Application of BP for optimization in structured spaces

Rajesh Sundaresan

Indian Institute of Science, Bangalore

NASIT 2018 (Texas A&M), May 2018
U Michigan (Ann Arbor), June 2018

Outline

- Basics of BP and BP for optimisation.
- BP for the assignment problem.
- Steps involved in making it rigorous.
- Other problems. Edge cover, traveling salesman problem, many-to-one matchings, etc.

Belief Propagation (BP)

- An *iterative* and *local* algorithm for computing the *marginal probabilities* of a *graphical probability model*

- Our interest is in probability models on n variables, denoted $x = (x_1, \ldots, x_n)$, with a certain dependence structure.

$$p(x_1, \ldots, x_n) = Z^{-1} \prod_{a \in F} Q_a(x_a).$$

- $Q_a(x_a)$ is a *factor* indexed by a subset $a \subseteq \{1, \ldots, n\}$ and involves the variables $x_a := (x_i, i \in a)$.

- F is the index set of factors, Z is a normalisation.

- Factors specify the dependence structure. Assumed known.

- Also called a graphical model or a Markov random field.

Markov chain

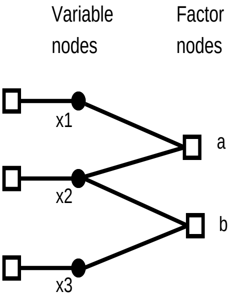
$$p(x_1, \ldots, x_n) = Q_1(x_1) \prod_{i=2}^{n} Q_{i,i-1}(x_i, x_{i-1}).$$

- Factor indices: $\{1\}, \{i, i-1\}_{i \geq 2}$.
- $Q_1(x_1)$ is the initial distribution.
- $Q_{i,i-1}(x_i, x_{i-1})$ is the transition probability matrix for the ith transition, more commonly written as $Q_{i|i-1}(x_i|x_{i-1})$.
- $Z = 1$.

Graphical model and marginal probabilities

- Example. Take $n = 3$. Each x_i is binary. Suppose:

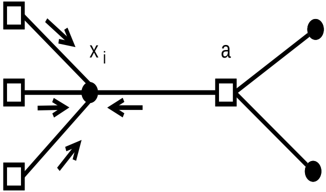
$$p(x_1, x_2, x_3) \quad \propto \quad \overbrace{Q_1(x_1) \cdot Q_2(x_2) \cdot Q_3(x_3)}^{\text{initial beliefs}} \cdot \overbrace{\mathbf{1}\{x_1 = x_2\}}^{a} \cdot \overbrace{\mathbf{1}\{x_2 = x_3\}}^{b}$$



- This is a "factor graph" representation of the model, with variable and factor nodes.

- Goal: compute the marginal probability $p(x_1)$.

Introducing BP

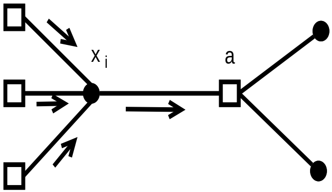


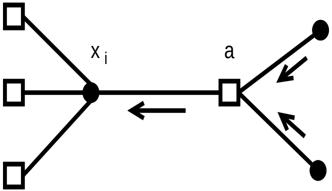
- If there were no variable nodes but x_i, by a suitable renormalisation, we can think of Q_a as probability distributions. Factor a's "opinion" on x_i's distribution.
- Then each factor imposes an "external field" on x_i, and we get the marginal as a "compromise":

$$p(x_i) = Z^{-1} \prod_{a \in F} Q_a(x_i)$$

- When there are other variable nodes, each factor node should convey the "effective" external field it will impose on x_i.

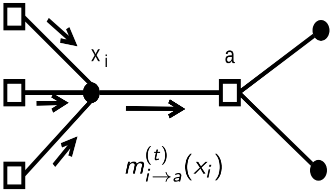
Introduce a cavity in the system

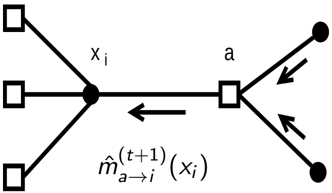




- Removing factor a and its associated edges breaks this graph into three components.
- Compute the associated variable node distributions, separately, on each component and pass them to the removed factor node along the corresponding removed edge.
- Then make the factor node pass, to x_i, its belief about x_i based on what's imposed by the other components.
- Do this repeatedly, and we have the BP algorithm.

BP : sum-product algorithm





The messages are distributions or beliefs. $y_a = ((y_{i'}, i' \in a, i' \neq i), y_i)$.

$$\text{Factor node} \quad : \quad \hat{m}^{(t+1)}_{a\to i}(x_i) = Z^{-1} \cdot \sum_{y_a : y_i = x_i} Q_a(y_a) \prod_{i' \sim a, i' \neq i} m^{(t)}_{i' \to a}(y_{i'}).$$

$$\text{Variable node} \quad : \quad m^{(t)}_{i\to a}(x_i) = Z^{-1} \cdot \prod_{a' \sim i, a' \neq a} \hat{m}^{(t)}_{a' \to i}(x_i).$$

$$\text{Marginal} \quad : \quad p^{(t)}(x_i) = Z^{-1} \cdot \prod_{a \sim i} \hat{m}^{(t)}_{a\to i}(x_i).$$

Three natural questions

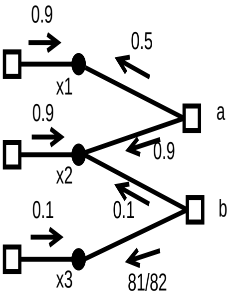
- Does the algorithm converge?
- Does it produce the correct answer?
- How many iterations?

BP works on trees

Theorem

On a tree of diameter d, BP converges after at most d steps to yield the correct marginals.

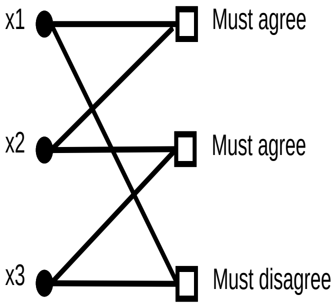
For our initial example ...



Converged marginal: $p(x_1 = 1) = 0.9$.

Problems

- Loops.



Locally consistent marginals, a belief of 0.5 for each, but these cannot be the marginals of any global probability distribution.

- Infinite trees. Nodes very far off, at infinity, may affect the marginal at a given node.

BP for optimisation

- Suppose we want to find the maximum-likelihood configuration:

$$x^* = \arg\max_x p(x).$$

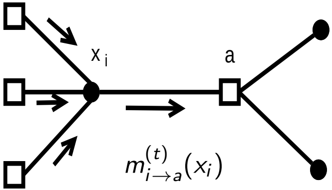
- Suppose we are able to compute max-marginals:

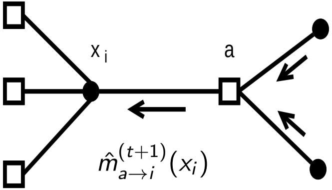
$$M_i(x_i) = \max_{y: y_i = x_i} p(y).$$

- Procedure to find ML configuration:
  - Find $M_1(\cdot)$. Find x_1^*.
  - New graphical model with $x_1 = x_1^*$. Compute max-marginals $M_2(\cdot)$. Find x_2^*.
  - ...

- So it suffices to compute max-marginals.
  How can BP be modified to do this?

Max-product algorithm





$$\text{Factor node} \quad : \quad \hat{m}_{a\to i}^{(t+1)}(x_i) = Z^{-1}\cdot \max_{y_a : y_i = x_i}\left[Q_a(y_a)\prod_{i'\sim a, i'\neq i} m_{i'\to a}^{(t)}(y_{i'})\right].$$

$$\text{Variable node} \quad : \quad m_{i\to a}^{(t)}(x_i) = Z^{-1}\cdot \prod_{a'\sim i, a'\neq a}\hat{m}_{a'\to i}^{(t)}(x_i).$$

$$\text{Max-marginal} \quad : \quad M^{(t)}(x_i) = Z^{-1}\cdot \prod_{a\sim i}\hat{m}_{a\to i}^{(t)}(x_i).$$

BP works on trees, again

Theorem

On a tree of diameter d, the max-product updates converge after at most d steps to yield the correct max-marginals (upto a scale factor).

But same issues as before - cycles, infinite trees.

The min-sum algorithm and the energy cavity equations

- The log transformation: $E_a(x_a) := -\frac{1}{\beta}\log Q_a(x_a)$.
- By writing the factors $Q_a(x_a) = e^{-\beta E_a(x_a)}$, we see that

$$p(x) = e^{-\beta \sum_{a\in F} E_a(x_a)}$$

- Maximum likelihood configuration is the one that minimises the "cost" or "energy" function:

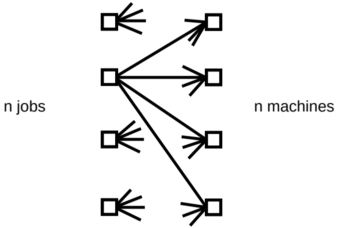
$$E(x) := \sum_{a\in F} E_a(x_a)$$

  Ground state.
- Replace beliefs by negative log-beliefs in the BP equations, and one gets what is known as the min-sum algorithm. The associated BP updates are called *energy cavity equations*.

Thus far ...

- Graphical models and factor graphs
- BP for marginals. The sum-product algorithm (via cavity)
- Works on trees. Questions when there are loops or the graph is infinite.
- BP for ML. The max-product algorithm
- BP for ML. The min-sum algorithm and energy cavity equations.

BP for optimisation : optimal assignment



- C_{ij} is cost of running job i on machine j.
- Goal: Each machine can take at most one job. Assign each job to a machine so that total cost is minimized.
- Minimum weight perfect matching on the weighted $K_{n,n}$. Solvable in (worst-case) $O(n^3)$ steps.
- On random instances, BP finds a near optimal solution with high probability in $O(n^2)$ steps. Each node executes only $O(n)$ steps.

The history of the assignment problem

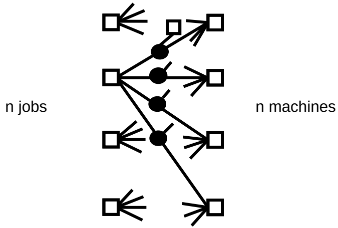
- Active since the 1930s. König (1931). Randomised setting since the 1960s. Kurtzberg (1962), Walkup (1979), Karp (1987), Goemans and Kodialam (1989).
- 1987. Mezard and Parisi showed via a nonrigorous method that the expected cost of minimum matching is $\zeta(2)$.
- 1992. Aldous showed that a limit exists.
- 2001. Aldous gave a rigorous proof that limit is $\zeta(2)$.
- 2005. Aldous and Bandopadhyay on "recursive distributional equations".
- 2009. Salez and Shah on BP.

Relaxed assignment: the factor graph

- Variable a_{ij}: 1 if job i assigned to machine j, 0 otherwise

$$p(\{a_{ij}\}) \quad \propto \quad \prod_{i,j} e^{-\beta a_{ij}(C_{ij}-2\gamma)} \cdot \prod_i \mathbf{1}\left\{\sum_{j'} a_{ij'} \leq 1\right\} \cdot \prod_j \mathbf{1}\left\{\sum_{i'} a_{i'j} \leq 1\right\}$$

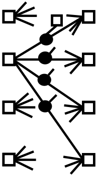
- As $\gamma \to \infty$, mass concentrates on perfect matchings
  As $\beta \to \infty$, mass further concentrates on minimum cost perfect matchings.



- Variable nodes indexed by ij. Factor nodes indexed by i, j, and ij.

- Goal: Sample from the distribution, or find mode (for large γ and β).

BP equations (sum-product)



- Message from right to left:

  Variable node:

$$m_{ij\to i}(a_{ij}) = Z^{-1} \cdot \hat{m}_{j\to ij}(a_{ij}) \cdot e^{-\beta a_{ij}(C_{ij}-2\gamma)}.$$

  Machine factor node:

$$\hat{m}_{j\to ij}(a_{ij}) = Z^{-1} \cdot \sum_{\{a_{i'j}\}_{i':i'\neq i}} \mathbf{1}\left\{a_{ij} + \sum_{i':i'\neq i} a_{i'j} \leq 1\right\} \cdot \prod_{i':i'\neq i} m_{i'j\to j}(a_{i'j}).$$

- Similarly for message from left to right.
- Some simplification is possible.
  - Variable node updates involve only one nontrivial factor node.
  - Work with log-likelihoods.

BP equations after simplification

Define: $\phi_{j\to i}$ as below, and $\phi_{i\to j}$ similarly.

$$\phi_{j\to i} := \gamma + \frac{1}{\beta}\log\left(\frac{\hat{m}_{j\to ij}(a_{ij}=1)}{\hat{m}_{j\to ij}(a_{ij}=0)}\right).$$

The BP equations simplify to the following.

- Left to right:

$$\phi_{i\to j} = -\frac{1}{\beta}\log\left[e^{-\beta\gamma} + \sum_{j':j'\neq j} e^{\beta(-C_{ij'}+\phi_{j'\to i})}\right]$$

- Right to left:

$$\phi_{j\to i} = -\frac{1}{\beta}\log\left[e^{-\beta\gamma} + \sum_{i':i'\neq i} e^{\beta(-C_{i'j}+\phi_{i'\to j})}\right]$$

The zero temperature limit

- Let $\gamma \to \infty$ first and then $\beta \to \infty$, we get:

$$
\begin{aligned}
\phi_{i \to j} &= \min_{j' : j' \neq j} \left[C_{ij'} - \phi_{j' \to i} \right] \\
\phi_{j \to i} &= \min_{i' : i' \neq i} \left[C_{i'j} - \phi_{i' \to j} \right]
\end{aligned}
$$

- Proposal:
  - Run the BP iterations as above until convergence.
  - Interpret the converged values to put out the matching. Each job i is matched to the minimising machine, i.e.,

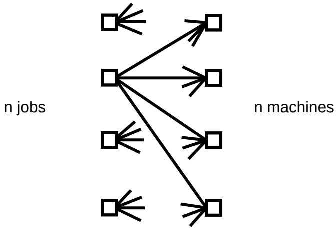
$$
\pi(i) = \arg\min_j \left[C_{ij} - \phi_{j \to i} \right]
$$

- The factor graph is full of loops, and our proposal is full of holes.

Hope in an ensemble viewpoint

- Random costs: $\{C_{ij}\}$ are independent with identical distribution, e.g., Uniform[0,1]
    - Beliefs, cavity variables, etc., are now random variables; they depend on the realisation $\{C_{ij}\}$
- What is the expected cost of the minimum weight matching?
- Further, let network size $n \to \infty$
- What is the *limiting* expected cost of the minimum weight matching?
- We have thrown in more complications. But there is hope in this random infinite setting.

Loops disappear in an appropriate topology



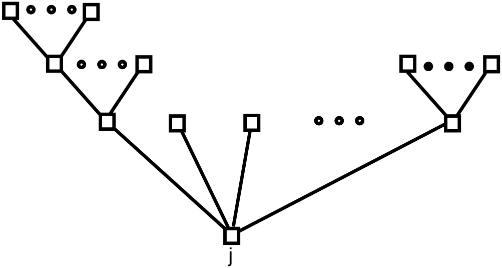
- C_{ij} independent and Uniform[0,1]

- From a typical job i's perspective, typical costs are $O(1)$; but
$$E\left[\min_j C_{ij}\right] = \frac{1}{n+1} = O\left(\frac{1}{n}\right)$$

- Only links with cost $O(1/n)$ matter

Locally tree-like

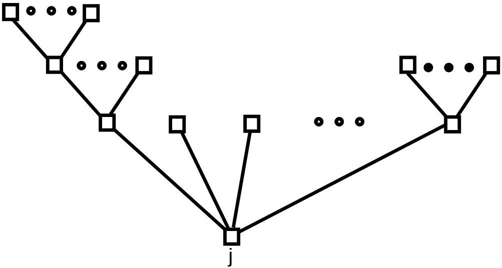
- Erase all links that cost more than, say, $10000/n$
- The picture from a typical node, after re-scaling of surviving links



- Loops disappear in the scale of interest

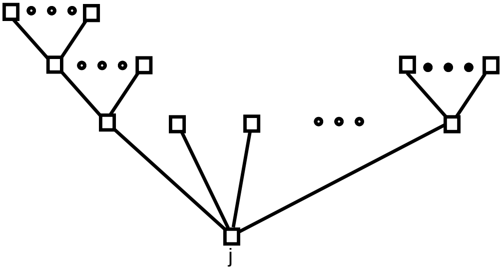
Locally tree-like on the scaled graph

- Alternatively, scale all link costs by n. E.g., Uniform $[0, n]$
- Erase all links that cost more than, this time, $\rho = 10000 = O(1)$
- The picture from a typical node



- Loops disappear when graph distances of only $O(1)$ are considered
- More precisely, $\Pr\{\text{there is no cycle of length } \leq \rho\} = 1 - O(1/n)$

What about number of neighbours of the root?



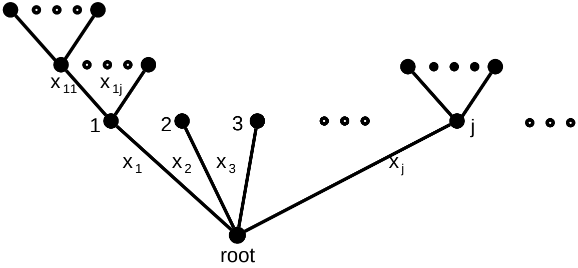
- Number of one-hop neighbours within distance ρ:

$$\sum_{i=1}^{n} \mathbf{1}\{nC_{ji} \leq \rho\} = \mathrm{Bin}(n, \rho/n) \to \mathrm{Poi}(\rho)$$

Local weak limit that describes the local neighbourhood

Theorem

The local neighbourhood from a typical node, on $K_{n,n}$ with weights scaled by n, has a limiting distribution identical to local neighbourhood of root on the Poisson Weighted Infinite Tree (PWIT).



The weights $x_1, x_2, \cdots$ are points of a unit rate PPP. Similarly, independent unit rate PPP at each descendent node.

This notion of convergence is called *local weak convergence*.

Thus far ...

- BP for optimisation.
  Want ground states or minimum energy configurations.
  Relaxation is to study configuration distribution at positive temperature.

- Assignment problem, BP iterates, and the cavity equations.

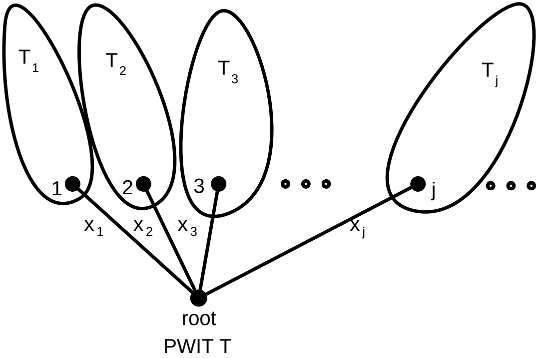
- Cavity equations at zero temperature.

- There are issues related to correctness. Our hope is in an ensemble view point.

- Loops disappear from a local perspective in the $O(1)$ scale. A locally tree-like structure emerges.

- Local weak limit is a Poisson Weighted Infinite Tree (PWIT).

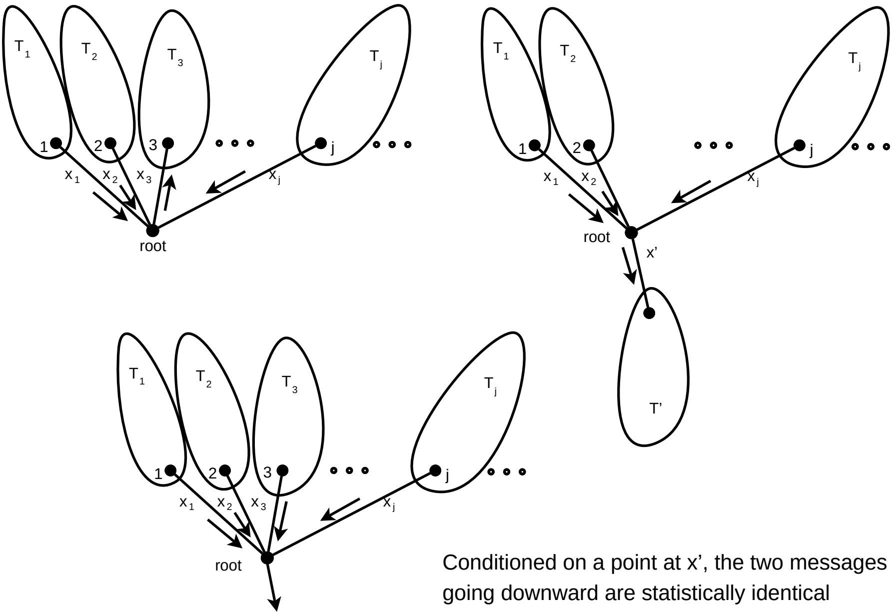
Look for symmetries



- Each of the subtrees $T_1, T_2, \ldots$ are identically distributed, with distribution identical to that of T.

- The distributions of $T_1, T_2, \ldots$ are independent.

The message going downward

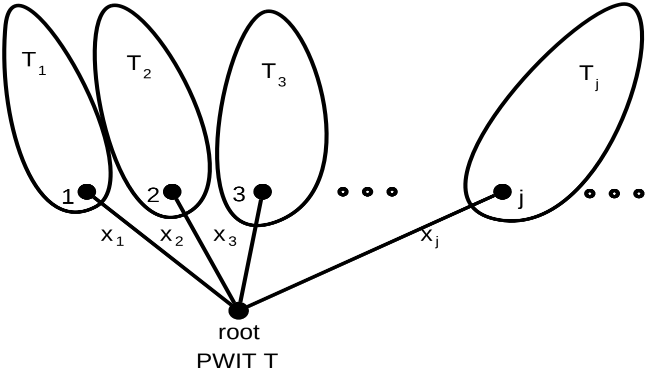


Conditioned on a point at x', the two messages going downward are statistically identical

Solve the problem on the PWIT by exploiting symmetry

- The cavity equations on the PWIT are:

$$\phi_{root} = \min_j (x_j - \phi_j) \, .$$



- Symmetry: ϕ_j are iid, and equal in distribution to ϕ_{root}.
- A recursive distributional equation (RDE).

Recursive distributional equation (RDE)

- Let $\phi_1, \phi_2, \ldots$ be iid $\sim F$.
- Let $x_1, x_2, \ldots$ be points of a unit rate PPP.
- The distribution of $\phi_{root} = \min_j \{x_j - \phi_j\}$ is also F.
- RDE : $\phi \stackrel{D}{=} \min_j \{x_j - \phi_j\}$.

Theorem

The unique solution to the above RDE is the logistic distribution $F(t) = 1/(1 + e^{-t})$.

Solving the RDE $\phi \stackrel{D}{=} \min_j (x_j - \phi_j)$

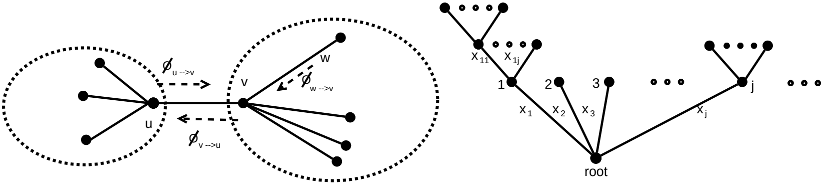
- Let F be the cdf of ϕ. Then $1 - F(t) = \Pr\{\min_j(x_j - \phi_j) > t\}$
- (x_j, ϕ_j) are points in $\mathbb{R}_+ \times \mathbb{R}$ of a Poisson process $\mathcal{P}$ with intensity $dx \times dF(\varphi)$.
- $\phi_{root} > t \iff$ no point in the set $A := \{(x, \varphi) : x - \varphi \leq t\}$.

$$\begin{aligned} 1 - F(t) = \Pr\{\text{no points in } A\} &= \exp\left\{-\int_0^\infty \int_{x-\varphi\leq t} dx dF(\varphi)\right\} \\ &= \exp\left\{-\int_0^\infty dx\ (1 - F(x-t))\right\} \\ &= \exp\left\{-\int_{-t}^\infty dx\ (1 - F(x))\right\} \end{aligned}$$

- Differentiate to get $F'(t) = (1 - F(-t))(1 - F(t))$.
- By symmetry of $F'(t) = F(t)(1 - F(t))$.

Solution: $F(t) = 1/(1 + e^{-t})$, logistic distribution

Recursive tree process

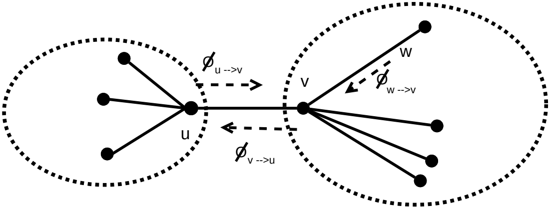


- With an explicit solution to the RDE, we can construct a tree process of the ϕ's on the PWIT

- The following holds on every directed edge:

$$\phi_{v \to u} = \min\{x_{v,w} - \phi_{w \to v},\ w \neq v, w \sim v\}$$

Finding a matching on the recursive tree process



- Match v to u if

$$x_{u,v} - \phi_{u \to v} = \min\{x_{w,v} - \phi_{u \to v},\ w \sim v\}$$

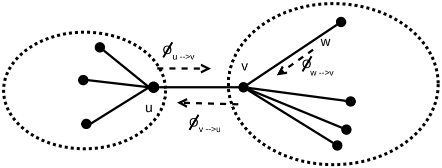
- This is equivalent to matching v to the u that satisfies

$$\phi_{u \to v} + \phi_{v \to u} > x_{uv}$$

  There is a unique such u.

- A pleasing symmetry: If u selects v, then v selects u.

This is indeed a consistent matching



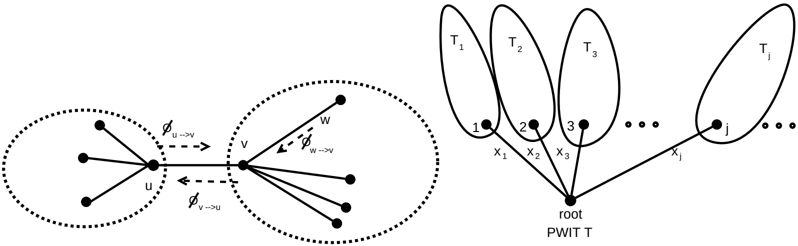
To see one way:

$$\begin{aligned} x_{u,v} - \phi_{u\to v} &= \min\{x_{w,v} - \phi_{w\to v},\ w \sim v\} \\ &< \min\{x_{w,v} - \phi_{w\to v},\ w \sim v, w \neq u\} \\ &= \phi_{v\to u}. \end{aligned}$$

To see the other way, if $z \sim v$ and $z \neq u$, then

$$\begin{aligned} x_{z,v} - \phi_{z\to v} &> \min\{x_{w,v} - \phi_{w\to v},\ w \sim v\} \\ &= \min\{x_{w,v} - \phi_{w\to v},\ w \sim v, w \neq z\} \\ &= \phi_{v\to z}. \end{aligned}$$

Two-crucial properties



- $\phi_{u \to v}$ and $\phi_{v \to u}$ are independent.
- Conditioned on the event that there is an edge of length x at u, say $\{u, v_x\}$, the quantities $\phi_{u \to v_x}$ and $\phi_{v_x \to u}$ are independent with the logistic distribution.

The $\zeta(2)$ result

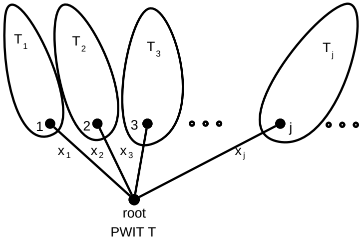
- Consider a matching M on $K_{n,n}$. New interpretation of total cost.

$$\begin{aligned} cost(M) &= \sum_{e \in M} C_e = \frac{1}{n} \sum_{e \in M} \tilde{C}_e \\ &= \frac{1}{2n} \sum_{j=1}^{2n} \tilde{C}_{j,M(j)} = \mathbb{E}[\tilde{C}_{root,M(root)}] \end{aligned}$$

- Next compute this expected cost on the optimal matching on the PWIT tree process.

$$\begin{aligned} \mathbb{E}[X_{root,M^*(root)}] &= \int_0^\infty x \Pr\{\phi_1 + \phi_2 > x\} dx \\ &= \frac{1}{2} \mathbb{E}[(\phi_1 + \phi_2)^2 \mathbf{1}\{\phi_1 + \phi_2 > 0\}] \\ &= \frac{1}{4} \mathbb{E}[(\phi_1 + \phi_2)^2] = \frac{1}{2} \mathbb{E}[\phi_1^2] = \frac{\pi^2}{6} = \zeta(2). \end{aligned}$$

Involution invariance

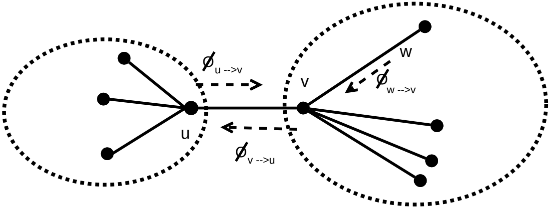


- Any ordinary matching on T won't do.
- Greedy has an expected cost of $1 < \pi^2/6$, but is not allowed.
- We must search among matchings M^* that are limits of M_n^*.
- The statistics must be identical when we move to the neighbour on the best matching, because it is so in the finite graph.
- "Involution invariance".

Thus far ...

- 1/4: BP algorithm, BP for optimisation, positive temperature relaxation, energy cavity equations at positive temperature.
- 2/4: The assignment problem, energy cavity equations, zero-temperature cavity equations, loops but with hope in an ensemble view, locally tree-like limit object, the PWIT.
- 3/4:
  - Symmetries of the PWIT and the recursive distributional equation (RDE).
  - Solution to the RDE, the logistic distribution.
  - The recursive tree process.
  - A good matching on the infinite tree, its consistency, involution invariance.
  - The local view from 'root' and the $\zeta(2)$ calculation.

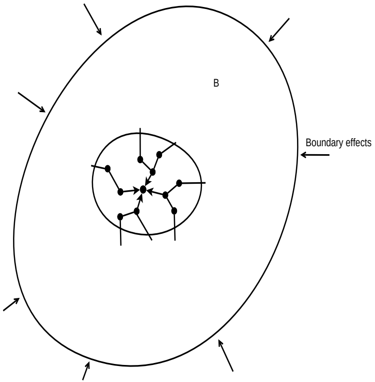
The BP iteration on the tree (and on $K_{n,n}$)



- Belief propagation algorithm.

$$\begin{aligned}
\text{Initialization :} \quad & \phi^{0}_{u\to v} \sim \text{i.i.d. Logistic} \\
\text{Update rule :} \quad & \phi^{(k+1)}_{u\to v} = \min_{w\neq u}\left(X_{v,w} - \phi^{(k)}_{w\to v}\right) \\
\text{Decision rule :} \quad & M^{(k)}(v) = \arg\min\left(X_{v,w} - \phi^{(k)}_{u\to v}\right) \\
& \text{“Matching”}\ \ M^{(k)} = \cup_{v}\{(v, M^{(k)}(v))\}.
\end{aligned}$$

Correlation decay



- The effect of happenings far away should be negligible: need *correlation decay*
- Example: As distance between root i and the boundary $\partial B \to \infty$,

$$\lim_{dist(i,\partial B)\to\infty} \mathbb{E}\left[\max_{x_{\partial B}, x'_{\partial B}} |p(a_{ij}=1|x_{\partial B}) - p(a_{ij}=1|x'_{\partial B})|\right] \to 0$$

Convergence of BP iterates on the PWIT

Theorem

- *On the PWIT, ϕ_{root} is a measurable function of the x's on the tree. (The RDE is endogenous.)*
- *Convergence of the BP iterates on the PWIT:*

$$M_T^k(root) \to M_T^*(root).$$

Proof via a version of "bivariate uniqueness"

- Let X_i be points of a PPP.
- For iid ϕ_i distributed F, let TF be the distribution of $\min_i\{X_i - \phi_i\}$.
- T is a mapping from the space of distributions on $\mathbb{R}$ to itself. The logistic distribution is a fixed point for the T map.
- Similarly $T^{(2)}$ map

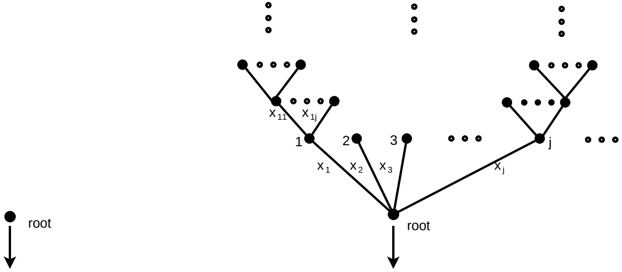
$$F^{(2)} \in \mathcal{P}(\mathbb{R}^2) \mapsto T^{(2)}F^{(2)} = \text{distribution} \begin{pmatrix} \min_i\{X_i - \phi_i^{(1)}\} \\ \min_i\{X_i - \phi_i^{(2)}\} \end{pmatrix},$$

where $(\phi_i^{(1)}, \phi_i^{(2)})_{i\geq 1}$ are iid $F^{(2)}$.

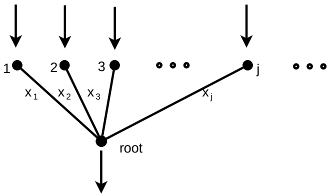
- Bivariate uniqueness if:

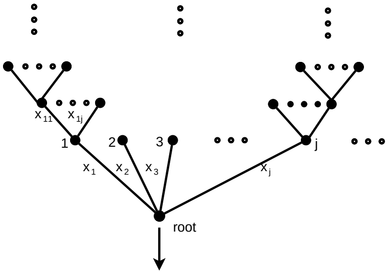
$$\lim_{k\to\infty} (T^{(2)})^k(\mathit{Logistic} \times \mathit{Logistic}) \text{ has } \Pr\{\phi^{(1)} = \phi^{(2)}\} = 1.$$

Bivariate uniqueness: funnelling through

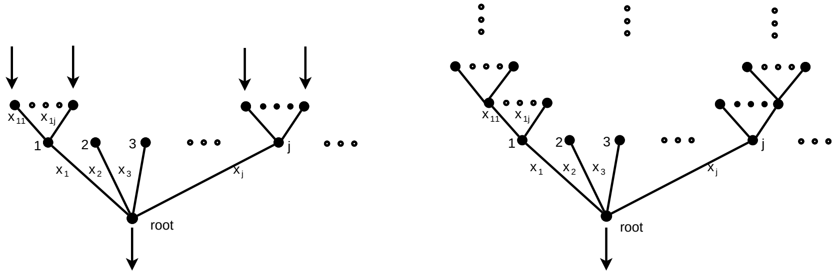


Bivariate uniqueness: funnelling through

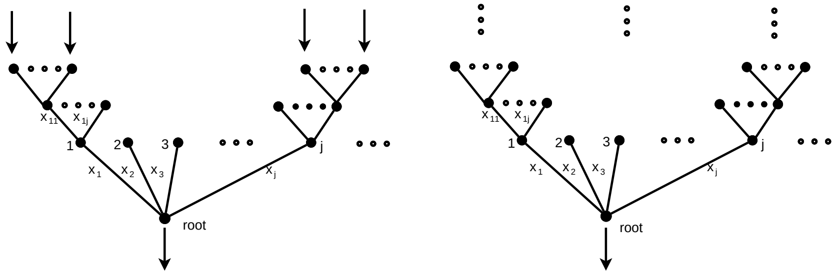




Bivariate uniqueness: funnelling through



Bivariate uniqueness: funnelling through



$$\lim_{k\to\infty} (T^{(2)})^k(\mathit{Logistic}\ \times\ \mathit{Logistic}) \text{ has } \Pr\{\phi^{(1)} = \phi^{(2)}\} = 1.$$

The route to proving correctness

$$
\begin{tikzcd}
M^{(k)}_{K_{n,n}} \arrow[r, dashed, "k \to \infty"] \arrow[d, "n \to \infty"'] & M^{*}_{K_{n,n}} \arrow[d] \\
M^{(k)}_{T} \arrow[r, dashed] & M^{*}_{T}
\end{tikzcd}
$$

Convergence of BP iterates on the PWIT is the bottom convergence.

Local weak limit of graphs with messages

Theorem

1. Convergences of the kth iterate and the optimal matching:

$$(a) \quad \phi^{(k)}_{u\to v}(K_{n,n}) \to \phi^{(k)}_{u\to v}(T) \qquad \textit{as } n\to\infty \textit{ in probability}$$

$$(b) \quad \Pr\left\{(u, M^*_{K_{n,n}}(u)) \neq (u, M^*_T(u))\right\} \to 0 \qquad \textit{as } n\to\infty.$$

2. The approximate matching can be turned into a perfect matching with negligible additional cost.

The route to proving correctness

$$
\begin{tikzcd}
M^{(k)}_{K_{n,n}} \arrow[rr, dashed, "k \to \infty"] \arrow[d, "n \to \infty"'] & & M^{*}_{K_{n,n}} \arrow[d] \\
M^{(k)}_{T} \arrow[rr, dashed] & & M^{*}_{T}
\end{tikzcd}
$$

The downward convergence on the left is of the kth iterate.

The downward convergence on the right is of the optimal matching.

Graphs with marks, and their convergence to respective limit objects.

Approximate to perfect matching

- *It suffices to solve the continuous relaxation of the assignment problem.*
- The adjacency matrix is almost doubly stochastic.
- Use this to compute a partial matching over a $(1 - \varepsilon)$ fraction of nodes.
- Assign unassigned machines to a well-chosen small subset of already assigned jobs, and then move the corresponding machines to handle the unassigned jobs. This can be done at low additional cost.

Matching, Edge cover, TSP, etc.

Let $x_1, x_2, \ldots$ be points of a unit rate Poisson point process.

- Matching: ϕ is a random variable taking values on $\mathbb{R}$ with

$$\phi \stackrel{d}{=} \min_j \left(x_j - \phi_j\right).$$

- Edge cover: ϕ is a random variable taking values on $\mathbb{R}_+$ with:

$$\phi \stackrel{d}{=} \min_j \left(x_j - \phi_j\right)_+.$$

- TSP: ϕ is a random variable taking values on $\mathbb{R}$ with

$$\phi \stackrel{d}{=} \text{second}\min_j \left(x_j - \phi_j\right).$$

- Many-to-one matching, load balancing, etc.

Summary

- BP for optimisation via positive temperature relaxation (graphical model with objective mapping to energy, and an inverse temperature parameter (or two)).
- Cavity equations at positive temperature, and at zero temperature.
- An ensemble perspective and passage to a local weak limit.
- Locally tree-like structure of the limiting object.
- A recursive distributional equation (RDE) and its solution exploiting the symmetries of the limit object.
- Existence of a recursive tree process.
- Endogeny to ensure correlation decay.
- Convergence of BP iterates on the tree. Pull back to $K_{n,n}$.

References

1. Mezard, M. and Montanari, A., 2009. Chapter 14 on "Belief Propagation". In Information, physics, and computation (pp. 291-326). Oxford University Press.

2. Aldous, D. and Steele, J.M., 2004. "The objective method: probabilistic combinatorial optimization and local weak convergence". In Probability on discrete structures (pp. 1-72). Springer, Berlin, Heidelberg.

3. Aldous, D., 1992. Asymptotics in the random assignment problem. Probability Theory and Related Fields, 93(4), pp.507-534.

4. Aldous, D.J., 2001. The $\zeta(2)$ limit in the random assignment problem. Random Structures & Algorithms, 18(4), pp.381-418.

5. Aldous, D.J. and Bandyopadhyay, A., 2005. A survey of max-type recursive distributional equations. The Annals of Applied Probability, 15(2), pp.1047-1110.

6. Salez, J. and Shah, D., 2009. Belief propagation: an asymptotically optimal algorithm for the random assignment problem. Mathematics of Operations Research, 34(2), pp.468-480.

7. Khandwawala, M. and Sundaresan, R., 2014. Belief propagation for optimal edge cover in the random complete graph. The Annals of Applied Probability, 24(6), pp.2414-2454.

8. Khandwawala, M., 2014. Belief propagation for minimum weight many-to-one matchings in the random complete graph. Electronic Journal of Probability, 19.

9. Wästlund, J., 2012. Replica symmetry of the minimum matching. Annals of Mathematics, pp.1061-1091.