Optimality of Belief Propagation for Random Assignment Problem

J. Salez [*] ***and D. Shah*** [†]

Abstract

The assignment problem concerns finding the minimum-cost perfect matching in a complete weighted $n \times n$ bipartite graph. Any algorithm for this classical question clearly requires $\Omega(n^2)$ time, and the best known one (Edmonds and Karp, 1972) finds solution in $O(n^3)$. For decades, it has remained unknown whether optimal computation time is closer to n^3 or n^2. We provide answer to this question for random instance of assignment problem. Specifically, we establish that Belief Propagation finds solution in $O(n^2)$ time when edge-weights are i.i.d. with *light tailed* distribution.

1 Introduction

Given a matrix of n^2 costs $(X_{i,j})_{1\leq i,j\leq n}$, the assignment problem consists of determining a permutation π of $\{1,\dots,n\}$ whose total cost $\sum_{i=1}^{n} X_{i,\pi(i)}$ is minimal. This is equivalent to finding a minimum-weight complete matching in the bipartite graph $G = (V_1 \cup V_2, E)$, with $|V_1| = |V_2| = n$ and $E = \{(i,j) : i \in V_1, j \in V_2\}$, edge $(i,j) \in E$ being assigned weight $X_{i,j}$. Recall that a complete matching on a graph is a subset of pairwise disjoint edges covering all vertices. In what follows, we consider the random model where the $(X_{i,j})$ are i.i.d. with c.d.f. denoted by H (i.e. $H(t) = \mathbb{P}(X_{i,j} \leq t)$). The resulting randomly weighted $n \times n$ bipartite graph will be denoted by $\mathcal{K}_{nn}$ and its optimal matching by $\pi^*_{\mathcal{K}_{nn}}$. We are interested in the convergence speed of the Belief Propagation heuristic for finding $\pi^*_{\mathcal{K}_{nn}}$.

1.1 Related Work

Although it seems cunningly simple, the assignment problem has led to rich development in combinatorial probability and algorithm design since the early 1960s. Partly motivated to obtain insights for better algorithm design, the question of finding asymptotics of the average cost of $\pi^*_{\mathcal{K}_{nn}}$ became of great in-

[*] INRIA and Ecole Normale Supérieure de Paris
[†] EECS, Massachussets Institute of Technology

terest (see [16, 8, 10, 11, 7]). In 1987, through replica method based calculations, Mézard and Parisi [13] conjectured that

$$\mathbb{E}\left[\sum_{i=1}^{n} X_{i,\pi^*_{\mathcal{K}_{nn}}(i)}\right] \xrightarrow[n\to\infty]{} \zeta(2).$$

This was rigorously established by Aldous [2] more than a decade later (2001), leading to the formalism of "the objective method" (see survey by Aldous and Steele [4]). The finite exact version of the above conjecture,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_{i,\pi^*_{\mathcal{K}_{nn}}(i)}\right] = \sum_{i=1}^{n} \frac{1}{i^2},$$

was independently established by Nair, Prabhakar and Sharma [14] and Linusson and Wåstlund [12] in 2003.

On the algorithmic aspect, the assignment problem has been extremely well studied and its consideration laid foundations for the rich theory of network flow algorithms. The best known algorithm is by Edmonds and Karp [9] and takes $O(n^3)$ operations for arbitrary instance. Concurrently, the statistical physics-based approach mentioned above suggested a non-rigorous decentralized iterative strategy which turned out to be an instance of the more general Belief Propagation (BP) heuristic, popular in artificial intelligence (see, book by Pearl [15] and work by Yedidia, Freeman and Weiss [17]). In a recent work, one of the authors of the present paper, Shah along with Bayati and Sharma [6], established correctness of this iterative scheme for any instance of the assignment problem, as long as the optimal solution is unique. More precisely, they showed exact convergence within at most $\lceil \frac{2n \max_{i,j} X_{i,j}}{\varepsilon} \rceil$ iterations, where ε denotes the difference of weights between optimum and second optimum. This bound is always greater than n, and typically scales like $O(n^2)$ as n goes to infinity – at least in the random model mentioned above. Since each iteration needs time $\Theta(n^2)$, the resulting computation cost does not seem competitive.

1.2 Our contribution

Simulation studies show much better performances on average than what is suggested by the worst case upper bound in [6]. Motivated by this, we consider here the question of determining the typical convergence rate of BP when running on large randomly generated cost matrices. Specifically, we establish that the number of iterations required in order to find an almost-optimal assignment remains bounded as $n \to \infty$. Thus, the total computation cost scales as $O(n^2)$ only. This is in sharp contrast to the best known bound of $O(n^3)$. Clearly, no algorithm can perform better than $\Omega(n^2)$. That is, BP is optimal for the random assignment problem.

2 Result and organization

2.1 BP algorithm

As we shall see later, the analysis of BP on $\mathcal{K}_{nn}$ as $n \to \infty$ will lead us to the study of the same dynamics on a limiting infinite tree. Therefore, we define the BP algorithm once and for all for an arbitrary weighted graph $G = (V, E)$. We use notation that the weight of $\{u, v\} \in E$ is $\|u, v\|_G$. By $w \sim v$, we denote that w is a neighbor of v in G. Also note that a complete matching on G can be equivalently seen as an involutive mapping π_G connecting each vertex v to one of its neighbors $\pi_G(v)$. We shall now onwards use this mapping representation rather than the edge set description.

The BP algorithm is distributed and iterative: in each iteration, it involves sending a real-valued message in both directions along each edge of the graph. Specifically, in iteration $k \geq 0$ every vertex $v \in V$ sends a message $\langle v \to w \rangle_G^k$ to each of its neighbor $w \sim v$. Initialization and update are as follows:

$$\langle v \to w \rangle_G^0 = 0\,; \tag{1}$$

$$\langle v \to w \rangle_G^{k+1} = \min_{\substack{u \sim v \\ u \neq w}} \left\{ \left\| u, v \right\|_G - \langle u \to v \rangle_G^k \right\} \tag{2}$$

Using these messages, every vertex $v \in V$ estimates the neighbor $\pi_G^k(v)$ to which it should be connected as follows:

$$\pi_G^k(v) = \arg\min_{w \sim v} \left\{ \left\| v, w \right\|_G - \langle w \to v \rangle_G^k \right\}. \tag{3}$$

When $G = \mathcal{K}_{nn}$, [6] ensures convergence of $\pi_{\mathcal{K}_{nn}}^k$ to the optimum $\pi_{\mathcal{K}_{nn}}^*$ as long as the latter is unique, which holds a.s. provided the cost c.d.f. H is continuous. The present paper asks about the typical speed of such a convergence, and more precisely its behavior as $n \to \infty$.

2.2 Result

We introduce the natural fraction-difference distance between two given assignments π, π' on a graph $G = (V, E)$:

$$d(\pi, \pi') = \frac{1}{|V|} \mathrm{card}\Big\{ v \in V, \pi(v) \neq \pi'(v) \Big\}.$$

Theorem 1. *Assume the cumulative distribution function H satisfies:*

A1. Regularity : H is continuous and $H'(0^+)$ exists and is non-zero;

A2. Light-tail property : as $t \to \infty$, $H(t) = 1 - \left(e^{-\beta t}\right)$ for some $\beta > 0$.

Then, $\lim_{k \to \infty} \lim_{n \to \infty} \mathbb{E}\Big[d\big(\pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*\big) \Big] = 0.$

In other words, given any $\varepsilon > 0$, there exists $k(\varepsilon), n(\varepsilon)$ such that the expected fraction of non-optimal row-to-column assignments after $k(\varepsilon)$ iterations of the BP algorithm on a random $n \times n$ cost array is less than ε, no matter how large $n \geq n(\varepsilon)$ is. Consequently, the probability to get more than any given fraction of errors can be made as small as desired within finite number of iterations, independently of n. Since each iteration requires $O(n^2)$ operations, the overall computation cost of the BP algorithm for the random assignment problem scales as $O(n^2)$ only. This applies for large class of cost distributions, including uniform over $[0, 1]$ or exponential.

2.3 Organization

The remaining of the paper is dedicated to proving Theorem 1. Although it is far from being an implication of the result by Aldous [2], it utilizes the machinery of local weak convergence, and in particular the Poisson Weighted Infinite Tree (PWIT) $\mathcal{T}$ appearing as the limit of the sequence $(\mathcal{K}_{nn})_{n \geq 1}$. These notions are recalled in Section 3. Figure 2.3 illustrates the three steps of our proof.

1. First (Section 4), we prove that BP's behavior on $\mathcal{K}_{nn}$ "converges" as $n \to \infty$ to its behavior on $\mathcal{T}$ – corresponding to the left vertical arrow in the Figure 2.3 and formally stated as Theorem 3.

2. Second (Section 5), we establish convergence of the recursive tree process describing BP's execution on $\mathcal{T}$ – corresponding to the bottom horizontal arrow in Figure 2.3 and summarized as Theorem 5.

We note that Theorem 5 resolves an open problem stated by Aldous and Bandyopadhay [3].

3. Third (Section 6), the connection between the fix-point on $\mathcal{T}$ and the optimal matching on $\mathcal{K}_{nn}$ is provided by the work by Aldous [2] – corresponding to the vertical right arrow and stated as Theorem 6. We use it to conclude our proof.

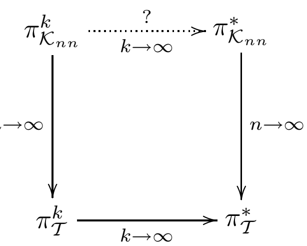

Figure 1. *Theorem 1 corresponds to establishing top-horizontal arrow; which is done by establishing the other three arrows in the above diagram.*

3 Preliminaries

We recall here the necessary framework introduced by Aldous in [2]. Consider a rooted, edge-weighted and connected graph G, with distance between two vertices being defined as the infimum over all paths connecting them of the sum of edge-weights along that path. For any $\varrho > 0$, define the $\varrho-$restriction of G as the sub-graph $\lceil G \rceil_\varrho$ induced by the vertices that are within distance at most ϱ from the root. Call G a *geometric graph* if $\lceil G \rceil_\varrho$ is finite for every $\varrho > 0$.

Definition 1 (local convergence). *Let $G, G_1, G_2, \ldots$ be geometric graphs. We say that $(G_n)_{n \geq 1}$ converges to G if for every $\varrho > 0$ such that no vertex in G is at distance exactly ϱ from the root:*

1. *$\exists n_\varrho \in \mathbb{N}$ s.t. the $\lceil G_n \rceil_\varrho, n \geq n_\varrho$ are all isomorphic[1] to $\lceil G \rceil_\varrho$;*
2. *One can chose the isomorphisms $\gamma^\varrho_n \colon \lceil G \rceil_\varrho \rightleftharpoons \lceil G_n \rceil_\varrho, n \geq n_\varrho$ so that for every edge $\{u, v\}$ in $\lceil G \rceil_\varrho$:*

$$\left\| \gamma^\varrho_n(u), \gamma^\varrho_n(v) \right\|_{G_n} \xrightarrow[n \to \infty]{} \left\| u, v \right\|_G .$$

The intuition behind this definition is that for large n, G_n should look very much like G in any arbitrarily large but fixed neighborhood of the root, in terms of both the structure (item 1) and the edge-weights (item 2). When each oriented edge (u, v) is also assigned a label $\lambda(u, v)$ taking values in some Polish space (Λ, d_Λ), we moreover require the isomorphisms γ^ϱ_n to satisfy, for every oriented edge (u, v) in $\lceil G \rceil_\varrho$,

$$\lambda_{G_n}\left(\gamma^\varrho_n(u), \gamma^\varrho_n(v)\right) \xrightarrow[n \to \infty]{d_\Lambda} \lambda_G\left(u, v\right).$$

With little work, one can define a distance that metrizes this notion of convergence and makes the space of geometric graphs complete and separable. As a consequence, one can import the usual machinery related to weak convergence of probability measures.

Next, we recall result by Aldous that showed $\mathcal{K}_{nn}$ local weak convergence to the so-called Poisson Weighted Infinite Tree. Before we state the result, we will need some notation that will be useful throughout the paper. Let $\mathcal{V}$ denote the set of all finite words over the alphabet $\mathbb{N}^*$, $\varnothing$ the empty word, "$\cdot$" the concatenation operation and for any

[1] An isomorphism from $G = (V, \varnothing, E)$ to $G' = (V', \varnothing', E')$, denoted $\gamma \colon G \rightleftharpoons G'$, is simply a bijection from V to V' preserving the root $\gamma(\varnothing) = \varnothing'$) and the structure ($\forall (x, y) \in V, \{\gamma(x), \gamma(y)\} \in E' \Leftrightarrow \{x, y\} \in E$).

$v \in \mathcal{V}^* := \mathcal{V} \setminus \{\varnothing\}$, $\dot{v}$ the word obtained from v by simply deleting the last letter. Set also $\mathcal{E} := \{\{v, v.i\}, v \in \mathcal{V}, i \geq 1\}$. The graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denotes an infinite tree with $\varnothing$ as root, all words of length 1 as the nodes at depth 1, words of length 2 as the nodes at depth 2, etc.

Theorem 2 (Aldous, [1, 2]). *To each edge $\{v, v.i\}$ of $\mathcal{T}$, assign weight ξ_i^v, where $(\xi^v = \xi_1^v, \xi_2^v \ldots)_{v \in \mathcal{V}}$ is a family of independent, ordered Poisson point processes with intensity 1 on $\mathbb{R}^+$. Then, under assumption A1 on H:*

$$nH'(0^+)\mathcal{K}_{nn} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{T}, \qquad (4)$$

in the sense of local convergence. The random geometric graph $\mathcal{T}$ is called the Poisson Weighted Infinite Tree (PWIT).

To get rid of scaling factors, we will multiply all edge-weights in $\mathcal{K}_{nn}$ by $nH'(0^+)$. Observe that both the optimal matching $\pi^*_{\mathcal{K}_{nn}}$ and the BP decisions $\pi^k_{\mathcal{K}_{nn}}, k \geq 0$ remain unaffected.

4 First step: convergence of dynamics of BP

In this section we deduce from Theorem 2 that the behavior of BP when running on $\mathcal{K}_{nn}$ *converges* as $n \to \infty$ to its behavior when running on $\mathcal{T}$.

Theorem 3 (Continuity of BP). *Consider BP's execution on $\mathcal{K}_{nn}$, with $\mathcal{K}_{nn}$ converging to $\mathcal{T}$ a.s. Then,*

$$\forall k \geq 0, \left(\mathcal{K}_{nn}, \langle\cdot\to\cdot\rangle^k_{\mathcal{K}_{nn}}, \pi^k_{\mathcal{K}_{nn}}\right) \xrightarrow[n\to\infty]{proba} \left(\mathcal{T}, \langle\cdot\to\cdot\rangle^k_{\mathcal{T}}, \pi^k_{\mathcal{T}}\right), \qquad (5)$$

in the sense of local convergence, $\pi^k_{\mathcal{K}_{nn}}$ being here viewed as the labeling function $(v, w) \mapsto \mathbf{1}_{\{w=\pi^k_{\mathcal{K}_{nn}}(v)\}}$.

Proof. In view of the recursive nature of the messages, a natural idea for proving (5) is to proceed by induction over k. The base case of $k = 0$ is trivial. However, when trying to go from step k to step $k+1$ one soon gets confronted to a major hinder. Indeed, one can not simply invoke convergence of each k-step incoming messages at a given vertex to deduce convergence of the resulting $k+1$-step out-coming message, because the update rule (1) is not continuous with respect to local convergence : it involves messages from unboundedly many neighbors as $n \to \infty$. Remarkably enough, it turns out that under assumption A2, only a uniformly bounded number of those neighbors do in fact really matter, as stated in the following technical lemma whose proof is omitted. □

Lemma 1. *Call $\varnothing$ the (uniformly chosen) root of $\mathcal{K}_{nn}$. Then for any $k \geq 0$:*

$$\sup_{n\geq 1} \mathbb{P}\left(\widehat{\pi}^k_{\mathcal{K}_{nn}}(\varnothing) \geq i\right) \xrightarrow[i\to\infty]{} 0, \qquad (6)$$

where $\widehat{\pi}^k_{\mathcal{K}_{nn}}(\varnothing)$ denotes the rank of vertex $\pi^k_{\mathcal{K}_{nn}}(\varnothing)$ in the set of neighbors $v \sim \varnothing$ ordered by increasing length of $\{\varnothing, v\}$

5 Second step: analysis of BP on PWIT

In light of Theorem 3, the analysis of BP on $\mathcal{K}_{nn}$ for large n naturally brings us to the study of BP's dynamics on the limiting PWIT. Formally, we are interested in asymptotics of the random process on $\mathcal{T}$ defined for all $v \in \mathcal{V}^*$ by the recursion:

$$\langle v\to\dot{v}\rangle^{k+1}_{\mathcal{T}} = \min_{i\geq 1}\left\{\left\|v, v.i\right\|_{\mathcal{T}} - \langle v.i\to v\rangle^k_{\mathcal{T}}\right\}, \qquad (7)$$

where the initial values $\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)_{v\in\mathcal{V}^*}$ are i.i.d. random variables independent of $\mathcal{T}$ (0 in the case of our algorithm).

First observe that at any given time k all $\langle v \to \dot{v} \rangle_T^k, v \in \mathcal{V}^*$ share the same distribution, owing to the natural spatial invariance of the PWIT. Moreover, if F denotes the corresponding common anti-c.d.f.[2] at a given time, a straightforward computation (see for instance [2]) shows that the new anti-c.d.f. obtained after a single application of update rule (7) is :

$$TF : x \mapsto \exp\left(-\int_{-x}^{+\infty} F(t)\,dt\right).$$

This defines an operator T on the space $\mathcal{D}$ of anti-c.d.f.'s, (i.e. left-continuous non-increasing functions $F\colon \mathbb{R} \to [0,1]$). T is known to have a unique fix-point (see [2]), the so-called *logistic* distribution:

$$F^* : x \mapsto \frac{1}{1+e^x}.$$

Our first step will naturally consist in studying the dynamics of T on $\mathcal{D}$.

5.1 Weak attractiveness.

Finding the domain of attraction of F^* under operator T is not known and has been listed as open problem by Aldous and Bandyopadhyay ([3, Open Problem # 62]). In what follows, we answer this question and more. We fully characterize the asymptotical behavior of the successive iterates $(T^kF)_{k\geq 0}$ for any initial distribution $F \in \mathcal{D}$. First observe that T is non-increasing in the following sense:

$$F \leq F' \text{ on } \mathbb{R} \Rightarrow TF' \geq TF \text{ on } \mathbb{R}.$$

This suggests considering the non-decreasing second iterate T^2. However, unlike T the second iterate T^2 admits an infinite number of fix-points. To see this, let θ_t ($t \in \mathbb{R}$) be the shift operator defined on $\mathcal{D}$ by $\theta_t F\colon x \mapsto F(x-t)$. Then,

$$T \circ \theta_t = \theta_{-t} \circ T, \tag{8}$$

Therefore, it follows that $T^2(\theta_t F^*) = \theta_t(T^2F^*) = \theta_t F^*$ for all $t \in \mathbb{R}$. Thus, $\theta_t F^*$ is fixed point of T^2 for all $t \in \mathbb{R}$. This observation leads us to introduce the key tool of our analysis:

Definition 2. *For $F \in \mathcal{D}$, define the transform $\widehat{F}$ as follows :*

$$\widehat{F}(x) = x + \ln\left(\frac{F(x)}{1-F(x)}\right).$$

The reason behind defining this transform is the following straightforward fact.

Lemma 2. *For any given $F \in \mathcal{D}$ and $x \in \mathbb{R}$, $F(x) = \theta_{\widehat{F}(x)} F^*(x)$. Further, $F \equiv \theta_x F^*$ if and only if $\widehat{F}$ is constant on $\mathbb{R}$ with value x.*

The above Lemma suggests that the maximal amplitude of the variations of $\widehat{F}$ on $\mathbb{R}$ tells something about the distance between F and the family of fix-points $\{\theta_t F^*, t \in \mathbb{R}\}$. Therefore, we will now focus on the variations of $\widehat{F}$, and especially the behavior of those variations under the action of T. We state three technical lemmas whose proofs can be found in the full version of this paper.

Lemma 3. *Let $F \in \mathcal{D}\setminus\{0\}$ be integrable at $+\infty$. Then, $\widehat{T^4F}$ is bounded on $\mathbb{R}$.*

Lemma 4. *If $F \in \mathcal{D}$ is such that $\widehat{F}$ is bounded, then $\widehat{T^2F}$ is bounded too and moreover :*

$$\sup_{\mathbb{R}} \widehat{T^2F} \leq \sup_{\mathbb{R}} \widehat{F}; \qquad \inf_{\mathbb{R}} \widehat{T^2F} \geq \inf_{\mathbb{R}} \widehat{F}.$$

Further, the above inequalities are strict if and only if $\widehat{F}$ is not constant on $\mathbb{R}$.

[2]The anti-c.d.f. of a real r.v. X is the function $F\colon x \to \mathbb{P}(X > x)$.

Lemma 5. *Let $F \in \mathcal{D}$ be such that $\widehat{F}$ is bounded. Then, $\widehat{T^k F}$ is continuously differentiable for $k \geq 2$, and the family $(\widehat{T^k F})', k \geq 3$ is uniformly integrable.*

We are now ready to state the main result of this section, which fully characterizes the asymptotics of $(T^k)_{k \geq 0}$.

Theorem 4 (Dynamics of T on $\mathcal{D}$). *Consider any $F \in \mathcal{D} \setminus \{0\}$ that is integrable at $+\infty$. Then*

$$\sup_{\mathbb{R}} \left| \widehat{T^k F} - (-1)^k \gamma \right| \underset{k \to \infty}{\searrow} 0,$$

for some constant $\gamma \in \mathbb{R}$ (dependent on F). In particular, the following convergence occurs uniformly on $\mathbb{R}$:

$$\begin{cases} T^{2k} F \xrightarrow[k \to \infty]{} \theta_\gamma F^*; \\ T^{2k+1} F \xrightarrow[k \to \infty]{} \theta_{-\gamma} F^*. \end{cases}$$

Remark 1. *Our assumption on F is minimal: if $F \equiv 0$ or $\int_0^\infty F = +\infty$, then the sequence $(T^k F)_{k \geq 1}$ trivially alternates between the 0 and 1 functions.*

Proof. Lemma 3 ensures existence of $M \geq 0$ such that for all $k \geq 4$,

$$\theta_{-M} F^* \leq T^k F \leq \theta_M F^*. \qquad (9)$$

By Lemma 4, the bounded real sequences $(\inf_{\mathbb{R}} \widehat{T^{2k} F})_{k \geq 2}$ and $(\sup_{\mathbb{R}} \widehat{T^{2k} F})_{k \geq 2}$ are monotone, hence converging, say to γ^- and γ^+ respectively. All we have to show is that $\gamma^- = \gamma^+$; convergence of $(\widehat{T^{2k+1} F})_{k \geq 2}$ to the opposite constant will then follow from Property (8) .

By Arzela-Ascoli theorem, the family of (clearly bounded and 1-lipschitzian) functions $(T^{2k} F)_{k \geq 4}$ is relatively compact with respect to uniform convergence on compact subsets. Thus, there exists a convergent sub-sequence:

$$T^{2\varphi(k)} F \xrightarrow[k \to \infty]{} F_\infty.$$

This implies convergence of $\widehat{T^{2\varphi(k)} F}$ to $\widehat{F_\infty}$ since on every fixed compact set of $\mathbb{R}$ the uniform bound (9) keeps all the values of the $T^{2\varphi(k)} F, k \geq 0$ within a compact subset of $]0, 1[$ over which the mapping $y \mapsto \ln \frac{y}{1-y}$ is uniformly continuous. Even better, the uniform integrability of variations stated in Lemma 5 allows us to turn this uniform convergence on compact subsets into a uniform convergence on all $\mathbb{R}$. In particular,

$$\inf_{\mathbb{R}} \widehat{F_\infty} = \gamma^- \text{ and } \sup_{\mathbb{R}} \widehat{F_\infty} = \gamma^+.$$

Now, the restriction of T to the subset $\{F \in \mathcal{D}, -M \leq \widehat{F} \leq M\}$ is clearly continuous with respect to uniform convergence on compact subsets. Therefore,

$$T^{2(\varphi(k)+1)} F \xrightarrow[k \to \infty]{} T^2 F_\infty.$$

But using exactly the same arguments as above, we obtain a similar conclusion:

$$\inf_{\mathbb{R}} \widehat{T^2 F_\infty} = \gamma^- \text{ and } \sup_{\mathbb{R}} \widehat{T^2 F_\infty} = \gamma^+.$$

By the second part of Lemma 4, it must be that $\gamma^- = \gamma^+$. □

5.2 Strong attractiveness.

So far, we have established the distributional convergence of the messages process. To complete the algorithm analysis, we now need to prove sample-path wise convergence. We note that Aldous and Bandyhopadhyay [3, 5] have studied the special case where the initial messages $(\langle v \to \dot{v} \rangle_T^0)_{v \in \mathcal{V}^*}$ are i.i.d. with distribution being the fix-point F^*. They established L^2-convergence of the message process to some unique stationary configuration, independent of the

F^*-distributed initial messages $(\langle v \to \dot{v}\rangle^0_{\mathcal{T}})_{v\in\mathcal{V}^*}$. They call this property *endogenity.* The main result of the present sub-section consists in extending such an *endogenity* to the case of F-distributed initial messages, where F is an (almost) arbitrary distribution. To this end, we construct an appropriate coupling on $\mathcal{T}$.

Theorem 5 (Convergence of BP on $\mathcal{T}$). *Assume that the i.i.d. initial messages $\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)_{v\in\mathcal{V}^*}$ satisfy*

$$\mathbb{E}\left[\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)^+\right] < \infty. \qquad (10)$$

Then, up to some additive constant $\gamma \in \mathbb{R}$, the recursive tree process defined by (7) converges to the unique stationary configuration $\langle \cdot \to \cdot\rangle^*_{\mathcal{T}}$ in the following sense: for every $v \in \mathcal{V}^*$,

$$\langle v \to \dot{v}\rangle^k_{\mathcal{T}} - (-1)^k\gamma \xrightarrow[k\to\infty]{L^2} \langle v \to \dot{v}\rangle^*_{\mathcal{T}}.$$

Further, defining $\pi^*_{\mathcal{T}}$ as the assignment induced by $\langle\cdot\to\cdot\rangle^*_{\mathcal{T}}$ according to rule (3), we have convergence of decisions at the root:

$$\pi^k_{\mathcal{T}}(\varnothing) \xrightarrow[k\to\infty]{proba} \pi^*_{\mathcal{T}}(\varnothing).$$

Remark 2. *Assumption (10) is minimal as otherwise the process becomes a.s. infinite after the very first iteration.*

Proof. Denote by F the anti-cdf of the initial messages and by γ the constant appearing in Theorem 4. First, observe that if we add a constant to all the initial messages then under the dynamics (7), the same constant is added to every even message $\langle v \to \dot{v}\rangle^{2k}_{\mathcal{T}}$ and subtracted from every odd message $\langle v \to \dot{v}\rangle^{2k+1}_{\mathcal{T}}$. Therefore, without loss of generality we may assume $\gamma = 0$. That is, for any $\varepsilon > 0$ there exists $k_\varepsilon \in \mathbb{N}$ so that

$$\theta_{-\varepsilon}F^* \le T^{k_\varepsilon}F \le \theta_\varepsilon F^*.$$

By the Skorohod's representation theorem, there exists a probability space $E' = (\Omega', \mathcal{F}', P')$, possibly differing from the original space $E = (\Omega, \mathcal{F}, P)$, on which is defined a random variable X^ε with distribution $T^{k_\varepsilon}F$ along with two other random variables X^- and X^+ with distribution F^*, in such a way that:

$$\text{a.s., } X^- - \varepsilon \le X^\varepsilon \le X^+ + \varepsilon. \qquad (11)$$

Now consider the product space $(\bigotimes_{v\in\mathcal{V}} E') \otimes E$ over which are jointly defined the PWIT $\mathcal{T}$ and independent copies $(X^-_v, X^\varepsilon_v, X^+_v)_{v\in\mathcal{V}}$ of the triple (X^-, X, X^+) for each vertex $v \in \mathcal{V}$. On $\mathcal{T}$, let us compare the configurations $\left(\langle\cdot \to \cdot\rangle^{k,-}_{\mathcal{T}}\right)_{k\ge 0}$, $\left(\langle\cdot \to \cdot\rangle^{k,\varepsilon}_{\mathcal{T}}\right)_{k\ge 0}$ and $\left(\langle\cdot \to \cdot\rangle^{k,+}_{\mathcal{T}}\right)_{k\ge 0}$ resulting from three different initial conditions, namely:

$$\forall v \in \mathcal{V}^*, \begin{cases} \langle v\to\dot{v}\rangle^-_{\mathcal{T}} & := & X^-_v; \\ \langle v\to\dot{v}\rangle^\varepsilon_{\mathcal{T}} & := & X^\varepsilon_v; \\ \langle v\to\dot{v}\rangle^+_{\mathcal{T}} & := & X^+_v. \end{cases}$$

Due to anti-monotony and homogeneity of the update rule (7), inequality (11) 'propagates' in the sense that for any $k \ge 0$ and $v \in \mathcal{V}^*$, when k is even,

$$\langle v\to\dot{v}\rangle^{k,-}_{\mathcal{T}} - \varepsilon \le \langle v\to\dot{v}\rangle^{k,\varepsilon}_{\mathcal{T}} \le \langle v\to\dot{v}\rangle^{k,+}_{\mathcal{T}} + \varepsilon\ ;$$

and when k is odd,

$$\langle v\to\dot{v}\rangle^{k,+}_{\mathcal{T}} - \varepsilon \le \langle v\to\dot{v}\rangle^{k,\varepsilon}_{\mathcal{T}} \le \langle v\to\dot{v}\rangle^{k,-}_{\mathcal{T}} + \varepsilon.$$

Now fix $v \in \mathcal{V}^*$. By construction :

$$\left(\langle v\to\dot{v}\rangle^{k+k_\varepsilon}_{\mathcal{T}}\right)_{k\ge 0} \stackrel{\mathcal{D}}{=} \left(\langle v\to\dot{v}\rangle^{k,\varepsilon}_{\mathcal{T}}\right)_{k\ge 0}.$$

In particular, for every $k \ge k_\varepsilon$ we have

$$\begin{aligned}&\sup_{s,t\ge k} \left\|\langle v\to\dot{v}\rangle^s_{\mathcal{T}} - \langle v\to\dot{v}\rangle^t_{\mathcal{T}}\right\|_{L^2} \\ &= \sup_{s,t\ge k-k_\varepsilon} \left\|\langle v\to\dot{v}\rangle^{s,\varepsilon}_{\mathcal{T}} - \langle v\to\dot{v}\rangle^{t,\varepsilon}_{\mathcal{T}}\right\|_{L^2} \\ &\le 2\sup_{t\ge k-k_\varepsilon} \left\|\langle v\to\dot{v}\rangle^{t,\pm}_{\mathcal{T}} - \langle v\to\dot{v}\rangle^*_{\mathcal{T}}\right\|_{L^2} + 2\varepsilon.\end{aligned}$$

Here the endogeneity property established by Aldous and Bandyhopadhyay [3, 5] for logistic distributions implies:

$$\sup_{t\geq k-k_\varepsilon} \left\| \langle v \to \dot{v} \rangle_{\mathcal{T}}^{t,\pm} - \langle v \to \dot{v} \rangle_{\mathcal{T}}^{*} \right\|_{L^2} \xrightarrow[k\to\infty]{} 0.$$

Thus, the sequence $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k}\right)_{k\geq 0}$ is Cauchy in L^2, hence convergent. It is not hard to check that the limiting configuration has to be stationary, i.e. is a fixed point for the recursion (7), and that the estimates $\pi_{\mathcal{T}}^k, k \geq 0$ do in turn converge to the estimate $\pi_{\mathcal{T}}^*$ associated with the limiting configuration. Note that endogeneity implies uniqueness of the stationary configuration, and therefore $\pi_{\mathcal{T}}^*$ is the infinite assignment studied in [2]. ▯

6 Third step: putting things together

Our last step utilizes the following remarkable result of Aldous.

Theorem 6 (Aldous, [2]). *Let $\pi_{\mathcal{T}}^*$ be the assignment associated with the unique stationary configuration $\langle \cdot \to \cdot \rangle_{\mathcal{T}}^*$. Then $\pi_{\mathcal{T}}^*$ is a perfect matching on $\mathcal{T}$, and*

$$\left(\mathcal{K}_{nn}, \pi_{\mathcal{K}_{nn}}^*\right) \xrightarrow{\mathcal{D}} \left(\mathcal{T}, \pi_{\mathcal{T}}^*\right), \qquad (12)$$

in the sense of local weak convergence.

Proof. **(of Theorem 1)** By Theorem 3 and Skorokhod's embedding Theorem, convergence (12) can be extended to include BP's answer at any fixed step k:

$$\left(\mathcal{K}_{nn}, \pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*\right) \xrightarrow[n\to\infty]{\mathcal{D}} \left(\mathcal{T}, \pi_{\mathcal{T}}^k, \pi_{\mathcal{T}}^*\right).$$

Therefore, the probability of an error at the root of $\mathcal{K}_{nn}$ (which by symmetry of $\mathcal{K}_{nn}$ is nothing but the expected fraction of errors), converges to the probability of an error at the root of $\mathcal{T}$: for all $k \geq 0$,

$$\mathbb{E}\left[d(\pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*)\right] = \mathbb{P}\left(\pi_{\mathcal{K}_{nn}}^k(\varnothing) \neq \pi_{\mathcal{K}_{nn}}^*(\varnothing)\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\pi_{\mathcal{T}}^k(\varnothing) \neq \pi_{\mathcal{T}}^*(\varnothing)\right).$$

Finally, Theorem 5 guarantees that the right-hand side vanishes as $k \to \infty$. ▯

7 Conclusion

In this paper we established[3] that the BP algorithm finds almost optimal solution to a random assignment problem in $O(n^2)$ time for a problem of size n with high probability. The natural lower bound of $\Omega(n^2)$ makes BP an (order) optimal algorithm for finding minimum cost matching in a bipartite graph. This result significantly improves over the bound proved by Bayati, Shah and Sharma [6] for the BP algorithm ; or for that matter the best known worst case bound on performance of algorithm by Edmonds and Karp [9].

Beyond the obvious practical interest of such an extremely efficient distributed algorithm for locally solving huge instances of the optimal assignment problem, we hope that the method used here – essentially replacing the asymptotical analysis of the iteration as the size of the underlying graph tends to infinity by its exact study on the infinite limiting structure revealed via local weak convergence – will become a powerful tool in the fascinating quest for a general mathematical understanding of loopy belief propagation. To the best of our knowledge, this is the first non-trivial use of local weak convergence framework for analyzing performance of algorithms.

[3]Due to space constraints, we have omitted certain technical proofs throughout the paper. The reader will find them in the full version of the present paper.

Bibliography

[1] D. Aldous. Asymptotics in the random assignment problem. *Prob. Th. Rel. Fields*, 93:507–534, 1992.

[2] D. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Struct. Algo.*, 18:381–418, 2001.

[3] D. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Ann. of App. Prob.*, 15(2):1047–1110, 2005.

[4] D. Aldous and J. Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. *Discr. Comb. Prob.*, 2003.

[5] A. Bandyopadhyay. Bivariate uniqueness in the logistic recursive distributional equation. Technical Report 629, Dept. of Stat., University of Cal., Berkeley, CA, 2002.

[6] M. Bayati, D. Shah, and M. Sharma. Max-product for maximum weight matching: convergence, correctness and lp duality. *IEEE Trans. Info. Th.*, March 2008.

[7] D. Coppersmith and G. Sorkin. Constructive bounds and exact expectations for the random assignment problem. *Random Struct. Algo.*, 15:113–144, 1999.

[8] M. Dyer, A. Frieze, and C.J.H. McDiarmid. On linear programs with random costs. *Math. Prog.*, 35:3–16, 1986.

[9] J. Edmonds and R. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *J. of the ACM*, 19:248–264, 1972.

[10] R. Karp. An upper bound on the expected cost of an optimal assignment. *Discr. Algo. Complexity: Proc. Japan- U.S. Joint Seminar, Academic Press*, pages 1–4, 1987.

[11] A. Lazarus. Certain expected values in the random assignment problem. *Oper. Res. Lett.*, 14:207–214, 1993.

[12] S. Linusson and J. Wåstlund. A proof of parisiŠs conjecture on the random assignment problem. *Prob. Th. Rel. Fields*, 128:419–440, 2004.

[13] M. Mézard and G. Parisi. On the solution of the random link matching problem. *J. of Physics*, 48:1451–1459, 1987.

[14] C. Nair, B. Prabhakar, and M. Sharma. Proofs of the parisi and coppersmith-sorkin random assignment conjectures. *Random Struct. Algo.*, 27(4):413–444, 2005.

[15] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Francisco, CA, 1988.

[16] D. Walkup. On the expected value of a random assignment problem. *SIAM J. Comp.*, 8:440–442, 1979.

[17] J. Yedidia, W. Freeman, and Y. Weiss. Generalized belief propagation. Technical Report TR-2000-26, Mitsubishi Elect. Res. Lab., Cambridge, MA, 2000.