Technical University of Crete

Functional Analysis Final Report

Panos Alevizos
Instructor: Alex Potamianos

July 29, 2013

1 Introduction

This work studies the sufficient conditions for the convergence of the *Sum-Product algorithm* (SPA) when the factors of the *factor graph* (FG) have pairwise interactions and variables have binary domain [1]. In a nutshell, FGs represent graphically the factorization of a global function into a product of local sub-functions. The global function is usually a multi-variable probability density/mass function (pdf/pmf), where the calculation of a marginal pdf/pmf is usually intractable. The SPA is applied on the FG through message-passing, i.e. exchange of functions, between the FG nodes in a distributed way; the output is a marginal pdf/pmf with respect to a variable of interest. Factor graph theory has several applications in many interdisciplinary fields, such as error correction coding theory, detection and estimation, wireless networking, artificial intelligence and many others. Further details regarding FGs and SPA can be found in [2–6].

The notation is the following. The set $\mathbb{B}$ denotes the set of binary numbers, i.e. $\mathbb{B} = \{x : x \in \{-1, 1\}\}$. The operator $|\cdot|$ stands for the cardinality of a set, i.e. $|\mathbb{B}| = 2$.

The rest of this work is organized as follows. In Section 2 are provided some preliminaries about FGs and SPA, Section 3 introduces the notation of [1]. Section 4 derives the sufficient conditions for the convergence of SPA for binary variables and pairwise interactions (factor nodes with at most 2 arguments). Finally, Section 5 concludes this work.

2 Background

2.1 Factor Graphs and the Sum-Product Algorithm

Let $X_1, X_2, ..., X_n$ be a set of variables, in which for each i, X_i takes values in some finite *domain* $\mathcal{X}_i$. Let $f(X_1, X_2, ..., X_n)$ be a real valued function of these variables, i.e. a function with domain

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n, \tag{1}$$

and range the set of real numbers $\mathbb{R}$. The domain $\mathcal{X}$ of f is called *configuration space* for the given set of variables $\{X_1, ..., X_n\}$, and each element of $\mathcal{X}$ is a particular configuration of the variables, i.e. an assignment of a value for each input of f. Knowing that the set of real numbers is closed over summation, we will associate n *marginal functions*[1] associated with function $f(X_1, ..., X_n)$, denoted as $g_{X_i}(x_i)$ for every i. For each $x_i \in \mathcal{X}_i$, the value $g_{X_i}(x_i)$ is obtained by summing the value of $f(X_1, ..., X_n)$ over all configurations of the input variables that have $X_i = x_i$.

The marginal of $f(X_1, ..., X_n)$ with respect to variable X_i is a function from $\mathcal{X}_i$ to $\mathbb{R}$ which is denoted $g_{X_i}(x_i)$, and it is obtained by summing over all other variables. More specifically, the marginal with respect to variable $x_i \in \mathcal{X}_i$ is given by

$$g_{X_i}(x_i) = \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_{i-1} \in \mathcal{X}_{i-1}} \sum_{x_{i+1} \in \mathcal{X}_{i+1}} \cdots \sum_{x_n \in \mathcal{X}_n} f(X_1 = x_1, ..., X_i = x_i, ..., X_n = x_n). \tag{2}$$

[1]Hereafter, it will be considered that the terms marginal and marginal function are equivalent terms

For notational convenience, instead of indicating the variables being summed over, we indicate those variables not being summed over and we will use the following shorthand

$$g_{X_i}(x_i) = \sum_{\sim\{x_i\}} f(x_1, ..., x_i, ..., x_n) \tag{3}$$

$$= \sum_{x_1,...,x_{i-1},x_{i+1},...,x_n} f(x_1, ..., x_i, ..., x_n). \tag{4}$$

Let $f(X_1, ..., X_n)$ factors into a product of several *local* functions, each having some subset of $\{X_1, ..., X_n\}$ as arguments, specifically, it is assumed that $(X_1, ..., X_n)$ can be factorized into K factors, namely,

$$f(X_1, ..., X_n) = \prod_{k=1}^{K} f_k(\mathbf{S}_k), \tag{5}$$

where $\mathbf{S}_k \subseteq \{X_1, ..., X_n\}$ is the subset of variables associated with the real-valued local factor f_k, i.e. its configuration space. Such factorization is not unique. Function $f(\cdot)$ itself is a trivial factorization, since it consist of 1 factor.

Factor graphs are bipartite graphs that represent the factorization of a global function to smaller local functions, e.g. as in expression 5. More formally, we provide the definition below:

Definition 1 (**Factor graph**). Let $f(X_1, ..., X_n)$ be a decomposable function with K factors, namely

$$f(X_1, ..., X_n) = \prod_{k=1}^{K} f_k(\mathbf{S}_k).$$

The factor graph $G(V, E)$ corresponding to global function f is a bipartite graph, where for every variable X_i, there is a variable node denoted with a solid circle, and for every factor f_j, there is a factor node denoted with a non-solid square. Furthermore, if variable X_i is in the domain of factor f_j an edge is created among them, namely $e_{ij} = (X_i, f_j)$. It is customary to write $X_i \in \mathcal{N}(f_j)$ or equivalently $f_j \in \mathcal{N}(X_i)$ to denote that variable X_i is argument of factor f_j or in "graph" words, variable node X_i is adjacent with factor node f_j. $\mathbf{S}_k$ stands for the subset of the variables of global function f associated with local function f_k. ▼

$\mathcal{N}(v)$ stands for the set of variable (factor) nodes that are adjacent with the factor (variable) node v.

Example 1. Consider a function f of 6 variables with identical (finite) domain, i.e. $\mathcal{X}_i = \mathcal{X},\ i = 1, ..., 6$:

$$f(X_1, X_2, X_3, X_4, X_5, X_6) = f_1(X_1, X_4) f_2(X_1, X_3, X_6) f_3(X_2, X_4, X_5) f_4(X_1).$$

The corresponding factor graph is given in Fig. 1. If we want to compute the marginal with respect to variable X_3, $g_{X_3}(x_3)$, we do the following:

$$g_{X_3}(x_3) = \sum_{\sim\{x_3\}} f_1(x_1, x_4) f_2(x_1, x_3, x_6) f_3(x_2, x_4, x_5) f_4(x_1).$$

■

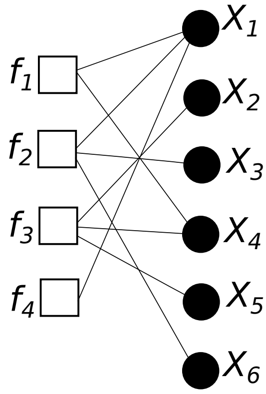


Figure 1: A factor graph that corresponds to the function of the example 1.

SPA is a powerful algorithm for the efficient computation of the marginals of a global factorisable function with finite domain.

The simultaneous computation of all marginal functions could be completed if we had computed all the messages passing across all the edges of FG. An outgoing message along an edge always depends on the incoming messages. SPA employs a synchronized factor/-variable node message passing scheduling, in order to calculate all the messages across the edges of the FG. SPA terminates when all messages have been passed along the edges of the FG.

Suppose we have a variable X_i with finite domain and a local function f_j.[2] Let $\mu_{X_i \longrightarrow f_j}(x_i)$ be denoting the message from a variable node X_i to a neighboring factor node f_j, and $\mu_{f_j \longrightarrow X_i}(x_i)$ be denoting the message from factor node f_j to a variable node X_i. According to the SPA [3], the message-passing update rules have the following form:

- **variable node to local function node update rule:**

$$\mu_{X_i \longrightarrow f_j}(x_i) = \prod_{f_k \in \mathcal{N}(X_i) \setminus \{f_j\}} \mu_{f_k \longrightarrow X_i}(x_i), \tag{6}$$

[2] The term local function is equivalent with the term factor and the term local factor.

- **local function node to variable node update rule:**

$$\mu_{f_j \longrightarrow X_i}(x_i) = \sum_{\sim\{x_i\}} \left(f_j(\mathbf{S}_j = \mathbf{s}_j) \prod_{X_l \in \mathcal{N}(f_j) \setminus \{X_i\}} \mu_{X_l \longrightarrow f_j}(x_l) \right), \tag{7}$$

where $\mathbf{S}_j = \{X_l : X_l \in \mathcal{N}(f_j)\}$. The backslash operator denotes the expression "except". Namely, the expression $X_l \in \mathcal{N}(f_j)\setminus\{X_i\}$ stands for all variable nodes X_l which are adjacent with factor node f_j except variable node X_i. Similarly, $f_k \in \mathcal{N}(X_i)\setminus\{f_j\}$ are all neighboring factors nodes to X_i , except $\{f_j\}$. During initialization phase every leaf factor node f_j (i.e. all factor nodes which are constituted from a single variable) conveys the message $\mu_{f_j \longrightarrow X_m} = f_j(x_m)$. In a similar manner, the messages from every leaf variable node X_i to the neighboring factor node ($f_l \in \mathcal{N}(X_i)$) is $\mu_{X_i \longrightarrow f_l}(x_i) = 1$ (by definition).

Every marginal $g_{X_i}(x_i)$, of a variable X_i results from the product of all incoming messages incident to variable node X_i, i.e.

$$g_{X_i}(x_i) = \prod_{f_k \in \mathcal{N}(X_i)} \mu_{f_k \longrightarrow X_i}(x_i). \tag{8}$$

If we want to find the marginal with respect to a cluster of variables $\mathbf{S}_j$ (that corresponds to a factor f_j, i.e. $\mathbf{S}_j = \{X_l : X_l \in \mathcal{N}(f_j)\}$), then

$$g_{S_j}(\mathbf{s}_j) = f_j(\mathbf{s}_j) \prod_{X_l \in \mathcal{N}(f_j)} \mu_{X_l \longrightarrow f_j}(x_l). \tag{9}$$

Fig. 3 depicts the sum-product algorithm update rule for factor nodes, while Fig. 2 illustrates the sum-product algorithm update rule for variable nodes. Finally, Fig. 4 shows the calculation of a marginal graphically. The calculations of each marginal is performed at every variable node at the last step of SPA (after the calculation of all messages across the edges of FG). The concurrent operation of SPA will be shown with an example, subsequently. One important observation regarding SPA is that every variable node of degree D must perform $D-1$ multiplications.

If we want to formulate everything in terms of messages $\mu_{f_k \longrightarrow X_i}(x_i)$, then from Eqs. 6, 7, we take

$$\mu_{f_j \longrightarrow X_i}(x_i) = \sum_{\sim\{x_i\}} \left(f_j(\mathbf{S}_j = \mathbf{s}_j) \prod_{X_l \in \mathcal{N}(f_j) \setminus \{X_i\}} \prod_{f_k \in \mathcal{N}(X_i) \setminus \{f_j\}} \mu_{f_k \longrightarrow X_l}(x_l) \right), \tag{10}$$

Finally it is presented an illustrative example that clarifies the message-passing operations of SPA and the respective time-scheduling. At each step the corresponding messages are shown graphically in Fig. 6.

Example 2. Consider the following factorization of a function f:

$$f(X_1, X_2, X_3, X_4, X_5) = f_1(X_1, X_3) f_2(X_3, X_4, X_5) f_3(X_2, X_4) f_4(X_2),$$

the corresponding FG for the above factorization is given in figure 5. For convenience, we will apply SPA in distinct steps. At every step the respective messages will be derived.

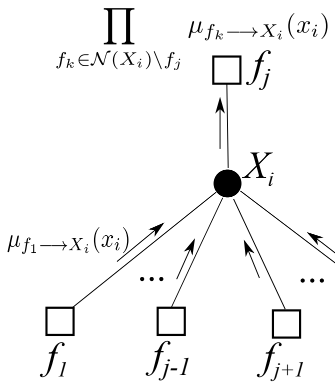


Figure 2: Update rule of the sum-product algorithm for a variable node.

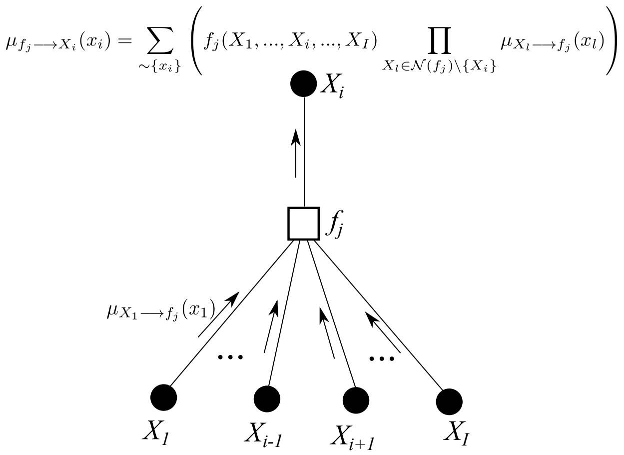


Figure 3: Update rule of the sum-product algorithm for a factor node.

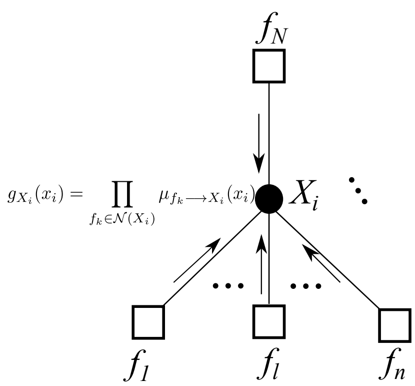


Figure 4: The marginal function for a variable node X_i with N adjacent factor nodes.

- Step 1:

$$\mu_{f_4 \longrightarrow X_2}(x_2) = \sum_{\sim\{x_2\}} f_4(x_2) = f_4(x_2),$$
$$\mu_{X_1 \longrightarrow f_1}(x_1) = 1,$$
$$\mu_{X_5 \longrightarrow f_2}(x_5) = 1.$$

The messages of this step are depicted in figure 6(a).

- Step 2:

$$\mu_{f_1 \longrightarrow X_3}(x_3) = \sum_{\sim\{x_3\}} f_1(x_1, x_3)\mu_{X_1 \longrightarrow f_1}(x_1),$$
$$\mu_{X_2 \longrightarrow f_3}(x_2) = \mu_{f_4 \longrightarrow X_2}(x_2).$$

Note that factor node f_2 has not received the incoming message from variable node X_3, in order to compute the outgoing message for variable nodes X_5 and X_4. Similarly, the incoming message from variable node X_4 is not available yet, therefore, f_2 can not send an outgoing message to variable nodes X_5 and X_3. Consequently, factor node f_2 at this step remains idle. The messages of this step are showed in figure 6(b).

- Step 3:

$$\mu_{X_3 \longrightarrow f_2}(x_3) = \mu_{f_1 \longrightarrow X_3}(x_3),$$
$$\mu_{f_3 \longrightarrow X_4}(x_4) = \sum_{\sim\{x_4\}} f_3(x_2, x_4)\mu_{X_2 \longrightarrow f_3}(x_2).$$

Factor node f_2 remains silent since its incoming messages have not arrived yet. Figure 6(c) illustrates the messages of this step.

- Step 4:

$$\mu_{X_4 \longrightarrow f_2}(x_4) = \mu_{f_3 \longrightarrow X_4}(x_4),$$
$$\mu_{f_2 \longrightarrow X_4}(x_4) = \sum_{\sim\{x_4\}} f_2(x_3, x_4, x_5)\big(\mu_{X_5 \longrightarrow f_2}(x_5)\mu_{X_3 \longrightarrow f_2}(x_3)\big).$$

The messages of this step are depicted in figure 6(d).

- Step 5:

$$\mu_{X_4 \longrightarrow f_3}(x_4) = \mu_{f_2 \longrightarrow X_4}(x_4),$$
$$\mu_{f_2 \longrightarrow X_3}(x_3) = \sum_{\sim\{x_3\}} f_2(x_3, x_4, x_5)\big(\mu_{X_5 \longrightarrow f_2}(x_5)\mu_{X_4 \longrightarrow f_2}(x_2)\big),$$
$$\mu_{f_2 \longrightarrow X_5}(x_5) = \sum_{\sim\{x_5\}} f_2(x_3, x_4, x_5)\big(\mu_{X_4 \longrightarrow f_2}(x_4)\mu_{X_3 \longrightarrow f_2}(x_3)\big).$$

- Step 6:

$$\mu_{X_3 \longrightarrow f_1}(x_3) = \mu_{f_2 \longrightarrow X_3}(x_3),$$
$$\mu_{f_3 \longrightarrow X_2}(x_2) = \sum_{\sim\{x_2\}} f_3(x_2, x_4)\mu_{X_4 \longrightarrow f_3}(x_4).$$

- Step 7:

$$\mu_{X_2 \longrightarrow f_4}(x_2) = \mu_{f_3 \longrightarrow X_2}(x_2),$$
$$\mu_{f_1 \longrightarrow X_1}(x_1) = \sum_{\sim\{x_1\}} f_1(x_1, x_3)\mu_{X_3 \longrightarrow f_1}(x_3).$$

- Termination:

$$g_{X_1}(x_1) = \mu_{f_1 \longrightarrow X_1}(x_1),$$
$$g_{X_2}(x_2) = \mu_{f_4 \longrightarrow X_2}(x_2)\mu_{f_3 \longrightarrow X_2}(x_2),$$
$$g_{X_3}(x_3) = \mu_{f_1 \longrightarrow X_3}(x_3)\mu_{f_2 \longrightarrow X_3}(x_3),$$
$$g_{X_4}(x_4) = \mu_{f_2 \longrightarrow X_4}(x_4)\mu_{f_3 \longrightarrow X_4}(x_4),$$
$$g_{X_5}(x_5) = \mu_{f_2 \longrightarrow X_5}(x_5).$$

In the last step we calculate concurrently the marginals with respect to all variables of function f. ■

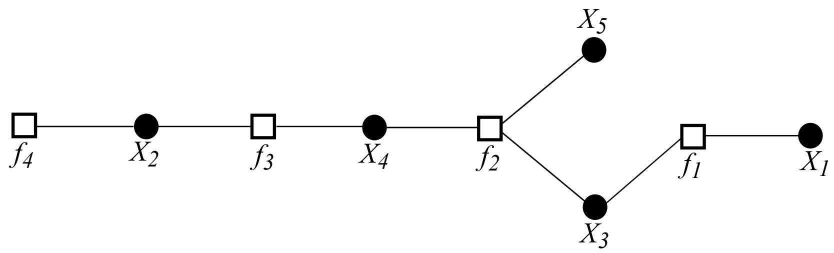
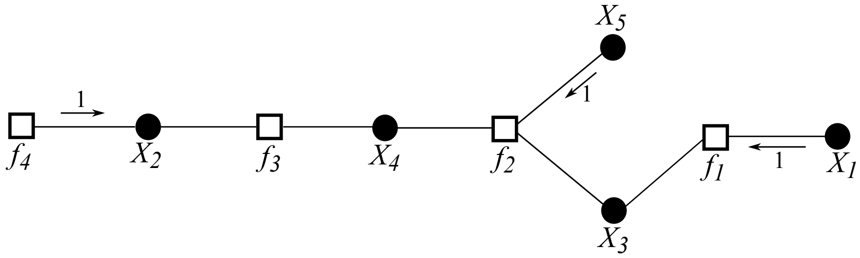


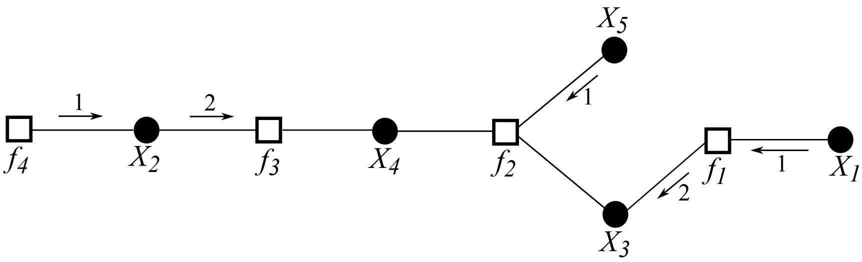
Figure 5: FG of the global function of example 2.

2.2 Normed Spaces, Contractions and Bounds

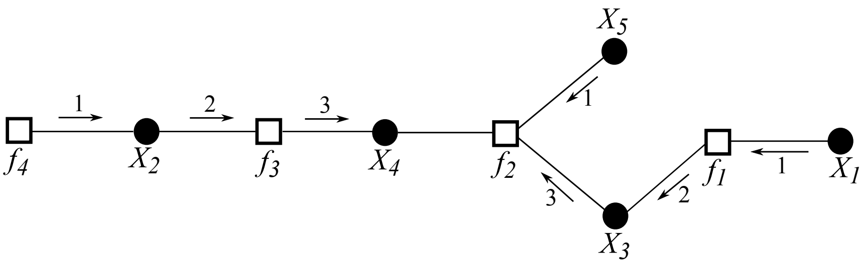
This section introduces some elementary and important properties of vector norms, matrix norms, contractions and the Mean Value theorem; all these constitute the basic mathematical tool for the subsequent Lemmas and Theorems.


X_5
1
f_4
1
X_2
f_3
X_4
f_2
f_1
1
X_1
X_3

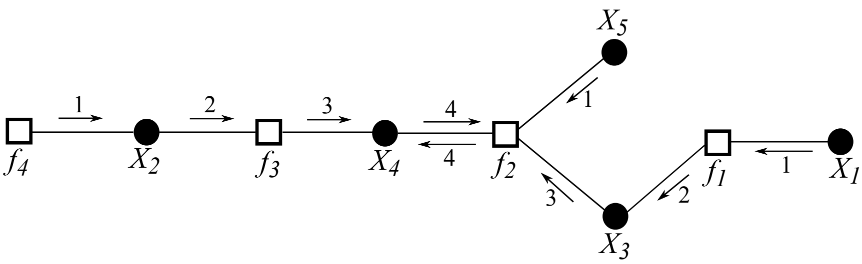
(a)


X_5
1
f_4
1
X_2
2
f_3
X_4
f_2
f_1
1
X_1
2
X_3

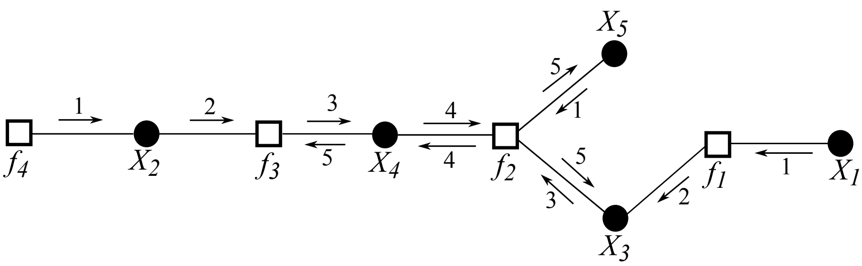
(b)


X_5
1
f_4
1
X_2
2
f_3
3
X_4
f_2
f_1
1
X_1
3
2
X_3

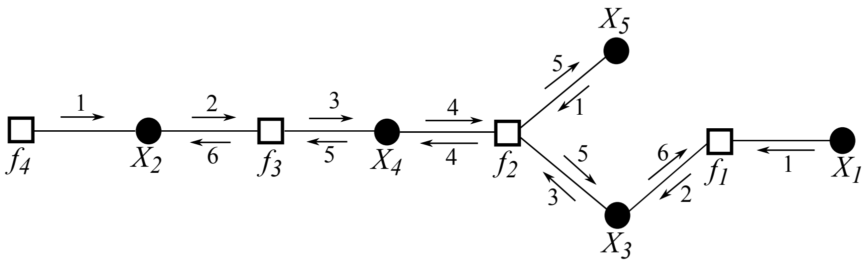
(c)


X_5
1
f_4
1
X_2
2
f_3
3
X_4
4
4
f_2
f_1
1
X_1
3
2
X_3

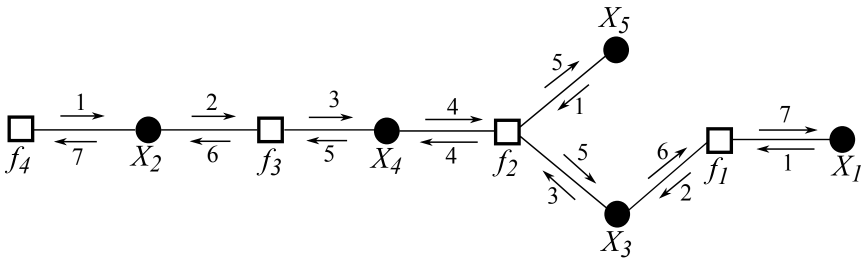
(d)


X_5
5
1
f_4
1
X_2
2
f_3
3
5
X_4
4
4
f_2
5
f_1
1
X_1
3
2
X_3

(e)



(f)



(g)

Figure 6: SPA and message-passing scheduling.

Some examples of vector norms in $\mathbb{R}^N$ are the l_1-norm and l_∞-norm which are defined as

$$||\mathbf{x}||_1 = \sum_{i=1}^{N} |x_i|, \tag{11}$$

$$||\mathbf{x}||_\infty = \max_{i \in \{1,\ldots,N\}} |x_i| \tag{12}$$

A norm on a vector space $(\mathbf{V}, ||\cdot||)$ induces a metric on $\mathbf{V}$ if $d(\mathbf{v}, \mathbf{u}) \triangleq ||\mathbf{v} - \mathbf{u}||$ satisfies the triangle inequality. The resulting metric space is complete, i.e. all Cauchy sequences converge therein. Assume $(\mathbf{X}, d)$ a metric space, a mapping $f : \mathbf{X} \mapsto \mathbf{X}$ is called contraction with respect to d if there exists a constant $L \in [0, 1)$ such that

$$d(f(\mathbf{x}), f(\mathbf{y})) \leq L \cdot d(\mathbf{x}, \mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{X} \tag{13}$$

A $||\cdot||$-contraction is called the contraction where a norm $||\cdot||$ induces d. If a metric space $(\mathbf{X}, d)$ is complete then

Theorem 1 ([1]). *Let f be a contraction of a complete metric space $(\mathbf{X}, d)$. Then f has a unique fixed point $\mathbf{x}_\infty \in \mathbf{X}$ and for any $\mathbf{x} \in \mathbf{X}$, the sequence $\mathbf{x}, f(\mathbf{x}), f(f(\mathbf{x}))$... obtained by iterating f converges to $\mathbf{x}_\infty$. The rate of convergence is at least linear, since $d(f(\mathbf{x}), \mathbf{x}_\infty) \leq L \cdot d(\mathbf{x}, \mathbf{x}_\infty)$ for all $\mathbf{x} \in \mathbf{X}$.*[3]

[3]Linear convergence means that the distance $d(f(\mathbf{x}), \mathbf{x}_\infty)$ decreases exponentially.

Proof. See Theorem 5.1.2 of [7] $\square$

Assume that $(\mathbf{V}, ||\cdot||)$ is a vector space. A *matrix norm* induced by norm, is a linear mapping $\mathbf{A} : \mathbf{V} \mapsto \mathbf{V}$, and defined as follows:

$$||\mathbf{A}|| = \sup_{\substack{\mathbf{v}\in\mathbf{V} \\ ||\mathbf{v}||\leq 1}} ||\mathbf{A}\mathbf{v}|| \tag{14}$$

Some well known matrix norms are the l_1-norm and l_∞-norm which on $\mathbb{R}^N$ induce the following norm

$$||\mathbf{A}||_1 = \max_{j\in\{1,\ldots,N\}} \sum_{i=1}^{N} |a_{i,j}|, \tag{15}$$

$$||\mathbf{A}||_\infty = \max_{i\in\{1,\ldots,N\}} \sum_{j=1}^{N} |a_{i,j}|, \tag{16}$$

where $|a_{i,j}|$ is the absolute value of the element in the ith row and jth column of $\mathbf{A}$.

Lemma 1 (Mean Value Theorem). *Let* $(\mathbf{V}, ||\cdot||)$ *be a normed space and* $f : \mathbf{V} \mapsto \mathbf{V}$ *with* f *being differentiable. Then* $\forall \mathbf{x}, \mathbf{y}$

$$||f(\mathbf{x}) - f(\mathbf{y})|| \leq ||\mathbf{x} - \mathbf{y}|| \cdot \sup_{\mathbf{z}\in[\mathbf{x},\mathbf{y}]} ||f'(\mathbf{z})||,$$

where $[\mathbf{x}, \mathbf{y}] = \{\lambda\mathbf{x} + (1-\lambda)\mathbf{y} : \lambda \in [0, 1]\}$, *i.e.* $[\mathbf{x}, \mathbf{y}]$ *is the line segment joining* $\mathbf{x}$ *and* $\mathbf{y}$.

Proof. See [8], Theorem 8.5.4. $\square$

In the above theorem $f'(\mathbf{z})$ denotes the $N \times N$ Jacobian derivative matrix. Finally if we combine Theorem 1 and Lemma 1 we obtain the following result which constitutes the basic tool hereafter.

Lemma 2. *Let* $(\mathbf{V}, ||\cdot||)$ *be a normed space,* $f : \mathbf{V} \mapsto \mathbf{V}$ *differentiable and suppose that*

$$\sup_{\mathbf{x}\in\mathbf{V}} ||f'(\mathbf{x})|| < 1.$$

Then the sequence $\mathbf{x}, f(\mathbf{x}), f(f(\mathbf{x})), \ldots$ *converges to a unique fixed point* $\mathbf{x}_\infty \in \mathbf{V}$.

Proof. By Lemma 1 f is $||\cdot||$-contraction, and hence by Theorem 1 converges to a unique fixed point $\mathbf{x}_\infty \in \mathbf{V}$. $\square$

For a square matrix $\mathbf{A}$, $\sigma(\mathbf{A})$ stands for its spectrum, namely the set of $\mathbf{A}$'s eigenvalues. By $\rho(\mathbf{A})$ is denoted the spectral radius of $\mathbf{A}$, and is designated as $\rho(\mathbf{A}) \triangleq \sup |\sigma(\mathbf{A})|$, namely, the largest modulus of eigenvalues of $\mathbf{A}$.

3 Notation of [1]

In this section it will be presented the notation of [1] for clarification and convenience in order to derive the sufficient conditions at the next Section. Firstly, the factors and the variables of the FG will be indexed by the sets $\mathcal{F}$ and $\mathcal{V} \triangleq \{1, ..., N\}$, respectively. Each variable $x_i,\ i \in \mathcal{V}$ has finite domain denoted by $\mathcal{X}_i$. The vector that contains all the variables is designated by $\mathbf{x} = [x_1, ..., x_N] \in \mathcal{X} = \prod_{i\in\mathcal{V}} \mathcal{X}_i$. Suppose $P : \mathcal{X} \mapsto \mathbb{R}_{++}$, a probability measure that can be factorized to local factors indexed by $\mathcal{F}$. Then its factorization is given by

$$P(x_1, ..., x_N) = \frac{1}{Z} \prod_{I\in\mathcal{F}} \psi_I(\mathbf{x}_I). \tag{17}$$

The factors ψ_I are indexed by subsets of $\mathcal{V}$, i.e. if $I \in \mathcal{F}$ then $I = \{i_1, ..., i_m\} \subseteq \mathcal{V}$. The variables associated with factor ψ_I are denoted as $\mathbf{x}_I = [x_{i_1}, \ldots, x_{i_m}] \in \prod_{i\in I} \mathcal{X}_i$. Furthermore, it is assumed that ψ_I is a positive function, i.e. $\psi_I : \prod_{i\in I} \mathcal{X}_i \mapsto \mathbb{R}_{++}$, (where $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$). Variable Z in Eq. 17 is a constant such that $\sum_{\mathbf{x}\in\mathcal{X}} P(\mathbf{x}) = 1$. Upper case letters will stand for indices of factors (i.e. $I, J, K... \in \mathcal{F}$), while lower case letters will be used for indices of variables (i.e. $i, j, k, ... \in \mathcal{V}$).

In FG each variable node $i \in \mathcal{V}$ is connected with all factors I that contain this variable as their argument. The set of neighbors of a variable node will be denoted as $\mathcal{N}(i) = \{I \in \mathcal{F} : i \in I\},\ \forall i \in \mathcal{V}$, while the set of neighbors of a factor node will be denoted by $\mathcal{N}(I) = \{i \in \mathcal{V} : i \in I\} = I,\ \forall I \in \mathcal{F}$. Finally, for each variable $i \in \mathcal{V}$ it is defined the set of neighboring variables, denoted by $\partial i = \mathcal{N}(\mathcal{N}(i))\backslash\{i\}$, i.e. the set of all variables that have in common with variable i at least one local factor.

The message from a factor node I to a variable node i will be denoted by $\mu^{I\to i}(x_i)$, whereas the message from a variable node i to factor node I will be denoted by $\mu^{i\to I}(x_i)$. Both messages are positive functions on $\mathcal{X}_i$, or equivalently, vectors in $\mathbb{R}_{++}^{|\mathcal{X}_i|}$ and are functions of variable x_i. A message from a node to another depends on the incoming messages. An outgoing message will be denoted as $\tilde{\mu}$ and will always depend on the incoming messages. Following this notation and observing Eqs. 6, 7, we obtain

$$\tilde{\mu}^{j\to I}(x_j) = \frac{1}{Z^{j\to I}} \cdot \prod_{J\in\mathcal{N}(j)\backslash j} \mu^{J\to j}(x_j), \tag{18}$$

$$\tilde{\mu}^{I\to i}(x_i) = \frac{1}{Z^{I\to i}} \cdot \sum_{\mathbf{x}_{I\backslash i}} \left(\psi_I(\mathbf{x}_I) \prod_{j\in I\backslash i} \mu^{j\to I}(x_j) \right), \tag{19}$$

where $Z^{j\to I}$ and $Z^{I\to i}$ are constants such that $\sum_{x_i\in\mathcal{X}_i} \tilde{\mu}^{j\to I}(x_j) = 1$ and $\sum_{x_i\in\mathcal{X}_i} \tilde{\mu}^{I\to i}(x_i)$ $= 1$. Throughout, it will be adopted the shorthand $J\backslash j$ instead of $J\backslash\{j\}$.

In FG framework, significant role plays if the FG has cycles or not. When the FG is cycle-free then SPA calculates exactly the marginals $\{P(x_i)\}_{i\in\mathcal{V}}$ and $\{P(\mathbf{x}_I)\}_{I\in\mathcal{F}}$, otherwise calculates approximations of them. Suppose an arbitrary FG and all messages, after SPA have converged to a fixed point $\boldsymbol{\mu}_\infty$, the marginals with respect to a variable $i \in \mathcal{V}$, or with respect to a local factor $I \in \mathcal{F}$ are given respectively by

$$b_i(x_i) = \frac{1}{Z_i} \cdot \prod_{I \in \mathcal{N}(i)} \boldsymbol{\mu}_\infty^{I \to i}(x_i) \approx P(x_i). \tag{20}$$

$$b_I(\mathbf{x}_I) = \frac{1}{Z_I} \cdot \psi_I(\mathbf{x}_I) \prod_{i \in I} \boldsymbol{\mu}_\infty^{i \to I}(x_i) \approx P(\mathbf{x}_I), \tag{21}$$

where Z_i and Z_I are normalization constants in order to obtain $\sum_{x_i \in \mathcal{X}_i} b_i(x_i) = 1$ and $\sum_{\mathbf{x}_i \in \mathcal{X}_I} b_I(\mathbf{x}_I) = 1$. It has been shown that always exists a fixed point [9]. However the existence of a fixed point does not necesarily means convergence towards the fixed point [10].

Finally, if we want the problem be formulated only in terms of messages $\mu^{I \to i}(\cdot)$, then, using Eqs. 18, 19 we obtain

$$\tilde{\mu}^{I \to i}(x_i) = \frac{1}{Z^{I \to i}} \cdot \sum_{\mathbf{x}_{I \setminus i}} \left(\psi_I(\mathbf{x}_I) \prod_{j \in I \setminus i} \prod_{J \in \mathcal{N}(j) \setminus I} \mu^{J \to j}(x_j) \right), \tag{22}$$

where $Z^{I \to i}$ is a constant, such that $\sum_{x_i \in \mathcal{X}_i} \tilde{\mu}^{I \to i}(x_i) = 1$. Having finished with the notation we proceed to the sufficient conditions.

4 Sufficient Conditions for Binary Variables with Pairwise Interactions

In this section are derived the sufficient conditions of the convergence of the SPA when the variables have binary domain, i.e. $x_i \in \mathcal{X}_i = \mathbb{B}$, $\forall i \in \mathcal{V}$, and all local factors have at most 2 variables as arguments. Since we assumed that all factors have pairwise interactions the set of factors can be rewritten as $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$, where $\mathcal{F}_1 = \mathcal{V}$ and $\mathcal{F}_2 \triangleq \big\{\{i, j\} : \exists \psi_I(x_i, x_j), I \in \mathcal{F}, i \neq j\big\}$, where ψ_I is a local factor that contains variables x_i and x_j. Then from Eq. 17 we take

$$P(x_1, ..., x_N) = \frac{1}{Z} \cdot \mathsf{exp} \left(\sum_{\substack{\{i,j\} \in \mathcal{F}_2 \\ i<j}} J_{i,j} x_i x_j + \sum_{i \in F_1} \theta_i x_i \right), \tag{23}$$

where θ_i and $J_{i,j}$ are such that $\psi_i(x_i) = \mathsf{exp}(\theta_i x_i)$, $\forall i \in \mathcal{F}_1$ and $\psi_{i,j}(x_i, x_j) = \mathsf{exp}(J_{i,j} x_i x_j)$, $\forall \{i, j\} \in \mathcal{F}_2$.

It is noted from Eq. 22 that the single variable factors $\mathcal{F}_1$ produce constant messages to their corresponding adjacent variables. Therefore, we focus on finding when messages converge, by studying only the factors with 2 variables. We define the following quantity, which will be employed for convenience

$$\nu^{i\to j} \triangleq \mathsf{tanh}^{-1}\left(\mu^{\{i,j\}\to j}(x_j=1) - \mu^{\{i,j\}\to j}(x_j=-1)\right). \tag{24}$$

The above message refers to the message from a variable i to a neighboring variable j via factor $\psi_{i,j}$. Note that

$$\mu^{\{i,j\}\to j}(x_j=1) - \mu^{\{i,j\}\to j}(x_j=-1) = \sum_{x_i\in\mathcal{X}_i}\left(\mu^{i\to\{i,j\}}(x_i,1) - \mu^{i\to\{i,j\}}(x_i,-1)\right) \tag{25}$$

After some algebraic manipulations (see reference [10] for intermediate operations), update equation 22 can written to the following form

$$\mathsf{tanh}(\tilde{\nu}^{i\to j}) \triangleq \mathsf{tanh}(J_{i,j})\cdot\mathsf{tanh}\left(\theta_i + \sum_{t\in\partial i\setminus j}\nu^{t\to i}\right). \tag{26}$$

where $\partial i = \{t\in\mathcal{V} : \{t,i\}\in\mathcal{F}_2\}$, that is, ∂i is the set of adjacent variables with respect to i.

It is further defined the set of order pairs $\mathcal{D}\triangleq\{i\to j : \{i,j\}\in\mathcal{F}_2\}$ and a mapping $\mathsf{f}:\mathbb{R}^{|\mathcal{D}|}\mapsto\mathbb{R}^{|\mathcal{D}|}$. We note that Eq. 26 associates a component $(\mathsf{f}(\boldsymbol{\nu}))^{i\to j}\triangleq\tilde{\nu}^{i\to j}$ in terms of the components of $\boldsymbol{\nu}\in\mathbb{R}^{|\mathcal{D}|}$. The goal is to derive sufficient conditions under which the mapping f is a contraction. The following examples clarifies all the concepts above.

Example 3. Let P be a probability measure of five variables that can be factorized to factors with pairwise interactions. Consider the following factorization (note Fig. 7):

$$\begin{aligned} P(x_1,...,x_5) &= \frac{1}{Z}\psi_1(x_1)\psi_2(x_2)\psi_3(x_3)\psi_4(x_4)\psi_5(x_5)\psi_{1,2}(\mathbf{x}_{1,2})\psi_{2,3}(\mathbf{x}_{2,3})\psi_{3,4}(\mathbf{x}_{3,4}) \\ &\quad\times\psi_{4,5}(\mathbf{x}_{4,5}) \qquad (27)\\ &= \frac{1}{Z}\mathsf{exp}\big(\theta_1x_1+\theta_2x_2+\theta_3x_3+\theta_4x_4+\theta_5x_5+J_{1,2}x_1x_2+J_{2,3}x_2x_3+J_{3,4}x_3x_4 \\ &\quad + J_{4,5}x_4x_5\big). \qquad (28)\end{aligned}$$

From the above factorization we obtain the following: $\mathcal{V}=\{1,2,3,4,5\}=\mathcal{F}_1$, while $\mathcal{F}_2=\big\{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{2,1\},\{3,2\},\{4,3\},\{5,4\}\big\}$. The set $\mathcal{D}$ is given by $\mathcal{D}=\big\{1\to 2,\ 2\to 3,\ 3\to 4,\ 4\to 5,\ 2\to 1,\ 3\to 2,\ 4\to 3,\ 5\to 4\big\}$. The vector $\boldsymbol{\nu}\in\mathbb{R}^8$ consist of all messages $\{\nu^{i\to j}\}_{i\to j\in\mathcal{D}}$, while the vector $\tilde{\boldsymbol{\nu}}\in\mathbb{R}^8$ consist of all messages $\{\tilde{\nu}^{i\to j}\}_{i\to j\in\mathcal{D}}$. The components of vector function $\mathsf{f}(\boldsymbol{\nu})$ are given by Eq. 26, namely

$$(\mathsf{f}(\boldsymbol{\nu}))^{i\to j} = \tilde{\nu}^{i\to j} = \mathsf{tanh}^{-1}\left(\mathsf{tanh}(J_{i,j})\cdot\mathsf{tanh}\left(\theta_i+\sum_{t\in\partial i\setminus j}\nu^{t\to i}\right)\right), \quad \forall\ i\to j\in\mathcal{D}.$$

■

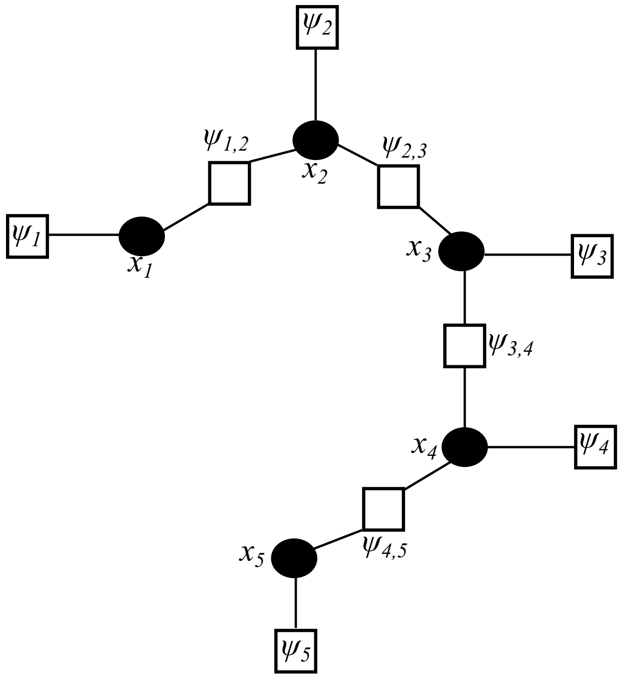


Figure 7: An FG with pairwise interactions.

After this example we present the sufficient conditions under which mapping f is a contraction mapping. Note that if mapping $\mathsf{f} : \mathbb{R}^{|\mathcal{D}|} \mapsto \mathbb{R}^{|\mathcal{D}|}$ satisfies the condition of Lemma 2 of Sec. 2 then f is a contraction. It must be noted that different choices of the vector norms in $\mathbb{R}^{|\mathcal{D}|}$ will yield different sufficient conditions for f to converge to a fixed point. Two specific examples of norms will be studied, specifically, the l_1- and l_∞- norms. The Jacobian matrix of f, f', is calculated as follows:

$$(\mathsf{f}'(\boldsymbol{\nu}))_{i\to j,k\to l} = \frac{\partial \tilde{\nu}^{i\to j}}{\partial \nu^{k\to l}} = \mathbf{A}_{i\to j,k\to l}\mathbf{B}_{i\to j}(\boldsymbol{\nu}) \tag{29}$$

where

$$\mathbf{B}_{i\to j}(\boldsymbol{\nu}) \triangleq \frac{1-\mathsf{tanh}^2(\theta_i + \sum_{t\in\partial i\backslash j}\nu^{t\to i})}{1-\mathsf{tanh}^2(\tilde{\nu}^{i\to j})} \cdot sign(J_{i,j}) \tag{30}$$

$$\mathbf{A}_{i\to j,k\to l} \triangleq \mathsf{tanh}(|J_{i,j}|)\cdot \delta_{i,l}\cdot \mathbb{I}_{\partial i\backslash j}(k). \tag{31}$$

$\mathbb{I}_{\partial i\backslash j}(k)$ is the indicator function of k to belong in set $\partial i\backslash j$, i.e. $\mathbb{I}_{\partial i\backslash j}(k) = 1$ if $k \in \partial i\backslash j$, otherwise it is 0. $\delta_{i,l}$ is the Kronecker delta function and it is 1 if $i = l$, or 0 otherwise. We derived Eq. 29 by using the facts that $\frac{\mathrm{d}\mathsf{tanh}(x)}{\mathrm{d}x} = 1-\mathsf{tanh}^2(x)$, $\frac{\mathrm{d}\mathsf{tanh}^{-1}(x)}{\mathrm{d}x} = \frac{1}{1-\mathsf{tanh}^2(x)}$, the chain rule of the derivative, and the fact that $\mathsf{tanh}(x) = sign(x)\cdot\mathsf{tanh}(|x|)$. One important observation is that matrix $\mathbf{B}$ absorbed the dependence on vector $\boldsymbol{\nu}$. The term $\mathbf{A}_{i\to j,k\to l}$ is nonnegative and hold the structure of the FG. Finally we observe that $\sup_{\boldsymbol{\nu}\in\mathbf{V}}|\mathbf{B}_{i\to j}(\boldsymbol{\nu})| = 1$

for all vectors in a vector space $\mathbf{V}$. This implies the following

$$\left|\frac{\partial \tilde{\nu}^{i\to j}}{\partial \nu^{k\to l}}\right| \leq \mathbf{A}_{i\to j,k\to l}, \tag{32}$$

for every vector $\boldsymbol{\nu}$ in $\mathbf{V}$.

Corollary 1. *For binary variables with pairwise interactions, if*

$$\max_{i\in\mathcal{V}}\left\{(|\partial i|-1)\cdot\max_{j\in\partial i}\{\mathsf{tanh}(|J_{i,j}|\}\right\}<1 \tag{33}$$

SPA is an l_∞- contraction and converges to a unique fixed point irrespective of the initial messages.

Proof.

$$\begin{aligned}
||\mathsf{f}'(\boldsymbol{\nu})||_\infty &\overset{(16)}{=} \max_{i\to j\in\mathcal{D}}\sum_{k\to l\in\mathcal{D}}\left|\frac{\partial \tilde{\nu}^{i\to j}}{\partial \nu^{k\to l}}\right| \\
&\overset{(32)}{\leq} \max_{i\to j\in\mathcal{D}}\sum_{k\to l\in\mathcal{D}}\mathbf{A}_{i\to j,k\to l} \\
&\overset{(31)}{=} \max_{i\to j\in\mathcal{D}}\sum_{k\to l\in\mathcal{D}}\mathsf{tanh}(|J_{i,j}|)\cdot\delta_{i,l}\cdot\mathbb{I}_{\partial i\setminus j}(k)
\end{aligned}$$

Now we use the fact that $\max\limits_{i\to j\in\mathcal{D}}\equiv\max\limits_{i\in\mathcal{V}}\max\limits_{j\in\partial i}$, and the fact that we sum over all k,l such that $k\in\partial i\setminus j$ and $l=i$, and thus we obtain

$$\begin{aligned}
||\mathsf{f}'(\boldsymbol{\nu})||_\infty &\leq \max_{i\in\mathcal{V}}\max_{j\in\partial i}\sum_{k\in\partial i\setminus j}\mathsf{tanh}(|J_{i,j}|) \\
&= \max_{i\in\mathcal{V}}\max_{j\in\partial i}(|\partial i|-1)\cdot\mathsf{tanh}(|J_{i,j}|) \\
&= \max_{i\in\mathcal{V}}\left\{(|\partial i|-1)\cdot\max_{j\in\partial i}\{\mathsf{tanh}(|J_{i,j}|\}\right\}.
\end{aligned}$$

Finally, we apply Lemma 2 and the proof is completed $\square$

Corollary 2. *For binary variables with pairwise interactions, if*

$$\max_{i\in\mathcal{V}}\max_{k\in\partial i}\sum_{j\in\partial i\setminus k}\mathsf{tanh}(|J_{i,j}|)<1 \tag{34}$$

SPA is an l_1- contraction and converges to a unique fixed point irrespective of the initial messages.

Proof. Similar to the proof of Corollary 2, we obtain

$$
\begin{aligned}
||\mathsf{f}'(\boldsymbol{\nu})||_1 &\overset{(15)}{=} \max_{k\to l\in\mathcal{D}} \sum_{i\to j\in\mathcal{D}} \left|\frac{\partial\tilde{\nu}^{i\to j}}{\partial\nu^{k\to l}}\right| \\
&\overset{32),(31)}{\leq} \max_{k\to l\in\mathcal{D}} \sum_{i\to j\in\mathcal{D}} \mathsf{tanh}(|J_{i,j}|)\cdot\delta_{i,l}\cdot\mathbb{I}_{\partial i\setminus j}(k) \\
&= \max_{i\in\mathcal{V}} \max_{k\in\partial i} \sum_{j\in\partial i\setminus k} \mathsf{tanh}(|J_{i,j}|)
\end{aligned}
$$

We apply Lemma 2 and the proof is completed. □

It is observed in such case, that condition 33 implies 34, but not conversely, hence l_1-norm results a tighter bound than the l_∞-norm.

Another way to derive sufficient conditions for convergence of SPA comes from spectral theory, and specifically from spectral radius of matrix in (31). The basic observation is that we must consider several iterations of SPA. This entails stronger condition for the convergence of SPA to a unique fixed point.

Lemma 3. *Let $(\mathbf{X}, d)$ be a metric space and a $\mathsf{f} : \mathbf{X} \mapsto \mathbf{X}$ a mapping. We define*

$$
\mathsf{f}^{(N)}(\cdot) \triangleq \underbrace{\mathsf{f}(\mathsf{f}(\cdots\mathsf{f}(\cdot)))}_{N\ times}.
$$

Suppose that $\mathsf{f}^{(N)}$ is a d-contraction for some $N \in \mathbb{N}$. Then f has a unique fixed point $\mathbf{x}_\infty$, and for any $\mathbf{x} \in \mathbf{X}$, the sequence $\mathbf{x}, \mathsf{f}(\mathbf{x}), \mathsf{f}^{(2)}(\mathbf{x}), ...$ converges to $\mathbf{x}_\infty$.

Proof. Take any $\mathbf{x} \in \mathbf{X}$ and further consider the N-sequences obtained by $\mathsf{f}^{(N)}$, by the following manner

$$
\begin{aligned}
&\mathbf{x}, \mathsf{f}^{(N)}(\mathbf{x}), \mathsf{f}^{(2N)}(\mathbf{x}), ... \\
&\mathsf{f}(\mathbf{x}), \mathsf{f}^{(N+1)}(\mathbf{x}), \mathsf{f}^{(2N+1)}(\mathbf{x}), ... \\
&\vdots \\
&\mathsf{f}^{(N-1)}(\mathbf{x}), \mathsf{f}^{(2N-1)}(\mathbf{x}), \mathsf{f}^{(3N-1)}(\mathbf{x}), ...
\end{aligned}
$$

We note that each sequence converges to $\mathbf{x}_\infty$ since $\mathsf{f}^{(N)}$ is d-contraction with fixed point $\mathbf{x}_\infty$. This implies that sequence $\mathbf{x}, \mathsf{f}(\mathbf{x}), \mathsf{f}^{(2)}(\mathbf{x}), ...$ converges to $\mathbf{x}_\infty$. □

Theorem 2. *Let $\mathsf{f} : \mathbb{R}^m \mapsto \mathbb{R}^m$ be a differentiable and suppose that $\mathsf{f}'(\mathbf{x}) = \mathbf{B}(\mathbf{x})\mathbf{A}$, where $\mathbf{A}$ has non negative entries and $\mathbf{B}(\mathbf{x})$ is diagonal, and further its diagonal entries are bounded, with $|\mathbf{B}_{i,i}(\mathbf{x})| \leq 1$. If $\rho(\mathbf{A}) < 1$, then for any $\mathbf{x} \in \mathbb{R}^m$, the sequence $\mathbf{x}, \mathsf{f}(\mathbf{x}), \mathsf{f}^{(2)}(\mathbf{x}), ...$ converges to $\mathbf{x}_\infty$, which does not depend on $\mathbf{x}$.*

Proof. For the proof we will use some basic algebraic operations. For a matrix $\mathbf{B}$, the matrix $|\mathbf{B}|$ will be designated the matrix with entries $|\mathbf{B}|_{i,j} = |\mathbf{B}_{i,j}|$. For 2 matrices $\mathbf{B}, \mathbf{C}$ we define the following properties:

$$\mathbf{B} \leq \mathbf{C} \Longleftrightarrow \mathbf{B}_{i,j} \leq \mathbf{C}_{i,j}, \quad \forall i, j, \tag{35}$$

$$|\mathbf{B}| \leq |\mathbf{C}| \overset{(35)}{\Longrightarrow} ||\mathbf{B}||_1 \leq ||\mathbf{C}_{i,j}||_1, \tag{36}$$

$$|\mathbf{BC}| \leq |\mathbf{B}| \cdot |\mathbf{C}|, \tag{37}$$

$$\text{if } \mathbf{0} \leq \mathbf{A} \text{ and } \mathbf{B} \leq \mathbf{C} \Longrightarrow \mathbf{AB} \leq \mathbf{AC} \text{ and } \mathbf{BA} \leq \mathbf{CA}. \tag{38}$$

Utilizing the above and the chain rule of the differentiation, $\forall n = 1, 2, 3, ...$ and any $\mathbf{x} \in \mathbb{R}^m$

$$\begin{aligned} |(\mathsf{f}^{(n)})'(\mathbf{x})| &= \left| \prod_{i=1}^{n} \mathsf{f}'(\mathsf{f}^{(i-1)}(\mathbf{x})) \right| \\ &\underset{\mathsf{f}'(\cdot)=\mathbf{B}(\cdot)\mathbf{A}}{\overset{(37)}{\leq}} \prod_{i=1}^{n} \left(\left| \mathbf{B} \left(\mathsf{f}^{(i-1)}(\mathbf{x}) \right) \mathbf{A} \right| \right) \\ &\underset{|\mathbf{B}_{i,i}(\mathbf{x})| \leq 1}{\overset{37),(38)}{\leq}} \mathbf{A}^n. \end{aligned}$$

Consequently, by property of Eq. 36 we obtain that $||(\mathsf{f}^{(n)})'(\mathbf{x})||_1 \leq ||\mathbf{A}^n||_1$. By Gel'fand spectral radius theorem that states

$$\lim_{n \longrightarrow \infty} \left(||\mathbf{A}^n||_1 \right)^{1/n} = \rho(\mathbf{A}),$$

we choose a positive ϵ such that $\rho(\mathbf{A}) + \epsilon < 1$, then $\exists N : ||\mathbf{A}^N||_1 \leq (\rho(\mathbf{A}) + \epsilon)^N < 1$. Hence, $\forall \mathbf{x} \in \mathbb{R}^m$, $||(\mathsf{f}^{(N)})'(\mathbf{x})||_1 < 1$, and by applying Lemma 2, we obtain that $\mathsf{f}^{(N)}$ is an l_1-contraction. Finally Lemma 3 is applied on $\mathsf{f}^{(N)}$ and the proof is completed. □

Corollary 3. *For binary variables with pairwise interactions, SPA converges to a unique fixed point irrespective of the initial messages, if the spectral radius of $|\mathcal{D}| \times |\mathcal{D}|$- matrix*

$$\mathbf{A}_{i \to j, k \to l} \triangleq \mathsf{tanh}(|J_{i,j}|) \cdot \delta_{i,l} \cdot \mathbb{I}_{\partial i \setminus j}(k)$$

is strictly smaller than 1.

Proof. If we combine Eqs. 29 (Eq. 29 indicates that we multiply every element in a specific row of matrix $\mathbf{A}$ with the corresponding element of the diagonal of $\mathbf{B}$, i.e. we have a matrix multiplication of $\mathbf{B}$ with $\mathbf{A}$), 30, 31 and Theorem 2 the proof is immediate. □

The last proof completes the sufficient conditions for convergence of SPA.

5 Conclusions

This work derived sufficient conditions for the convergence of the SPA to a fixed point, when the factors of FG have pairwise interactions and variables have binary domain. The report was for the graduate course Functional Analysis and studied the aforementioned problem using the principles learned from theory lessons. The text was based on [1] with respect to Sections 3, 4 and subsection 2.2, whereas regarding subsection 2.1 the text was based on [5].

References

[1] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *IEEE Trans. Inf. Theor.*, vol. 53, no. 12, Dec. 2007.

[2] T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge University Press, 2008.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transaction on Information Theory*, vol. 47, pp. 498–519, 1998.

[4] H. Loeliger, "An Introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[5] P. N. Alevizos, "Factor graphs: Theory and applications," Chania, Tech. Rep., Sep. 2012. [Online]. Available: http://users.isc.tuc.gr/~palevizos/palevizos_thesis.pdf

[6] R. G. Gallager, *Low-Density Parity-Check Codes.* MIT Press, 1963.

[7] E. Jreyszig, *Introductory Functional Analysis with Applications.* John Willey & Sons. Inc., 1978.

[8] J. Dieudonne, *Foundations of Modern Analysis.* New York: Academic Press, 1969.

[9] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theor.*, vol. 51, pp. 2282–2312, 2005.

[10] J. M. Mooij and H. J. Kappen, "On the properties of the bethe approximation and loopy belief propagation on binary networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.