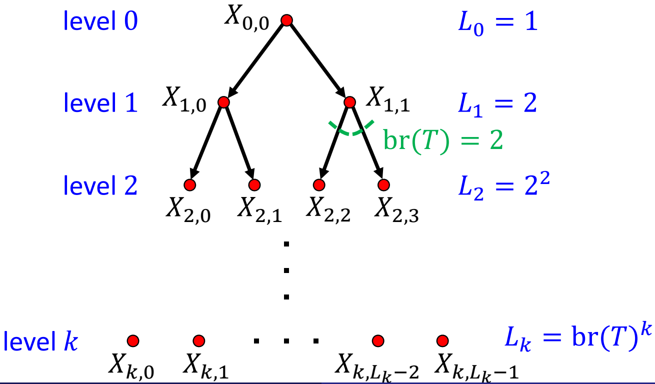
Application 3: Broadcasting on Trees

- Fix infinite oriented tree T with branching number $\mathrm{br}(T) := d$.



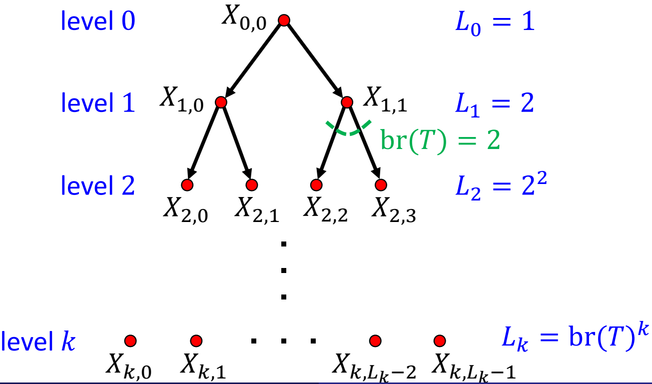
Application 3: Broadcasting on Trees

- Fix infinite oriented tree T with branching number $\mathrm{br}(T) := d$.
- Root $X_{0,0} \sim \mathrm{Unif}\{\pm 1\}$



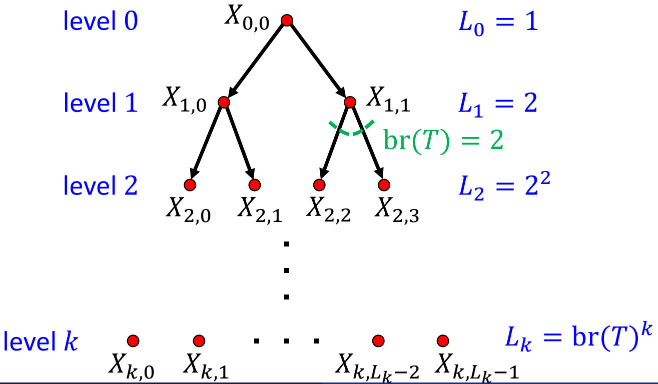
Application 3: Broadcasting on Trees

- Fix infinite oriented tree T with branching number $\mathrm{br}(T) := d$.
- Root $X_{0,0} \sim \mathrm{Unif}\{\pm 1\}$
- For any edge $u \to v$ set $X_v = \begin{cases} X_u, & \text{w.p. } 1-\delta \\ -X_u, & \text{w.p. } \delta \end{cases}$

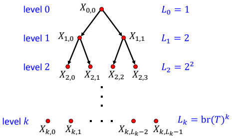


Application 3: Broadcasting on Trees

- Fix infinite oriented tree T with branching number $\mathrm{br}(T) := d$.
- Root $X_{0,0} \sim \mathrm{Unif}\{\pm 1\}$
- For any edge $u \to v$ set $X_v = \begin{cases} X_u, & \text{w.p. } 1-\delta \\ -X_u, & \text{w.p. } \delta \end{cases}$
- Goal: Reconstruct $X_{0,0}$ from $X_k = \left(X_{k,0}, \ldots, X_{k,d^k-1}\right)$.

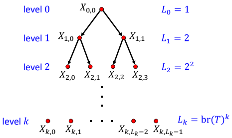


Broadcasting on Trees and BP fixed point



- Root variable $X_{0,0}$ is the information source
- It spreads along a tree of binary symmetric channels, BSC_δ.
- Question: How to estimate $X_{0,0}$ from a vector of far-away leaves X_k?

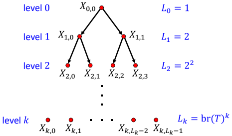
Broadcasting on Trees and BP fixed point



- Root variable $X_{0,0}$ is the information source
- It spreads along a tree of binary symmetric channels, BSC_δ.
- Question: How to estimate $X_{0,0}$ from a vector of far-away leaves X_k?
- A standard algorithm (belief propagation or BP) computes

$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$

Broadcasting on Trees and BP fixed point



- Root variable $X_{0,0}$ is the information source
- It spreads along a tree of binary symmetric channels, BSC_δ.
- Question: How to estimate $X_{0,0}$ from a vector of far-away leaves X_k?
- A standard algorithm (belief propagation or BP) computes

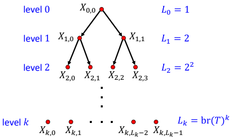
$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$

- Let μ_k denote the distribution of $R^{(k)}$ conditioned on $X_k = +1$. Then:

$$R^{(k+1)} = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,, \qquad R_i \overset{iid}{\sim} \mu$$

Thus: $R^{(k+1)} \sim \mu_{k+1} \triangleq \mathcal{Q}\mu_k$.

Broadcasting on Trees and BP fixed point



- Root variable $X_{0,0}$ is the information source
- It spreads along a tree of binary symmetric channels, BSC_δ.
- Question: How to estimate $X_{0,0}$ from a vector of far-away leaves X_k?
- A standard algorithm (belief propagation or BP) computes

$$R^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | X_k]}{\mathbb{P}[X_{0,0} = -1 | X_k]}$$

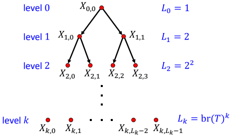
- Let μ_k denote the distribution of $R^{(k)}$ conditioned on $X_k = +1$. Then:

$$R^{(k+1)} = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,, \qquad R_i \overset{iid}{\sim} \mu$$

Thus: $R^{(k+1)} \sim \mu_{k+1} \triangleq \mathcal{Q}\mu_k$.

Can show that $R^{(k)} \xrightarrow{d} \mu^*$ – BP fixed point.

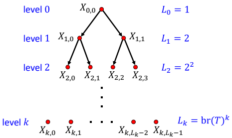
Broadcasting on Trees: Robust Variation



- How to estimate $X_{0,0}$ from the **noisy** observation Y_k of the leaves X_k?

$$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1 - \tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

Broadcasting on Trees: Robust Variation



- How to estimate $X_{0,0}$ from the **noisy** observation Y_k of the leaves X_k?

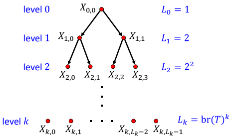
$$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1 - \tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

- As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | Y_k]}{\mathbb{P}[X_{0,0} = -1 | Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq \mathcal{Q}^k \tilde{\mu}_0$$

Broadcasting on Trees: Robust Variation



- How to estimate $X_{0,0}$ from the **noisy** observation Y_k of the leaves X_k?

$$\forall i \in \{1, \ldots, d^k - 1\} : \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1-\tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

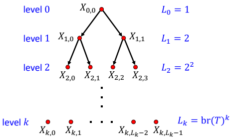
- As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | Y_k]}{\mathbb{P}[X_{0,0} = -1 | Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq \mathcal{Q}^k \tilde{\mu}_0$$

$\tilde{\mu}_0 = (1-\tau)\delta_{-c} + \tau\delta_c$, $c = \log \frac{\tau}{1-\tau}$.

Broadcasting on Trees: Robust Variation



- How to estimate $X_{0,0}$ from the **noisy** observation Y_k of the leaves X_k?

$$\forall i \in \{1, \ldots, d^k - 1\}: \quad Y_{k,i} = \begin{cases} X_{k,i}, & \text{w.p. } 1-\tau \\ -X_{k,i}, & \text{w.p. } \tau \end{cases}$$

- As before:

$$\tilde{R}^{(k)} \triangleq \log \frac{\mathbb{P}[X_{0,0} = +1 | Y_k]}{\mathbb{P}[X_{0,0} = -1 | Y_k]}$$

$$\tilde{R}^{(k)} \sim \tilde{\mu}_k \triangleq \mathcal{Q}^k \tilde{\mu}_0$$

$\tilde{\mu}_0 = (1-\tau)\delta_{-c} + \tau\delta_c$, $c = \log \frac{\tau}{1-\tau}$.

- Our main result: $\mathcal{Q}^k \tilde{\mu}_0 \to \tilde{\mu}_\infty$ and $\tilde{\mu}_\infty$ does not depend on τ (!!)

Proof ideas

Binary symmetric (BMS) channels

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- Example 1: BSC_δ channel is $Y = (-1)^{\mathsf{Bern}(\delta)} X$

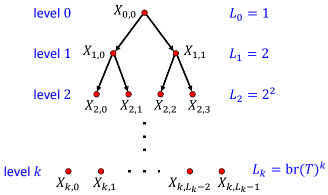
Binary symmetric (BMS) channels

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- Example 1: BSC_δ channel is $Y = (-1)^{\mathsf{Bern}(\delta)} X$
- Example 2: the tree channel $W_k : X_0 \to X_k$



Binary symmetric (BMS) channels

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- Example 1: BSC_δ channel is $Y = (-1)^{\mathsf{Bern}(\delta)} X$
- Example 2: the tree channel $W_k : X_0 \to X_k$

Binary symmetric (BMS) channels

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- Example 1: BSC_δ channel is $Y = (-1)^{\mathrm{Bern}(\delta)} X$
- Example 2: the tree channel $W_k : X_0 \to X_k$
- Blackwell preorder: We say that $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$

Binary symmetric (BMS) channels

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- Example 1: BSC_δ channel is $Y = (-1)^{\mathsf{Bern}(\delta)} X$
- Example 2: the tree channel $W_k : X_0 \to X_k$
- Blackwell preorder: We say that $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$

For example: $W_{k+1} \preceq W_k$

BMS $\leftrightarrow$ μ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- **Question:** How can we compactly represent a BMS?

BMS $\leftrightarrow \mu$ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- Sufficient statistic: Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X=+1]}{\mathbb{P}[Y|X=-1]}$$

BMS $\leftrightarrow \mu$ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- Sufficient statistic: Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X=+1]}{\mathbb{P}[Y|X=-1]}$$

- $\Rightarrow$ only need distribution of R under $X=+1$ and under $X=-1$.

BMS $\leftrightarrow \mu$ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- Sufficient statistic: Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X=+1]}{\mathbb{P}[Y|X=-1]}$$

- $\Rightarrow$ only need distribution of R under $X = +1$ and under $X = -1$.
- ... But due to symmetry only one is needed!

$$\mu \triangleq \text{Law of } R \quad \text{under } X = +1\,.$$

BMS $\leftrightarrow$ μ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- **Question:** How can we compactly represent a BMS?
- Sufficient statistic: Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X=+1]}{\mathbb{P}[Y|X=-1]}$$

- $\Rightarrow$ only need distribution of R under $X = +1$ and under $X = -1$.
- ... But due to symmetry only one is needed!

$$\mu \triangleq \text{Law of } R \ \text{ under } X = +1\,.$$

Tech note: $\mu(-dr) = e^{-r}\mu(dr)$ and all μ's in this work are such!

BMS $\leftrightarrow \mu$ equivalence

Definition (BMS channel)

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

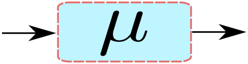
- **Question:** How can we compactly represent a BMS?
- Sufficient statistic: Log-likelihood ratio

$$R = \log \frac{\mathbb{P}[Y|X=+1]}{\mathbb{P}[Y|X=-1]}$$

- $\Rightarrow$ only need distribution of R under $X = +1$ and under $X = -1$.
- ... But due to symmetry only one is needed!

$$\mu \triangleq \text{Law of } R \text{ under } X = +1 \,.$$

Tech note: $\mu(-dr) = e^{-r}\mu(dr)$ and all μ's in this work are such!

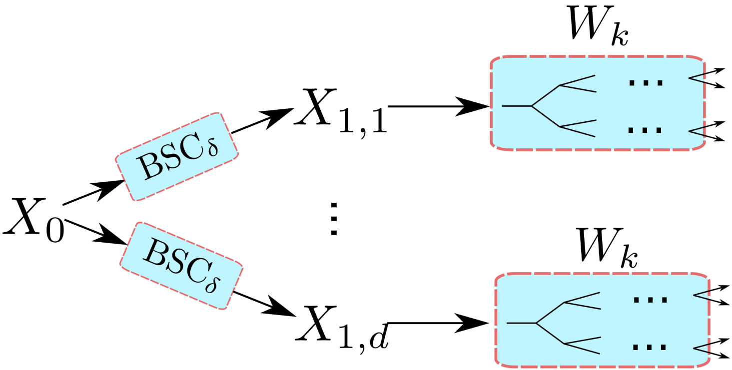
- Denote a BMS corresp. to μ as 

Tree channels recursion

In BOT we can build W_{k+1} channel from W_k and BSC_δ:

Tree channels recursion

In BOT we can build W_{k+1} channel from W_k and BSC_δ:



Understanding $\mathcal{Q}\mu$ graphically

Recall: BP-operator $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \mathrm{Bern}(\delta), S_0 \sim \mu_0$$

Understanding $\mathcal{Q}\mu$ graphically

Recall: BP-operator $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \mathrm{Bern}(\delta), S_0 \sim \mu_0$$

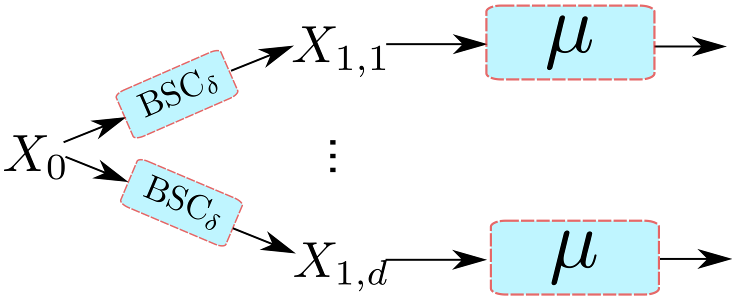
We can understand the BMS $\mathcal{Q}\mu$ graphically as:

Understanding $\mathcal{Q}\mu$ graphically

Recall: BP-operator $\mathcal{Q}\mu =$ the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0 \,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

We can understand the BMS $\mathcal{Q}\mu$ graphically as:

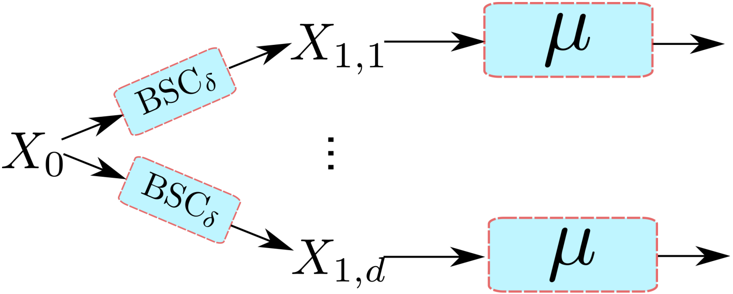


Understanding $\mathcal{Q}\mu$ graphically

Recall: BP-operator $\mathcal{Q}\mu$ = the law of

$$R = \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i) + S_0\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta), S_0 \sim \mu_0$$

We can understand the BMS $\mathcal{Q}\mu$ graphically as:



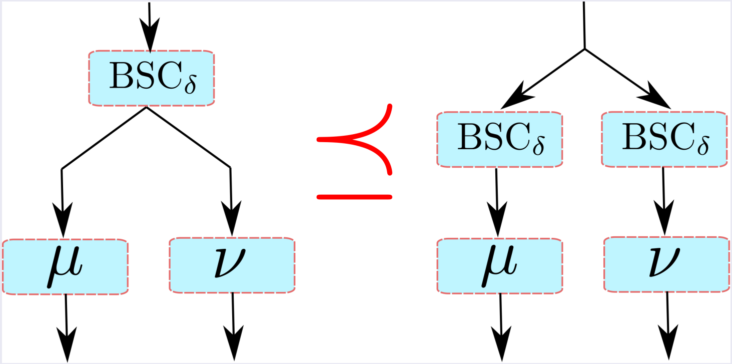
Infinite divisibility: a fixed point channel has property of not changing upon d-fold copying of its δ-noisy version.

With these preparations, we can appreciate the following result [EKPS'2000].

With these preparations, we can appreciate the following result [EKPS'2000].

Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman'2000))

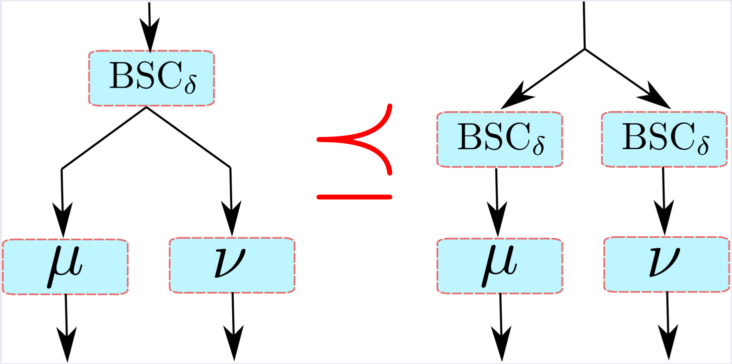
For all μ and ν we have:



With these preparations, we can appreciate the following result [EKPS'2000].

Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman'2000))

For all μ and ν we have:

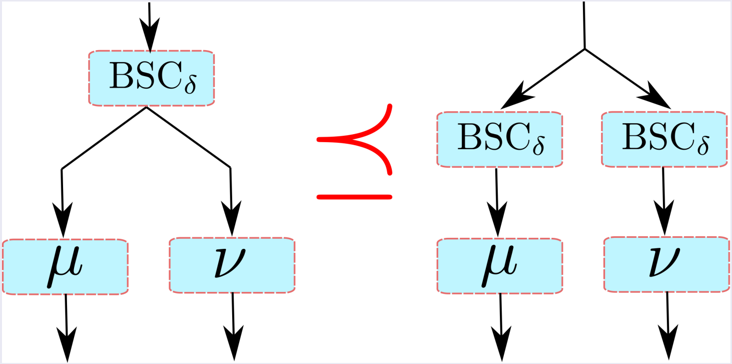


Copying input bit before adding noise results in a better channel.

With these preparations, we can appreciate the following result [EKPS'2000].

Theorem (Stringy tree lemma (Evans-Kenyon-Peres-Schulman'2000))

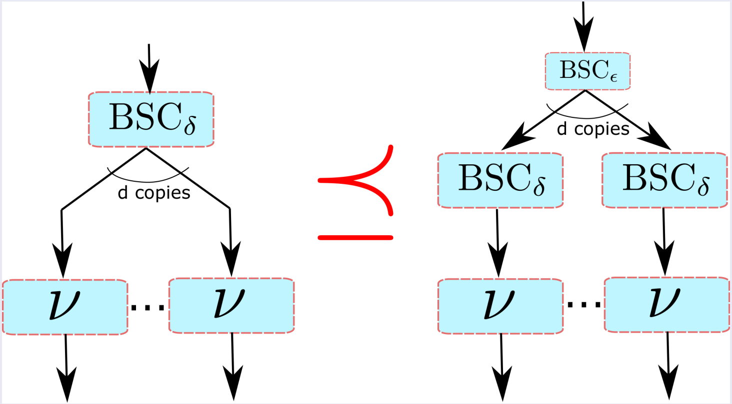
For all μ and ν we have:



Copying input bit before adding noise results in a better channel.
Applying iteratively $1 - 2\delta_k = (1 - 2\delta)^k$ [EKPS'2000] bound W_k and showed $\mathcal{Q}^k \delta_\infty \to \delta_0$ if $(1 - 2\delta)^2 d < 1$.

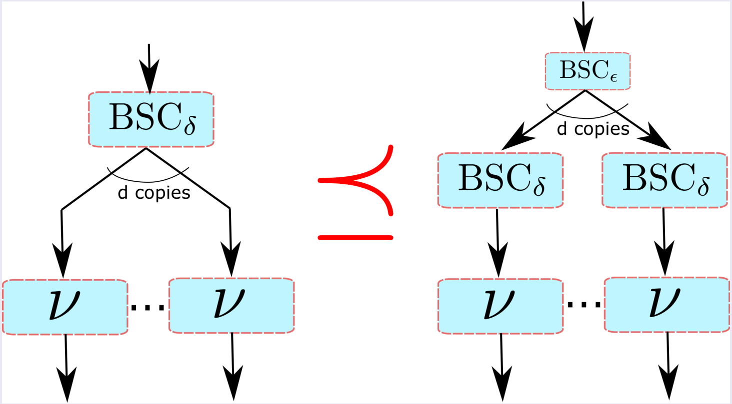
Theorem (Main technical discovery of Yu-P.'2022)

For every ν, $\delta \in (0,1)$ *and* $d \geq 3$ *there exists* $\epsilon > 0$ *such that:*



Theorem (Main technical discovery of Yu-P.'2022)

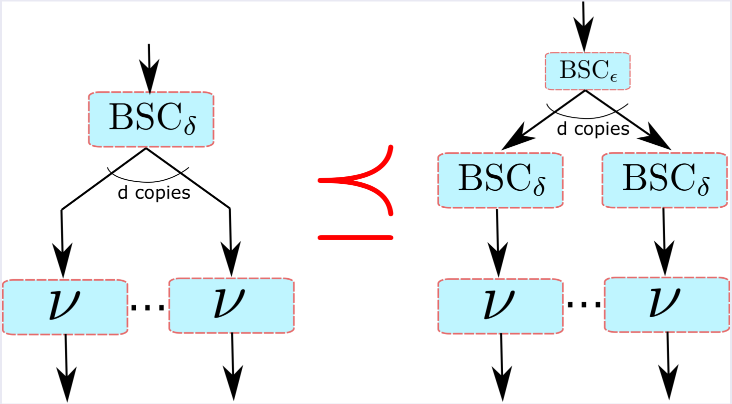
For every ν, $\delta \in (0,1)$ *and* $d \geq 3$ *there exists* $\epsilon > 0$ *such that:*



For $d = 2$ this is wrong, but holds for a depth-2 (binary) tree.

Theorem (Main technical discovery of Yu-P.'2022)

For every ν, $\delta \in (0,1)$ and $d \geq 3$ there exists $\epsilon > 0$ such that:



For $d = 2$ this is wrong, but holds for a depth-2 (binary) tree.

Let us see how this result implies unicity of BP-fixed point.

Proof of the main result I

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathsf{BSC}_\phi$).

Proof of the main result I

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \mathsf{BSC}_\phi)$.
- Define degradation index

$$\phi^*(\mu,\nu) = \inf\{\phi : \mathcal{B}_\phi \mu \preceq \nu\}\,.$$

$=$ least amount of noise needed to make μ worse than ν.

Proof of the main result I

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \mathrm{BSC}_\phi)$.
- Define degradation index

$$\phi^*(\mu,\nu) = \inf\{\phi : \mathcal{B}_\phi \mu \preceq \nu\}\,.$$

  = least amount of noise needed to make μ worse than ν.
- If $\phi^* > 0$ then (we will show next) our improved STL implies for some $\epsilon > 0$

$$\mathcal{B}_{\phi^*} \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$

Proof of the main result I

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \mathsf{BSC}_\phi)$.
- Define degradation index

$$\phi^*(\mu,\nu) = \inf\{\phi : \mathcal{B}_\phi \mu \preceq \nu\}\,.$$

  = least amount of noise needed to make μ worse than ν.
- If $\phi^* > 0$ then (we will show next) our improved STL implies for some $\epsilon > 0$

$$\mathcal{B}_{\phi^*} \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$

- So if μ and ν are two fixed points, then

$$\mathcal{B}_{\phi^*} \mu \preceq \mathcal{B}_\epsilon \nu\,,$$

  violating minimality of ϕ^*

  Here we also need ν – non-trivial, so that $\phi^* < 1/2$

Proof of the main result I

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Define degradation index

$$\phi^*(\mu, \nu) = \inf\{\phi : \mathcal{B}_\phi \mu \preceq \nu\}\,.$$

  = least amount of noise needed to make μ worse than ν.
- If $\phi^* > 0$ then (we will show next) our improved STL implies for some $\epsilon > 0$

$$\mathcal{B}_{\phi^*} \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$

- So if μ and ν are two fixed points, then

$$\mathcal{B}_{\phi^*} \mu \preceq \mathcal{B}_\epsilon \nu\,,$$

  violating minimality of ϕ^*

  Here we also need ν – non-trivial, so that $\phi^* < 1/2$
- So $\phi^* = 0$, which implies $\mu \preceq \nu$. From symmetry $\nu = \mu$ then.

Proof of the main result II

- Define $\mathcal{B}_\phi\mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \mathsf{BSC}_\phi)$.

Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathsf{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

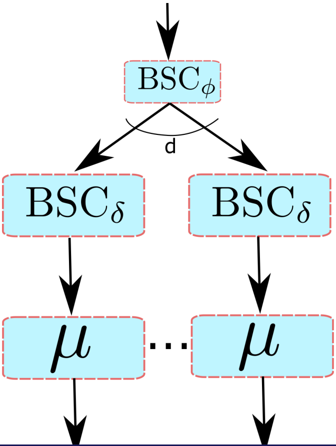
$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu \,.$$

Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$

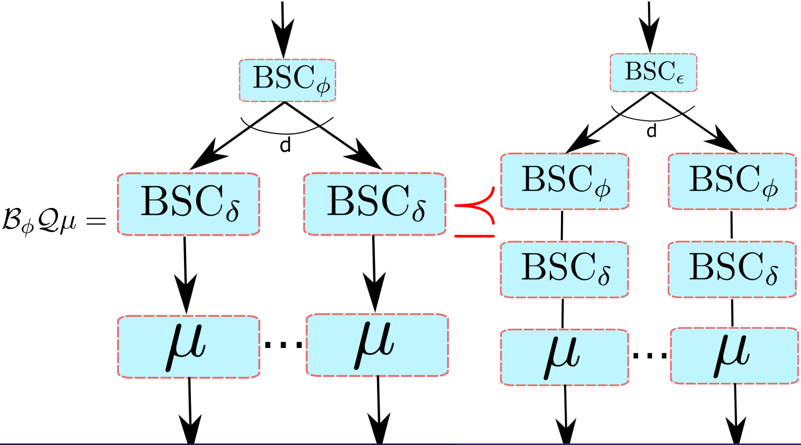
$\mathcal{B}_\phi \mathcal{Q}\mu =$



Proof of the main result II

- Define $\mathcal{B}_\phi\mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel $(\mu \circ \mathrm{BSC}_\phi)$.
- Remains to show: If $\mathcal{B}_\phi\mu \preceq \nu$ then $\exists\epsilon > 0$

$$\mathcal{B}_\phi\mathcal{Q}\mu \preceq \mathcal{B}_\epsilon\mathcal{Q}\nu\,.$$

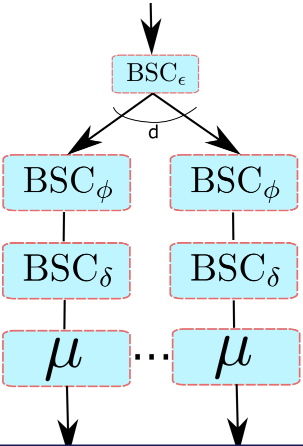


Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu \,.$$

$\mathcal{B}_\phi \mathcal{Q}\mu \preceq$

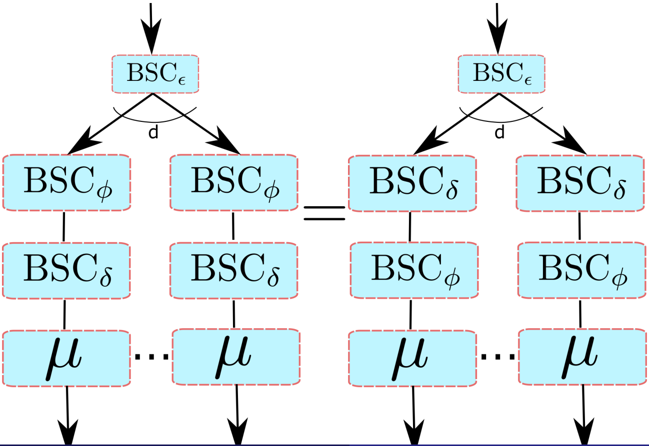


Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$

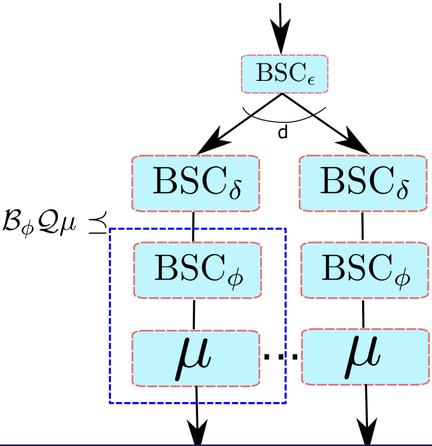
$\mathcal{B}_\phi \mathcal{Q}\mu \preceq$



Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

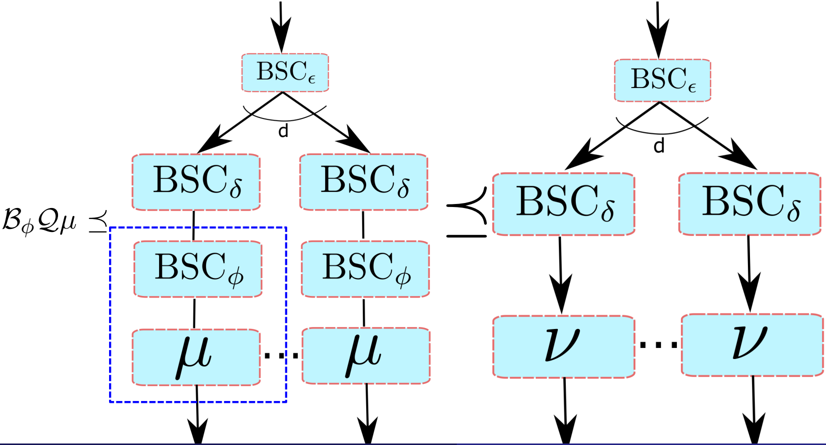
$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu \,.$$



Proof of the main result II

- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

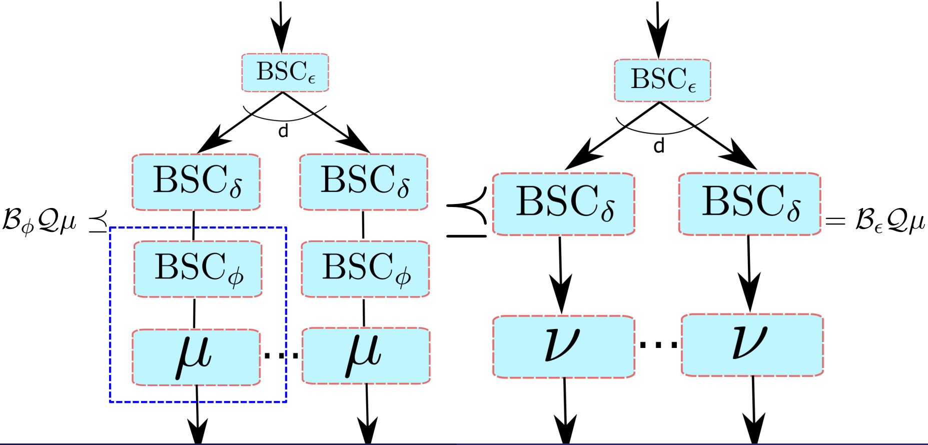
$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu \,.$$

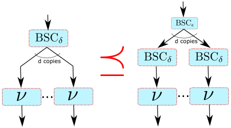


Proof of the main result II

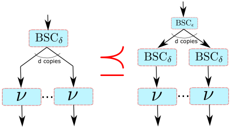
- Define $\mathcal{B}_\phi \mu(dr) = (1-\phi)\mu(dr) + \phi\mu(-dr)$.
  Note: this is the law of channel ($\mu \circ \mathrm{BSC}_\phi$).
- Remains to show: If $\mathcal{B}_\phi \mu \preceq \nu$ then $\exists \epsilon > 0$

$$\mathcal{B}_\phi \mathcal{Q}\mu \preceq \mathcal{B}_\epsilon \mathcal{Q}\nu\,.$$



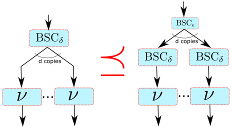


- One word about the proof of improved STL



- One word about the proof of improved STL
- Blackwell preorder: $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$



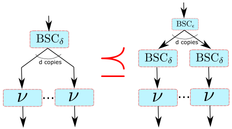
- One word about the proof of improved STL
- Blackwell preorder: $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$

- In [EKPS'2000] to prove STL authors constructed the coupling explicitly
- Idea 1: Our trick was to avoid the (hard) job of constructing $P_{Y|Z}$:

$$\mu \preceq \nu \iff \beta(t;\mu) \le \beta(t;\nu) \qquad \forall t \ge 0\,,$$

where $\beta(t;\mu) = \mathbb{E}_{R\sim\mu}[\tanh \frac{|R|}{2} \vee t]$



- One word about the proof of improved STL
- Blackwell preorder: $P_{Y|X} \preceq P_{Z|X}$ if there is $P_{Y|Z}$ s.t.

$$P_{Y|X}(y|x) = \sum_z P_{Z|X}(z|x) P_{Y|Z}(y|z) \quad \iff \quad X \xrightarrow{P_{Y|X}} Z \xrightarrow{P_{Y|Z}} Y$$

- In [EKPS'2000] to prove STL authors constructed the coupling explicitly
- Idea 1: Our trick was to avoid the (hard) job of constructing $P_{Y|Z}$:

$$\mu \preceq \nu \iff \beta(t;\mu) \le \beta(t;\nu) \qquad \forall t \ge 0\,,$$

where $\beta(t;\mu) = \mathbb{E}_{R\sim\mu}[\tanh \frac{|R|}{2} \vee t]$

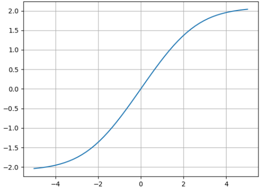
- Idea 2: Any μ is a mixture of elementary ones:

$$\mathcal{B}_\tau = (1-\tau)\delta_{-c} + \tau\delta_c, \qquad c = \log\frac{\tau}{1-\tau}\,.$$

So key inequalities are checked for BSC$_\tau$'s in place of generic ν.

Conclusion

$$F_\delta(x) =$$



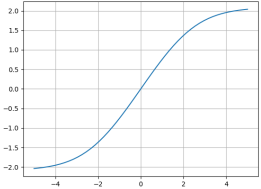
$$\mathcal{Q}\mu \triangleq \text{Law of} \quad \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta)$$

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* of $\mathcal{Q}$.

- Unusual method based on (new) properties of channel compositions.

Conclusion

$F_\delta(x) =$ 

$$\mathcal{Q}\mu \triangleq \text{Law of} \quad \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,, \qquad R_i \stackrel{iid}{\sim} \mu, X_i \stackrel{iid}{\sim} \text{Bern}(\delta)$$

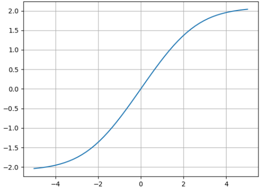
Theorem (Main result)

There exists at most one non-trivial fixed point μ^* of $\mathcal{Q}$.

- Unusual method based on (new) properties of channel compositions.
- We construct metric $d(\cdot,\cdot)$ s.t. $d(\mathcal{Q}\mu, \mathcal{Q}\nu) < d(\mu,\nu)$ unless both are trivial. This implies $\mathcal{Q}^k \mu_0 \to \mu^*$.

Conclusion

$$F_\delta(x) =$$



$$\mathcal{Q}\mu \triangleq \text{Law of} \quad \sum_{i=1}^{d} (-1)^{X_i} F_\delta(R_i)\,, \qquad R_i \overset{iid}{\sim} \mu, X_i \overset{iid}{\sim} \text{Bern}(\delta)$$

Theorem (Main result)

There exists at most one non-trivial fixed point μ^* of $\mathcal{Q}$.

- Unusual method based on (new) properties of channel compositions.
- We construct metric $d(\cdot,\cdot)$ s.t. $d(\mathcal{Q}\mu, \mathcal{Q}\nu) < d(\mu,\nu)$ unless both are trivial. This implies $\mathcal{Q}^k \mu_0 \to \mu^*$.
- **Future work:** Extending to Potts (q-ary) models.

Thank You!

The draft is available here:
https://www.mit.edu/~ypol