Self-Guided Belief Propagation – A Homotopy Continuation Method

Christian Knoll, Associate Member, IEEE, Adrian Weller, and Franz Pernkopf, Senior Member, IEEE

Abstract—Belief propagation (BP) is a popular method for performing probabilistic inference on graphical models. In this work, we enhance BP and propose self-guided belief propagation (SBP) that incorporates the pairwise potentials only gradually. This homotopy continuation method converges to a unique solution and increases the accuracy without increasing the computational burden. We provide a formal analysis to demonstrate that SBP finds the global optimum of the Bethe approximation for attractive models where all variables favor the same state. Moreover, we apply SBP to various graphs with random potentials and empirically show that: (i) SBP is superior in terms of accuracy whenever BP converges, and (ii) SBP obtains a unique, stable, and accurate solution whenever BP does not converge.

Index Terms—Graphical models, belief propagation, probabilistic inference, sum-product algorithm, partition function, inference algorithms.

1 INTRODUCTION

Computing the marginal distributions and evaluating the partition function are two fundamental problems in the context of probabilistic graphical models. Both problems can be solved efficiently for tree-structured models but are NP-hard if the graphical model contains loops [7]. Since loops are present in many problems of relevance there is a need for efficient approximation methods.

Belief propagation (BP) provides a way to approximate the marginal distribution and may be viewed as an approach to try to minimize the Bethe free energy $F_B$ [61]. BP has a long success story in many applications, including computer vision, speech processing, social network analysis, and error-correcting codes [19], [25], [35]. But, despite its success, BP does not always approximate the marginals well and may even fail to converge. Reasons for BP’s failure include the existence of multiple fixed points of varying accuracy (where it depends on implementation details to which one BP converges) and of unstable fixed points that cause BP to oscillate far away from any fixed point [16], [33], [51]. These limitations motivate the search for modifications of BP that have better convergence properties and find more accurate marginals.

One alternative is to perform a different optimization problem that minimizes the Bethe free energy $F_B$, instead of applying BP [61]. While, for certain problem classes, polynomial-time algorithms exist, it is even problematic to approximate the global minimum for models with arbitrary potentials [5], [42], [55]. Hence, the pursuit for methods that approximate the marginals well – ideally with both run-time and convergence guarantees – is still ongoing.

In this work, we present self-guided belief propagation (SBP) to help address this gap. The observation that strong pairwise potentials often reduce accuracy and worsen the convergence properties [21] inspired us to construct a homotopy; i.e., we first consider only local potentials (where BP is exact) and subsequently modify the model by gradually increasing the pairwise potentials to their given values. SBP thus solves a deterministic sequence of models that iteratively refines the Bethe approximation towards a solution that is uniquely defined by the initial model.

We empirically evaluate SBP for grid-graphs, complete graphs, and random graphs with random potentials. Compared to BP, we observe superior performance in terms of accuracy; in fact, SBP achieves more accurate results than Gibbs sampling in a fraction of run-time. Additionally, SBP exhibits favorable convergence properties and excels for hard problems, for which – despite BP failing to converge – SBP provides accurate results. We also analyze SBP theoretically and demonstrate optimality of the selected fixed point for attractive models with unidirectional local potentials.1

It is worth emphasizing that SBP is a relatively straightforward modification to standard BP that can be easily added to existing BP implementations. We thus expect that the ease of use lowers the hurdle for practical applications.

The paper is structured as follows: Section 2 provides background on probabilistic graphical models, belief propagation, and methods that minimize the free energy. Moreover, we present a unified taxonomy of model-classes according to their complexity. Our proposed algorithm is presented in Section 3. We evaluate SBP and discuss empirical observations in Section 4 and provide a more formal analysis in Section 5. Finally, we conclude the paper in Section 6.

2 BACKGROUND

In this section, we briefly introduce probabilistic graphical models and specify the models considered in this work. We further

1. An attractive model is a probabilistic graphical model where all pairwise interactions are specified by positive couplings; a general model also contains pairwise interactions with negative couplings. Unidirectional local potentials are potentials that all favor the same state (see Section 2.2).
introduce the BP algorithm and its connection to the Bethe approximation. Finally, we summarize important properties regarding the solution space of BP as a foundation for our theoretical analysis of SBP.

2.1 Probabilistic Graphical Models
Let us consider an undirected graph $G = (X, E)$, where $X = \{X_1, \ldots, X_N\}$ is the set of nodes, and $E$ is the set of undirected edges. Then, two nodes $X_i$ and $X_j$ are joined by an edge if $(i, j) \in E$; note that we consider each edge only once, i.e., $(i, j) = (j, i)$. We further denote the set of neighbors of $X_i$ by $N(i) = \{X_j \in X : (i, j) \in E\}$.

Let us define a probabilistic graphical model $U = (G, \Psi)$, where $\Psi = \{\Phi(x_1), \ldots, \Phi(x_K)\}$ is the set of all $K$ potentials and $X$ is the set of random variables. In this work we focus on pairwise models where all potentials consist of two variables at most, so that the joint distribution $P_X(x)$ factorizes according to

$$P_X(x) = \frac{1}{Z} \prod_{(i,j)\in E} \Phi(x_i, x_j) \prod_{i=1}^N \Phi(x_i), \quad (1)$$

where the potentials specify the energy $E(x)$.

We consider the following two problems: First, evaluating the partition function $Z$ (i.e., the normalization function of the joint distribution). Consider any distribution $P_X(x)$ and the (Gibbs) free energy $\mathcal{F} = \mathcal{F}(Q_X(x)) = \sum_x Q_X(x) (E(x) + \ln Q_X(x))$; then, evaluating the partition function and minimizing the free energy are equivalent since the minimum of $\mathcal{F}$ corresponds to $Z$ according to $\min \mathcal{F} = \mathcal{F}^* = -\ln Z$ (cf. [61]).

Second, we consider computing the marginal distribution

$$P_Y(y) = \sum_{x_i: x_i \in (X \setminus Y)} P_X(x), \quad (2)$$

where $Y \subset X$ may be any set of RVs. The identities of the random variables are often obvious from the notation of the values; therefore, we often use the shorthand notation $P(y)$ for the marginal distribution. Note that the problems of evaluating the partition function and computing the marginal distributions are inherently linked, as $\mathcal{F}$ obtains its minimum precisely for the marginal distributions $P_X(x)$.

In general, both problems are intractable [7]. Consequently one often relaxes the problems and only approximates the marginal distributions and the partition function. This allows for an elegant iterative algorithm that was discovered multiple times in different fields. It is known as belief propagation (BP) in computer science, the sum-product algorithm in information theory, and the cavity- or the Bethe-method in physics; we refer the reader to [26], [30] for a good overview of the underlying principles.

2.2 MODEL SPECIFICATION
We consider binary pairwise models where every random variable $X_i$ takes values from $\mathcal{X} = \{-1, +1\}$ (i.e., Ising models). For these models, instead of considering the singleton marginals $P(x_i)$ and the pairwise marginals $P(x_i, x_j)$ explicitly, it is often more convenient to work with the means $m_i$ and the correlations $\chi_{ij}$ defined according to

$$m_i = \mathbb{E}(X_i) = P_X(+1) - P_X(-1), \quad (3)$$

$$\chi_{ij} = \mathbb{E}(X_i X_j) = \sum_{x_i, x_j} x_i x_j P(x_i, x_j). \quad (4)$$

Let us define couplings $J_{ij} \in \mathbb{R}$ that are assigned to each edge $(i, j) \in E$ and local fields $\theta_i \in \mathbb{R}$ that act on each variable $X_i \in X$. These parameters then define the pairwise potentials $\Phi(x_i, x_j) = \exp(J_{ij} x_i x_j)$ and the local potentials $\Phi(x_i) = \exp(\theta_i x_i)$ so that the corresponding joint distribution from (1) can be written as

$$P_X(x) = \frac{1}{Z} \exp \left( \sum_{(i,j)\in E} J_{ij} x_i x_j + \sum_{i=1}^N \theta_i x_i \right). \quad (5)$$

There are two different types of interactions between random variables: if $J_{ij}$ is negative then the edge $(i, j)$ is repulsive; if $J_{ij}$ is positive then the edge $(i, j)$ is attractive. Accordingly, one distinguishes between general models that contain both attractive and repulsive edges and attractive models that contain only attractive edges.

Attractive models are often more well-behaved than general models but, rather than only distinguishing between these two model-types, we find it beneficial to consider an even finer-grained distinction into the following three model-classes. In doing so, we additionally unify different naming conventions in the literature and provide a consistent taxonomy of model-classes.

2.2.1 Frustrated, Balanced, and Unidirectional Models
First, frustrated models are models that contain cycles for which the product over all couplings equates to a negative number, i.e., cycles with an odd number of repulsive edges (see Fig. 1a).\(^3\) Note that if a model is frustrated, it must be a general model per definition.

Second, balanced models are models that are not frustrated. For a balanced model, it is possible to flip a certain subset of variables without affecting the represented distribution such that the resulting model is attractive (cf. [52]). To flip a variable $X_i$, one must switch its two states, and: reverse the sign of the corresponding local field $\theta_i$; and reverse the signs of the couplings $J_{ij}$ for all incident edges $(i, j) : X_j \in N(i)$. For an illustrative example, consider the model in Fig. 1b. This model is a general one in its current form; reversing the sign of the rightmost variable further implies changing both attached edges from repulsive to attractive, which results in an equivalent attractive model (see Fig. 1c).

2. Note that attractive models are also known as ferromagnetic models [30] or log-supermodular models [40].
3. Frustrated models are also referred to as spin glasses in the physics literature. These are some of the most complex models considered of the form (5).
Finally, unidirectional models are attractive models with unidirectional local potentials, i.e., either all $\theta_i \leq 0$ or $\theta_i \geq 0$ (see Fig. 1d). In a similar manner as for balanced models, certain general models have equivalent unidirectional models, i.e., by flipping variables to render the model attractive all local potentials are mirrored so that the attractive model is unidirectional; we call these models quasi-unidirectional. All theoretical findings for unidirectional models thus readily extend to quasi-unidirectional models.

The major advantage of considering unidirectional models is that they are relatively well understood. Thus unidirectional models are of particular relevance for our work – as well as for studying inference algorithms in general.

Intuitively, in unidirectional models, we expect that all variables will energetically favor the same state as indicated by the local potentials. Additionally, since all edges are attractive, no contradictions between neighboring variables will occur, thus further reinforcing this behavior. Moreover, various research directions indicate the relevance of this model class. We will list a couple of these indications below:

On the one hand, it is generally often easier to study homogeneous models, where all pairwise potentials are the same (i.e., $J_{ij} = J$) and all local potentials are the same (i.e., $\theta_i = \theta$). Provided that the model is unidirectional, the findings from homogeneous models carry over readily [12], [21]. Alternatively, one often considers models of the form (5) in an external field $\theta$, i.e., where all local potentials are the same (or have the same sign) [17]. Keeping the physical interpretation in mind, we might hope that models that are placed in an external field are easier to understand. Let us use the concept of flipping variables for transforming balanced models into attractive models again. Interestingly, if we consider models in an external field only, then flipping variables is not a viable option as this would imply reversing the sign of some $\theta_i$. Thus for models in a constant external field, the class of attractive models reduces precisely to the class of unidirectional models.

On the other hand, the complexity of approximating the partition function also depends on the model class and is in line with our distinction. That is for unidirectional models the partition function can be approximated efficiently in polynomial time [17]; for frustrated models, the partition function cannot be approximated efficiently; balanced models are of intermediate complexity [12].

Finally, it is sometimes beneficial to study models without local potentials; in fact, one can find an auxiliary model without local potentials for every model of the form (5) by adding one variable to the graph (and connecting it to all existing variables), where the auxiliary model has the same marginals and the same partition function (up to a constant factor of two) [18], [41], [53]. Provided that the original model was attractive, then the auxiliary model remains attractive (or can be rendered attractive by flipping the additional variable) if and only if the original model was unidirectional.

2.3 BELIEF PROPAGATION (BP)

BP approximates the marginals by recursively exchanging messages between random variables. The messages $\mu_{ij}^{n+1}(x_j)$ from $X_i$ to $X_j$ at iteration $n + 1$ are updated according to (6) and are normalized so that $\sum_{x_j \in X} \mu_{ij}^{n+1}(x_j) = 1$.

$$\mu_{ij}^{n+1}(x_j) \propto \sum_{x_i \in X} \Phi(x_i, x_j) \prod_{k \in N(i) \setminus j} \mu_{ki}^{n}(x_i).$$  

Let $\mu^n = \{\mu_{ij}^n(x_j) : e_{ij} \in E, x_j \in X\}$ be the set of all messages at iteration $n$ and let the mapping induced by (6) be denoted as $\mu^{n+1} = B\mu^n$. If all successive messages remain unchanged, i.e., if $\mu^{n+1} = \mu^n$, then BP is converged to a fixed point $\mu^\infty$. We further write $\mu^\infty = B\mu^{\infty \! \! \! \circ}$, where $\mu^{\infty \! \! \! \circ}$ updates the messages until convergence. If BP fails to converge and the messages oscillate, one can try to achieve convergence by either changing the update-rule [10], [23], [47], or by replacing the messages with a convex combination of the last messages [34]. The latter method is known as damping (BP$\rho$) where a damping parameter $\epsilon \in [0, 1)$ specifies the new update rule

$$\mu^{n+1} = B\mu^n + (1 - \epsilon)\mu^n + \epsilon\mu^n.$$

After convergence, the singleton marginals $\tilde{P}(x_i)$ and pairwise marginals $\tilde{P}(x_i, x_j)$ are approximated by

$$\tilde{P}(x_i) = \frac{1}{Z_i} \Phi(x_i) \prod_{k \in N(i)} \mu_{ki}^\infty(x_i),$$  

$$\tilde{P}(x_i, x_j) = \frac{1}{Z_{ij}} \Phi(x_i) \Phi(x_j) \prod_{k \in N(i) \setminus j} \mu_{ki}^\infty(x_i) \prod_{l \in N(j) \setminus i} \mu_{lj}^\infty(x_j).$$

where $Z_i, Z_{ij} \in \mathbb{R}_{+}^*$ guarantee that all probabilities sum to one. We refer to the set of singleton and pairwise marginals as pseudomarginals $\tilde{P}_B$, where

$$\tilde{P}_B = \{\tilde{P}(x_i), \tilde{P}(x_i, x_j) : x_i \in X, (i, j) \in E\}.$$

We further denote the pseudomarginals obtained at a fixed point of BP by $\tilde{P}^\infty_B$.  

2.4 THE BETHE APPROXIMATION & RELATED WORK

BP is closely connected to variational methods and concepts from statistical physics. In particular, a direct relationship between performing BP and minimizing the Bethe free energy exists (cf. [61]).

The Bethe free energy $F_B(\tilde{P}_B)$ is evaluated over the pseudomarginals (i.e., the singleton- and pairwise marginals) and is a

5. On the one hand, we need to allow for a flip of the additional variable for the case that all $\theta_i < 0$; on the other hand, this makes the statement agnostic to the definition (i.e., it holds irrespective if we have $X = \{-1, +1\}$ or $X = \{0, 1\}$).

6. With slight abuse of notation, we will refer to both the pseudomarginals $\tilde{P}_B$ and the messages $\mu^n$ as fixed points throughout this work. Strictly speaking, however, the pseudomarginals are only evaluated at a fixed point rather than being a fixed point themselves.
function of the average energy $E_B(\tilde{P}_B)$ and the Bethe entropy $S_B(\tilde{P}_B)$ that are defined by

$$E_B(\tilde{P}_B) = -\sum_{x_i} \sum_{x_j} \tilde{P}(x_i) \ln \Phi(x_i) - \sum_{(i,j) \in E} \sum_{x_i,x_j} \tilde{P}(x_i,x_j) \ln \Phi(x_i,x_j).$$

(11)

$$S_B(\tilde{P}_B) = -\sum_{(i,j) \in E} \sum_{x_i,x_j} \tilde{P}(x_i,x_j) \ln \tilde{P}(x_i,x_j) + \sum_{x_i} \left(\left|N(i)\right| - 1\right) \sum_{x_j} \tilde{P}(x_i) \ln \tilde{P}(x_i).$$

(12)

This subsequently defines the Bethe free energy according to

$$\mathcal{F}_B(\tilde{P}_B) = E_B(\tilde{P}_B) - S_B(\tilde{P}_B).$$

(13)

Let us define the local polytope $L$, i.e., the set of consistent pseudomarginals, where

$$L = \{\tilde{P}_B : \sum_{x_i} \tilde{P}(x_i) = 1, \sum_{x_j} \tilde{P}(x_i,x_j) = \tilde{P}(x_i)\}.$$

Minimizing $\mathcal{F}_B$ over the local polytope gives us the desired global minimum $\mathcal{F}_B = \min_L \mathcal{F}_B(\tilde{P}_B)$. This, however, is not straightforward as the constrained $\mathcal{F}_B$ is often non-convex.

The main reason for considering the Bethe free energy is its immediate connection with the fixed points of BP $\mu^\ast$ (and the associated pseudomarginals $\tilde{P}_B$); all stationary points $\mathcal{F}_B$ are in a one-to-one correspondence with the fixed points of BP. More precisely, we have

$$\mathcal{F}_B = \mathcal{F}_B(\tilde{P}_B).$$

(14)

In practice, we are particularly interested in stable fixed points (for which BP converges if initialized close enough); we index all stable fixed points $\tilde{P}_B^s$ by $s = 1, \ldots, S$ and denote the set of all stable fixed points by $S$. Likewise, we consider the set of all fixed points that constitute minima of the Bethe free energy $M$ and index them by $m = 1, \ldots, M$. Note that every stable fixed point $\tilde{P}_B^s$ must be a minimum of $\mathcal{F}_B$; the converse, however, need not be the case, i.e., not every local minimum corresponds to a stable fixed point, so that $S \subseteq M$.

This correspondence between BP and $\mathcal{F}_B$ led to a better understanding of BP and inspired plenty of methods that minimize $\mathcal{F}_B$ directly [58], [62]. The minimization, however, is still highly non-trivial and requires good approximation methods in practice. Since strong pairwise potentials often reduce the accuracy of the Bethe approximation and are responsible for its non-convexity [20], one can correct the entropy term (12) by accounting for the strong potentials; this admits convex relaxations that provide provable convergent message passing algorithms [11], [15], [28], [29], [49]. There is, however, a trade-off between convergence-properties and accuracy in general. That is, if $\mathcal{F}_B$ is to be minimized, the Bethe approximation often provides accurate results and outperforms convex free energy approximations [29], [55]. Thus, it is a relevant problem to directly approximate $\mathcal{F}_B$ in a way that allows for efficient minimization. Polynomial run-time algorithms exist that approximate $\mathcal{F}_B$ for restricted models: these include sparsity constraints [42] or require attractive models [55]. If both properties are fulfilled, i.e., for locally tree-like attractive models the Bethe approximation is exact and can be optimized efficiently [9]. Note that $\mathcal{F}_B$ provides an upper bound on $\mathcal{F}$ for attractive models [40], [56], [60].

We aim to estimate $\mathcal{F}_B$ and $P_B$; while the approximation of $\mathcal{F}_B$ in [55] is $\varepsilon$-accurate, no run-time guarantees exist for general models. Our work, on the contrary, provides an estimate in constant run-time (see Theorem 6 in Section 5); the approximation error, however, cannot be made arbitrarily small in general. Both methods overcome their respective disadvantages when restricting the models; i.e., both methods do efficiently minimize the Bethe approximation for unidirectional models.3

2.5 Solution Space of Binary Pairwise Models

Over the years, a substantial amount of work has been devoted to getting a good understanding of how the solution space is affected by the specification of the graphical model. We are particularly interested in how the performance of BP and the energy landscape of $\mathcal{F}_B$ vary, depending on the specification of the potentials. Here, we briefly summarize the most important results with a particular focus on how the solution space changes when considering increasingly stronger couplings.

This discussion serves two purposes: on the one hand, it provides a good intuition and motivates the working principle of SBP discussed in Section 3; on the other hand, it prepares the required background for our theoretical analysis in Section 5. Moreover, the difference in the behavior of BP is often only considered between general and attractive models. Instead, we consider the difference in BP’s behavior between frustrated, balanced, and unidirectional models. We are convinced that this distinction is not only relevant for analyzing SBP in this work but that this is a crucial distinction one must make in the pursuit of understanding what makes some problems inherently hard to solve.

For as long as the couplings $J_{ij}$ are sufficiently small (relative to $\theta_i$), the Bethe free energy $\mathcal{F}_B$ is convex, BP has a unique fixed point, and BP converges [20], [33]. As the couplings increase, the performance tends to suffer as multiple fixed points emerge and as BP may fail to converge (even if the fixed point is still unique) [21], [33].8 The precise effect of the strong couplings on the behavior of BP and the shape of $\mathcal{F}_B$ mainly depends on the respective model-class. Thus, we will split the following discussion accordingly.

For frustrated models, the number of stationary points scales exponentially with the number of variables. Often, multiple stationary points coexist with the same value of $\mathcal{F}_B$ but possibly completely different marginals; it may even happen that there is not a single unambiguous global minimum. Moreover, for frustrated models, the Bethe free energy is not lower bounded by the exact free energy, so that the global minimum need not approximate the free energy best. Such complex energy landscapes render any attempt of minimizing the Bethe free energy futile. Moreover, in particular for models with strong couplings, BP often fails to converge for frustrated models.

For balanced models, the number of stationary points still scales exponentially with the number of variables. Often, multiple stationary points coexist with the same value of $\mathcal{F}_B$ but possibly completely different marginals; it may even happen that there is not a single unambiguous global minimum. Moreover, for balanced models, the Bethe free energy is not lower bounded by the exact free energy, so that the global minimum need not approximate the free energy best. Such complex energy landscapes render any attempt of minimizing the Bethe free energy futile. Moreover, in particular for models with strong couplings, BP often fails to converge for frustrated models.

For balanced models, the number of stationary points still scales exponentially with the number of variables. Often, multiple stationary points coexist with the same value of $\mathcal{F}_B$ but possibly completely different marginals; it may even happen that there is not a single unambiguous global minimum. Moreover, for balanced models, the Bethe free energy is not lower bounded by the exact free energy, so that the global minimum need not approximate the free energy best. Such complex energy landscapes render any attempt of minimizing the Bethe free energy futile. Moreover, in particular for models with strong couplings, BP often fails to converge for frustrated models.

7. The restriction to unidirectional models is actually too conservative for [55], where the approximation of $\mathcal{F}_B$ takes polynomial run-time for balanced models as well.

8. Note that the differences in behavior coincide with different phases in statistical mechanics (cf. [30], [48], [64]).
easier to solve. Since every balanced model can be transformed into an equivalent attractive model, the Bethe free energy is lower bounded by the exact free energy. Moreover, many methods that minimize $\mathcal{F}_B$ are particularly efficient for attractive models and thus for balanced models as well. One of the properties that we will use throughout is that for attractive models all fixed points corresponding to local minima of $\mathcal{F}_B$ are stable [50, Theorem 6], [36], [46]. Although BP may still fail to converge on balanced models, it is thus at least guaranteed that BP will converge if the messages are initialized close enough to a fixed point.

For unidirectional models two fixed points exist at most. This not only renders approximating the free energy and the marginals relatively straightforward but also allows for a thorough evaluation of various algorithms. Since we devote a great part of our theoretical analysis of SBP to its performance on unidirectional models, we will now discuss the solution space of unidirectional models in more detail and prepare some foundations for the later analysis.

Most properties result from generalizing the properties of homogeneous models. As discussed in Section 2.2.1 these models qualitatively behave the same.

First, we treat the special case where all local potentials are specified by $\theta_i = 0$ and where both fixed points – if they exist – are symmetric.

**Lemma 1.** Let us consider a unidirectional model with loops, with $\theta_i = 0$, and with couplings strong enough to admit multiple (i.e., two) fixed points. Then, both fixed points $s$ and $t$ are symmetric, i.e., $m_{is} = -m_{is}^*$ and both fixed points have the same Bethe free energy, i.e., $\mathcal{F}_B = \mathcal{F}_B$.

**Proof:** Given that the local potentials are the same for both states, the message update rule from (6) simplifies to $\mu_{ij}^{n+1}(x_j) \propto \sum_{x_i} \Phi(x_i, x_j) \prod_{x_k} \mu_{ik}^{n}(x_i)$. Consequently, if $\mu_{ij}^*(x_j) = \mu_{ij}^*(x_j)$ are fixed point messages, so are the symmetric messages $\mu_{ij}^*(x_j) = 1 - \mu_{ij}^*(x_j)$. By computing the pseudo-marginals according to (8)-(9) it follows immediately that the means are symmetric around 0 and that both fixed points have the same Bethe free energy (see Example 1).

We will now present an illustrative example that describes the behavior of models with $\theta = 0$. These models have been well-studied in the literature [20], [32], [57]. We present Example 1 to summarize the most important properties that serve as the basis for our theoretical analysis of SBP in Section 5.

**Example 1.** Consider a homogeneous model on a complete graph with $N = 4$ variables, positive couplings $J_{ij} = J > 0$, and local potentials specified by $\theta_i = 0$. For computing the Bethe free energy $\mathcal{F}_B$ according to (13), we enforce the symmetry of the model by setting all singleton marginals to the same value $\tilde{P}(x_i) = \tilde{P}$. We compute the Bethe-optimal pairwise marginals according to [59, Lemma 1] (see also [57]). This allows us to compute the pairwise marginals given only the singleton marginals.\(^9\)

The effect of the coupling strength is illustrated in Fig. 2. Note that for $\theta = 0$, $\tilde{P} = (0.5, 0.5)$ is a fixed point for every value of $J$; we will often refer to this fixed point in terms of its means $m_j = 0.10$. Consequently, there is also a corresponding stationary point of $\mathcal{F}_B$. For small values of $J$, $\mathcal{F}_B$ is convex and this fixed point is unique (see Fig. 3a); as $J$ increases, $\mathcal{F}_B$ becomes flat (see 9. Note that an alternative representation of (5) is used in [57], [59]; there is, however, a straightforward mapping from one representation to the other.

10. This fixed point exists whenever all local potentials are specified by $\theta = 0$ and can be computed analytically [32].

Fig. 2. Evolution of the Bethe free energy $\mathcal{F}_B$ for Example 1. We consider a complete graph with $N = 4$ variables with homogeneous potentials, attractive edges, and $\theta_i = 0$. $\mathcal{F}_B$ is evaluated for $P(x_i) = P$ for (a) $J = 0.1$, (b) $J = 0.55$, and (c) $J = 0.7$.

Fig. 3b), before turning into a local maximum with two symmetric minima branching off (see Fig. 3c).

Second, when all local potentials are specified by $\theta_i > 0$ the fixed points are not symmetric anymore. For later reference, we summarize some important properties of the fixed points in the following Lemma.

**Lemma 2.** Let us consider a unidirectional model with $\theta_i > 0$ and couplings strong enough to admit multiple (i.e., two) fixed points. Then, one fixed point has its means aligned with the local potentials, i.e., $m_{is}^* > 0$; we call this fixed point consistent or state-preserving (cf. [12], [22]). Similar, as in Lemma 1, the means of the second fixed point have opposing signs, i.e., $m_{is}^* < 0$. Both fixed points are not symmetric, however, and the consistent fixed point has larger means $|m_{is}^*| > |m_{is}^*|$ and larger correlations $\chi_{is}^* > \chi_{is}^*$. Accordingly, the consistent fixed point constitutes the global minimum of the Bethe free energy, i.e., $\mathcal{F}_B = \mathcal{F}_B$.

**Proof:** First, the existence of a consistent fixed point, i.e., $m_{is}^* > 0$, is an immediate consequence of bounding the location of the fixed points of BP in [54, Theorem 3], [61]. Second, the existence of a second point with opposing means, i.e., $m_{is}^* < 0$, follows from the fact that only a single fixed point exists with positive means [9, Lemma 2.3]; thus the second fixed point, if it exists, must have negative means (see Example 2). Third, it is an immediate consequence of the update rule in (6) that the consistent fixed point is energetically favored, i.e., $\mu_{ij}^*(x_j) > \mu_{ij}^*(x_j)$. Then, computing the marginals according to (8) implies that the consistent fixed point has larger means, i.e., $|m_{is}^*| > |m_{is}^*|$. The same line of reasoning shows that the correlations are larger as well. Finally, the consistent fixed point has a lower average energy, i.e., $E_B(\tilde{P}_B) < E_B(\tilde{P}_B)$ (see (11)). Therefore (and by symmetry of the Bethe entropy), it follows that $\mathcal{F}_B^* = \mathcal{F}_B^*$.\(\square\)

**Example 2.** Consider a unidirectional homogeneous model on a complete graph with $N = 4$, positive couplings $J_{ij} = J > 0$, and local potentials specified by $\theta_i = \theta = 0.02$. Again, we enforce symmetry by setting all singleton marginals to the same value and compute the pairwise marginals and the Bethe free energy accordingly.

Similar as in Example 1, we illustrate the effect of the coupling strength on the shape of $\mathcal{F}_B$. In some sense we see a similar picture emerging; i.e., $\mathcal{F}_B$ is convex for small values of $J$ (see Fig. 3a) and non-convex for larger values of $J$ (see Fig. 3c).\(\square\)

There is one crucial difference however: because of $\theta_i > 0$ the minimum does not turn into a maximum with two minima branching of. Instead, we see that the global minimum persists\(11\). Contrary to Example 1, no closed-form solution exists for computing the fixed points because of the local potentials.
for all values of $J$, whereas the energy landscape becomes flat before two additional stationary points emerge somewhere else (see Fig. 3b).

We will see in Section 5 that unidirectional (non-homogeneous) models behave qualitatively the same; consequently we will make repeated use of the behaviour illustrated in Example 2 for analysing the behaviour of SBP.

Finally, note that every fixed point that corresponds to a local minimum is a stable fixed point of BP for unidirectional models. This is a direct consequence of [50, Theorem 6] and the fact that unidirectional models are attractive per definition. Even more importantly, BP will always converge for unidirectional models: therefore note that BP always converges for unidirectional models with $\theta_i = 0$ [33] and that the existence of local potentials with $\theta_i \neq 0$ can only enhance the convergence properties [21]. Recently, it has also been shown that for unidirectional models BP will converge reasonably fast, if initialized properly [9], [24].

2.5.1 Combination of Fixed Points

If multiple fixed points exist, the performance of BP often varies considerably between different fixed points. Here we briefly introduce the RSB (replica symmetry breaking) assumption that states how all these fixed points can be combined to compute the exact marginals. Therefore, let us consider all local minima of the Bethe free energy $F_B$, the pseudomarginals $\tilde{P}_B$ at the corresponding fixed point, and the associated partition function approximation $Z_B^{m}$.

**Conjecture 1 (RSB Assumption).** Let us consider all $M$ local minima of the Bethe free energy $F_B$, the pseudomarginals $\tilde{P}_B$ at the corresponding fixed point, and the associated partition function approximation $Z_B^{m}$.

**Conjecture 1.** Let us consider all $M$ local minima of the Bethe free energy $F_B$, the pseudomarginals $\tilde{P}_B$ at the corresponding fixed point, and the associated partition function approximation $Z_B^{m}$. Then, the exact singleton marginals $P(x_i)$ are given by a convex combination of all fixed points at minima of $F_B$ according to

$$P(x_i) = \frac{1}{\sum_{m=1}^{M} Z_B^m} \sum_{m=1}^{M} \tilde{P}_B^m(x_i) Z_B^m.$$  

(15)

This representation belongs to a set of postulates in [31]. One underlying assumption is that the system does exhibit multiple fixed points (unique fixed points would falsely imply exact marginals otherwise). Despite its non-rigorous flavor, Conjecture 1 has been verified for a wide range of problems [30, Ch.19]. Example 1 and Fig. 2c in particular give an intuitive illustration of the RSB assumption: for this model the local maximum (i.e., the unstable fixed point) corresponds to the exact marginals; yet, both local minima do not approximate the marginals well. Combining both minima according to (15), however, gives the exact singleton marginals because of the model’s symmetry properties.

Obtaining all minima of the Bethe free energy, however, is a complex task that is only possible for models with certain structures or potentials [6], [20], [36], [64]. Besides the application to optimization problems [4] the RSB assumption is thus still of limited practical relevance for estimating the marginals; nonetheless, it provides a powerful concept for assaying and comparing the marginals given a selection of fixed points [22].

3 Self-guided belief propagation (SBP)

The main aim of SBP is to select an accurate fixed point, i.e., a fixed point that approximates the singleton marginals well, or – if none are stable – to approximate an accurate fixed point. In this section we present an intuitive justification of the proposed method and subsequently introduce SBP in detail. We further present practical considerations and pseudocode of SBP. A formal treatment of SBP is presented in Section 5.

The current understanding of BP is that strong pairwise potentials negatively influence BP in the sense of deteriorating the convergence behavior [20], [32] and the approximation quality of the marginals [20], [22], [57] and that incorporating the potentials slowly [3] may reduce the overall number of iterations. Inspired by the recent observation that strong local potentials increase accuracy and lead to better convergence properties [21], we aim to reduce the influence of the pairwise potentials that negatively influence BP. In doing so, we hope that an accurate fixed point emerges if we start from a simple model with independent random variables and slowly increase the potentials’ strength.

SBP starts from a simple model with independent random variables and slowly incorporates the edge potentials’ strength. More precisely, SBP is a homotopy method that solves the simple problem first and – by repetitive application of BP – keeps track of the fixed point as the interaction strength is increased by a scaling term.

Formally, SBP considers an increasing length-$M$ sequence \( \{\zeta_m\} \) where \( m = 1, \ldots, M \) such that \( \zeta_m < \zeta_{m+1} \) and \( \zeta_m \in [0, 1] \) with \( \zeta_1 = 0 \) and \( \zeta_M = 1 \). This further indexes a sequence of probabilistic graphical models \( \{U_m\} \) that converges to the model of interest \( U_M = U \). Every probabilistic graphical model has a set of potentials \( \Phi_m = \{\Phi_m(x_i, x_j), \Phi_m(x_i)\} \) associated, where \( \Phi_m(x_i) = \Phi(x_i) \) and the pairwise potentials at index \( m \) are exponentially scaled by

$$\Phi_m(x_i, x_j) = \exp(J_{ij} \zeta_m x_i x_j) = \Phi(x_i, x_j)^{\zeta_m}.$$  

(16)

We further write \( \mu_m \) to clarify that we consider a BP fixed point for the model \( U_m \). If multiple fixed points exist, the initialization of BP plays a crucial role in determining to which fixed point BP converges and how it performs. SBP always provides a favorable initialization by the preceding fixed point and aims to perform the composite function

$$\mu_M = BP_M^0 \left( BP_{M-1}^0 \left( \cdots BP_1^0 \left( \mu_1^{init} \right) \right) \right).$$  

(17)

12. The underlying assumption is the existence of a continuous path that connects the initial messages \( \mu_1^{init} \) to the final fixed point \( \mu_M^0 \), where all intermediate initializations \( \mu_m^{init} \) lie along this path. For now, we assume a fixed number of models \( M \) for practical reasons but we will consider SBP’s behavior in the limit of \( M \to \infty \) in our formal analysis in Section 5.
This may lead to problems if the fixed point becomes unstable for some value $m < M$, in which case we cannot rely on BP to keep track of the fixed point anymore. Instead, SBP provides the last stable fixed point in that case, i.e., $\mu_m \circ$ as the final estimate.

**Example 3.** We illustrate how SBP approximates the marginals for a frustrated model, for which BP fails to converge, in Fig. 4. Initially, SBP obtains the pseudomarginals for $\zeta = 0$ by running BP on the simple model. As $\zeta$ increases, SBP estimates the pseudomarginals along the solution path by running BP to keep track. For $\zeta > 0.7$, however, BP does not converge anymore; SBP consequently terminates and provides the marginals of the last stable fixed point as the final approximate solution. Note that the approximated marginals are already close to the exact ones in this example; experiments show that this is often the case (see Section 4).

In other words, SBP relaxes the problem of minimizing $F_B$ by making all variables independent (and the Bethe approximation exact). Then the problem is deformed into the original one by increasing $\zeta$ from zero to one. Thereby, a stationary point $F_B^\circ$ emerges as a well-behaved path (see Property 1 in Section 5) and SBP keeps track of it with BP constantly correcting the stationary point.

### 3.1 Practical considerations

In practice the run-time of SBP is influenced by the difference between two successive fixed points $\mu_m \circ$ and $\mu_{m-1} \circ$ – the difference is primarily determined by the number of steps $M$. Ideally $M$ should be as large as possible. This, however, increases the runtime (see Theorem 6); in practice, we would choose $M$ as small as possible but as large as necessary. Moreover, one can adaptively increase the step size if two successive fixed points are close, i.e., if $\mu_m \circ \approx \mu_{m-1} \circ$ (cf. [43, pp.23], [1]). Our experiments show that it is sufficient to use rather coarse steps; we used $M \leq 10$ for all reported experiments and using more steps would not have improved the accuracy.

Additionally, instead of initializing $BP_m$ with its preceding fixed point messages, i.e.,

$$\mu_m^{\text{init}} = \mu_{m-1} \circ \circ$$

one can (e.g., by spline extrapolation) estimate $\mu_m = f(\mu_{m-1} \circ, \mu_{m-2} \circ, \ldots, \mu_{m-k} \circ)$ so that $\mu_m^{\text{init}} \approx \mu_m \circ$ to reduce the overall number of iterations. We empirically observed that the benefit diminishes for $k > 3$.

### 3.2 Pseudocode

Pseudocode of SBP is presented in Algorithm 1. Note that we initialize the messages $\mu_m \circ$ randomly. The sequence of messages over all models $\{\mathcal{M}_m\}$ is collected in $\{\mu_m \circ\} = \{\mu_1 \circ, \ldots, \mu_m \circ\}$. ExtrapolateMsg applies cubic spline extrapolation to estimate the initial messages of the subsequent model.

We further present the pseudocode for the adaptive step size controller in Algorithm 2. The step size is increased depending on $k$, where $k$ is the number of preceeding fixed point messages that are $\varepsilon$-close to the current fixed point messages $\mu_m \circ$.

#### Algorithm 1: Self-Guided Belief Propagation (SBP)

**input**: graph $\mathcal{G} = (\mathcal{X}, \mathcal{E})$, potentials $\Psi$, initial step-size $\text{step}_{\text{init}}$, maximum number of iterations $N_{BP}$

**output**: fixed point messages $\mu_m \circ$

1. $\mu_1^{\text{init}} \leftarrow \mu_1$
2. $m \leftarrow 1$
3. $\zeta_1 \leftarrow 0$
4. while $\zeta_m \leq 1$
5. \hspace{1em} $\Psi(\zeta_m) \leftarrow \text{ScalePotentials}(\Psi, \zeta_m)$
6. \hspace{1em} $(\mu, \text{iterations}) \leftarrow \text{BP}(\mu_m^{\text{init}}, \Psi(\zeta_m), N_{BP})$
7. \hspace{1em} if $\text{iterations} < N_{BP}$ then
8. \hspace{2em} $\mu_m \circ \leftarrow \mu$
9. \hspace{1em} else
10. \hspace{2em} break
11. \hspace{1em} if $\text{adaptive step size}$ then
12. \hspace{2em} $\zeta_{n+1} \leftarrow \zeta_n + \text{AdaptiveStepSize}(\mu_m \circ, \text{step}_{\text{init}}, m)$
13. \hspace{1em} else
14. \hspace{2em} $\zeta_{n+1} \leftarrow \zeta_n + \text{step}_{\text{init}}$
15. \hspace{2em} $\mu_{m+1}^{\text{init}} \leftarrow \text{ExtrapolateMsg}(\mu_m \circ, \{\zeta_m\})$
16. \hspace{2em} $m \leftarrow m + 1$
17. $\mu_m \circ \leftarrow \mu_{m-1} \circ$

#### Algorithm 2: Adaptive Step Size Controller

**input**: sequence of messages $\{\mu_m \circ\}$, $\text{step}_{\text{init}}, m$

**output**: stepsize $\text{step}$

1. $\text{step} \leftarrow \text{step}_{\text{init}}$
2. $k \leftarrow 1$
3. while $|\mu_m \circ - \mu_{m-k} \circ| < \varepsilon$ do
4. \hspace{1em} $k \leftarrow k + 1$
5. \hspace{1em} $\text{step} \leftarrow \text{step} + \text{step}_{\text{init}} \cdot k$

#### 4 Experiments

We apply SBP to attractive (Section 4.2) and general (Section 4.3) models on $n \times n$ grid graphs of different sizes, on complete graphs with $N = 10$ random variables, and on random graphs (i.e., Erdos Renyi graphs with an average degree $d = 3$) with $N = 10$ random variables. These graphs are considered in order to render the computation of the exact marginals feasible and to make the results comparable to previous work [29], [44], [45], [57].
4.1 Experimental settings

We evaluate SBP and compare it to BP, BP<sub>D</sub> (BP with damping), and Gibbs sampling (run for 10<sup>5</sup> iterations) using the mean squared error (MSE) between the approximate marginals \( \tilde{P}_X \) and the exact marginals \( P_X \), where, for binary random variables, MSE = \( \frac{1}{N} \sum_{i=1}^{N} |P_X_{-}(+1) - P_X_{+}(+1)|^2 \). We further compare the run-time by counting the overall number of BP iterations and the number of iterations for Gibbs sampling.\(^{13}\)

For BP and SBP we set the maximum number of iterations to \( N_{BP} = 10^3 \) and use random scheduling. We randomly initialize \( \mu_{init} \) and use an adaptive step size with an initial step size of \( \epsilon = 10^{-3} \). The overall number of steps is thus \( M \leq 10 \).

For BP<sub>D</sub> we choose a large damping factor \( \epsilon = 0.9 \) in order to prioritize convergence over run-time and therefore increase the maximum number of iterations to \( N_{BP} = 10^4 \). Carefully selecting a damping factor that depends on a given model may reduce the number of iterations until convergence but cannot increase the accuracy; moreover, if chosen too small, BP<sub>D</sub> may fail to converge at all.

The initial messages are randomly initialized 100 times for each model, before applying BP with and without damping. We consider BP (and BP<sub>D</sub>) as converged for a model if at least a single message initialization (out of 100) exists for which BP failed to converge at all.

\(^{13}\) Computing the acceptance-probability requires similar run-time as one BP message update.
once divided by the overall number of experiments (i.e., 100).

The reported error (MSE) and the number of iterations are averaged over all convergent runs of BP and BP₀ (i.e., BP and BP₀) while all runs that did not converge are discarded. On the contrary, we average the error and the number of iterations over all models for SBP (SBP₀), Gibbs sampling (Gibbs₀), and for minimization of the Bethe approximation (Fₐ, all).

### 4.2 Attractive models

We consider grid graphs with N = 10 × 10 random variables, random graphs with N = 10 random variables, and complete graphs with N = 10 random variables. We generate L = 100 models for every value of β ∈ {0, 0.5, ..., 5} and sample the potentials according to θᵢ ∼ U(−0.5, 0.5) and Jᵢ,j ∼ U(0, β); i.e., overall we consider 1100 different models for each individual graph-structure. Note that BP is randomly initialized 100 times for every considered model. We compute the MSE for every value of β and visualize the mean and the standard deviation of the MSE as well as the number of iterations in Fig. 5a.

BP (orange) converges rapidly for all graphs considered; hence, there is no additional benefit in using BP₀ (green) that only increases the number of iterations. SBP (blue) slightly increases the number of iterations as compared to BP but converges in fewer iterations than BP₀. Note that SBP is guaranteed to capture the global optimum when applied to unidirectional models (see Theorem 9-10). But even if we do allow for random local potentials, we empirically observe that SBP consistently outperforms BP with respect to accuracy; specifically for models with strong couplings. These models exhibit multiple stable fixed points [20] such that, depending on the initialization, BP often converges to inaccurate fixed points while SBP selects an accurate one.

### 4.3 General models

General models with random potentials are often frustrated and traditionally pose problems for BP and other methods that aim to minimize the Bethe approximation.

First, in order to evaluate the performance of SBP we consider uniform local potentials θᵢ = θ ∈ {0, 0.1, 0.4} and draw the couplings with equal probability from Jᵢ,j ∈ {−1, 1}; the results are summarized in Tab. 1. Although BP and BP₀ fail to converge for most models we observe that SBP stops after only a few iterations and significantly outperforms BP in terms of accuracy. In fact, SBP achieves accuracy competitive with Gibbs sampling but requires three orders of magnitude fewer iterations.

Second, we apply SBP to general graphs and evaluate whether SBP provides a good approximation of the pseudomarginals P*ₐ that correspond to the global minimum of the Bethe free energy Fₐ. Therefore we consider grid graphs (of size 5 × 5), which still allows us to approximate the global minimum Fₐ – and the related pseudomarginals P*ₐ – reasonably well by [55]. The results are summarized in Tab. 1 and show that SBP approximates P*ₐ within the accuracy of our reference method (MSEₐ). We further report the number of times where SBP obtains the terminal fixed point, i.e., for UM, in Tab. 1 i.e., (Fₐ(ζₐ) equals SBP). It becomes obvious that SBP approximates the terminal fixed point reasonably well, despite frequently stopping for ζₐ < 1. Moreover, the MSE reveals that SBP does not only approximate the pseudomarginals P*ₐ well but concurrently provides an accurate approximation of the exact marginals Pₐ.

Third, we investigate how the approximation quality depends on the scaling term ζₐ. Therefore, we depict the evolution of MSEₐ (to the approximate solution) and of the MSE (to the exact solution) in Fig. 6. We observe that MSEₐ (blue) decreases monotonically with every iteration, which empirically verifies that SBP proceeds along a well-behaved solution path (see Property 1). Note that MSEₐ decreases rapidly in the first iterations and SBP spends a major part of the overall run-time for slight improvements. The MSE to the exact solution, on the other hand, decreases first until it increases again as SBP incorporates stronger couplings. Stronger couplings tend to degrade the quality of the Bethe approximation in loopy graphs and lead to marginals that are increasingly biased towards one state [20], [51]. This explains why the MSE to the exact solution increases as SBP converges towards the terminal fixed point. One could exploit this behavior and restrict the run-

### Table 1

Results for general models with Jᵢ,j ∈ {−1, 1} on grid graphs (N = 25 and N = 100), complete graphs (N = 10), and random graphs (N = 10). We report the MSE to the exact marginals and the MSEₐ to the Bethe approximation, convergence ratio, and the overall number of BP iterations. Only converged runs are considered for BP₀ and BP₀ but all runs are considered for SBP₀, Gibbs₀, and Fₐ, all.
5 THEORETICAL PROPERTIES OF SBP

Here, we present some more formal arguments and discuss the properties of SBP to understand under which conditions the algorithm (presented in Section 3) can be expected to perform well. We begin by formally defining the notion of a solution path in Section 5.1 before we discuss the properties of SBP’s solution path in Section 5.2. Finally, we restrict ourselves to unidirectional models, for which we analyze the accuracy of SBP in Section 5.3.

5.1 Solution Path: Definition

First, we fix our notation: For the model \( U_m \) with its potentials \( \Psi_m \), we refer to the pseudomarginals by \( \tilde{P}_B(\zeta_m) \), and, with slight abuse of notation, we refer to the corresponding stationary point of the Bethe free energy by \( F_B(\zeta_m) = F_B(\tilde{P}_B(\zeta_m)) \). It is beneficial to study the behavior of SBP as \( M \) tends towards infinity. Therefore we consider the unit interval \( \zeta \in [0, 1] \) to be the compact support of the functions \( F_B(\zeta) \) and \( \tilde{P}_B(\zeta) \). SBP is inspired by the idea to proceed along a so-called solution path as \( \zeta \) increases from zero to one in order to obtain the marginal distributions for the model of interest. Therefore, we consider a continuous homotopy function \( H(\mu, \zeta) : \mathbb{R}^{|\mu| + 1} \rightarrow \mathbb{R}^{|\mu|} \). Note that the scaling factor modifies the graphical model \( U = (G, \Psi(\zeta)) \), for which the pairwise potentials are specified according to (16); we define the homotopy accordingly by

\[
H(\mu, \zeta) = \mu - BP(\mu) \quad \text{where} \quad \Psi = \Psi(\zeta).
\]

Consequently, for a given value \( \zeta \), \( H(\mu, \zeta) \) is zero precisely for the fixed points \( \mu^0 \) of \( U = (G, \Psi(\zeta)) \). It follows that a solution path

\[
c(\zeta) : H(\mu, \zeta) = 0
\]

exists that (i) has a start point \( c(\zeta = 0) = \mu : H(\mu, \zeta = 0) = 0 \), (ii) an endpoint \( c(\zeta = 1) = \mu : H(\mu, \zeta = 1) = 0 \) , and (iii) is continuous over \( \zeta \in [0, 1] \), i.e., the solution path connects the start- with the endpoint.

SBP then proceeds along a solution path \( c(\zeta) \) that implicitly defines the pseudomarginals \( \tilde{P}_B(\zeta) \) by (8)-(9). In particular, we refer to the start- and endpoint by \( \tilde{P}_B(\zeta = 0) \) and \( \tilde{P}_B(\zeta = 1) \) respectively.

The following example illustrates the solution path for a unidirectional model.

Example 4. Consider a unidirectional grid graph of size \( 3 \times 3 \) with \( J = 2 \) and \( \theta = 0.1 \). For this model, we can compute the fixed points of BP; i.e., the solutions to \( H(\mu, \zeta) = 0 \) with the numerical homotopy continuation method as in [20].

Fig. 7 shows how the solutions evolve as \( \zeta \) goes from zero to one. Similar to Example 2, this illustrates how increasing the interactions between variables changes the energy landscape. That is, for small values of \( \zeta \) we have a unique fixed point until – as \( \zeta \) increases – two additional fixed points emerge; one of which is a minimum and one a maximum of \( F_B \). Although multiple solutions to \( H(\mu, \zeta) = 0 \) exist for larger values of \( \zeta \), all except for one lack a start point and are therefore of no relevance for any method that proceeds along a solution path defined by the homotopy in (18).

5.2 Solution Path: Properties

The following proposition summarizes the main properties of the solution path that is specified and followed by SBP.

Properties 1 (Convergence Properties). SBP proceeds along a well-defined solution path \( c(\zeta) \). More precisely, we have:
BP has a unique fixed point \( \mu^\zeta_i \) for \( \zeta = 0 \), so that SBP has a unique start point \( \tilde{P}_B^0(\zeta = 0) \) (see Theorem 3).

A smooth (i.e., continuous) solution path originates from \( \tilde{P}_B^0(\zeta = 0) \) (see Theorem 4).

SBP proceeds along this (unique) solution path \( c(\zeta) \) and terminates after a fixed number of iterations (see Theorem 6).

In the sequel, we validate all points of Property 1.

### 5.2.1 Uniqueness of the Start Point

First, we show that the initial model with \( \zeta = 0 \) has a unique fixed point. This is an immediate (and known) consequence of the fact that all variables are independent; we still include the following Theorem for the sake of completeness.

**Theorem 3 (Property 1.1).** For the initial model \( \mathcal{U}_I \) with parameters \( \Psi(\zeta = 0) \), BP has a unique fixed point \( \mu^0_i \). That is, only one single start point exists for the solution path of SBP. Moreover, this start point is exact, i.e., \( \tilde{P}_B^0(\zeta = 0) = P_B(\zeta = 0) \), and BP is guaranteed to converge.15

**Proof:** First, we show that all messages converge to the identical value whenever all \( J_{ij} = 0 \). Therefore, note that we can omit the pairwise potential in the update equation of BP and rewrite (6) to

\[
\mu^{n+1}_{ij}(x_j) \propto \sum_{x_i \in X^i} \Phi(x_i) \prod_{k \in N(i) \backslash j} \mu^n_{kj}(x_k).
\]

It follows that (20) is independent of the state \( x_j \), i.e., the computed messages are the same for all states of \( X_j \). Thus, after normalization, \( \mu^0_j(x_j) = \frac{1}{\tilde{Z}_i} = 0.5 \) for all edges \( (i, j) \in E \).

Second, the pseudomarginals are then obtained according to

\[
\tilde{P}(x_i) = \frac{1}{\tilde{Z}_i} \Phi(x_i) \prod_{k \in N(i)} \mu^0_k(x_k) = e^{x_i \cdot \theta_i} / (e^{\theta_i} + e^{-\theta_i}).
\]

Finally, we must show that the exact marginals are the same, i.e., that \( \tilde{P}(x_i) = \tilde{P}_B(x_i) \). Therefore, note that the joint in (5) factorizes only over the local potentials, which allows us to reorder the summation and compute the exact marginals according to

\[
P(x_i) = \frac{1}{Z} e^{x_i \cdot \theta_i} \sum_{x_j \in (X \backslash X_i)} \prod_{j \in (X \backslash X_i)} e^{x_j \cdot \theta_j} = e^{x_i \cdot \theta_i} / (e^{\theta_i} + e^{-\theta_i}).
\]

Theorem 3 thus reduces the problem of initializing SBP to computing the fixed point messages \( \mu^0_i \) of the initial model with \( \zeta = 0 \); this can be done in linear time. Moreover, a unique start point implies that only a single solution path emerges from it.

### 5.2.2 Smooth Solution Path

Second, we show that – for any model – the Bethe free energy is a smooth function of the scaling parameter \( \zeta \) in the sense that no discontinuities exist. This implies the existence of a smooth solution path; let us make this notion of a smooth solution path more precise:

\[15. \text{Note that this concurs with the sandwich-bound [54, Th.4] that reduces to} \ P_{X_i}(+1) = \theta_i / (e^{\theta_i} + e^{-\theta_i}) = \tilde{P}_{X_i}(+1) \text{for} \ J_{ij} = 0.\]

**Theorem 4 (Property 1.2).** Consider the pseudomarginals \( \tilde{P}_B^0(\zeta) \) along the (one) solution path that originates from the start point \( \tilde{P}_B^0(\zeta = 0) \). Then, both the pseudomarginals \( \tilde{P}_B^0(\zeta) \) and the associated stationary point \( F_B^0(\zeta) \) are continuous functions of the scaling parameter \( \zeta \) along the whole solution path, i.e., for \( \zeta \in [0, 1] \).

**Proof:** First, we show that the Bethe free energy \( F_B(\tilde{P}_B^0(\zeta)) \) is continuously differentiable. Therefore, the derivative with respect to \( \zeta \) must exist and be continuous as well. Note that \( F_B(\tilde{P}_B^0(\zeta)) \) is a high-dimensional function defined over \( \tilde{P}_B \in \mathcal{L} \), i.e., over all locally consistent marginals. Now, let us consider (13) with the pairwise potentials defined by (16). Then, when taking the derivative with respect to \( \zeta \), all terms depending only on singleton marginals vanish and the derivative is given according to

\[
\frac{\partial F_B(\zeta)}{\partial \zeta} = \sum_{(i, j) \in E} \sum_{x_i, x_j} \tilde{P}(x_i, x_j) \ln \Phi(x_i, x_j)
\]

Since (24) is a finite sum over finite terms, this proves that \( F_B(\zeta) \) is continuously differentiable.16 More precisely, (24) proves that \( F_B(\tilde{P}_B^0, \zeta) \) is continuously differentiable with respect to \( \zeta \) for every point on the energy surface.

We specifically consider the local minimum \( F_B^m(\zeta) \) along the solution path, i.e., the minimum that emerges from the unique start point \( F_B^m(\zeta = 0) \).

Given that the Bethe free energy varies in a continuous fashion along this solution path (for \( \zeta \in [0, 1] \)), the local minimum \( F_B^m(\zeta) \) and its position vary in a continuous fashion as well [8]. For existence of this fixed point, note that minima of \( F_B \) persist when increasing \( \zeta \) (similar to increasing the inverse temperature; see [63]). Consequently, because of the one-to-one correspondence between stationary points of the Bethe free energy \( F_B^m(\zeta) \) and fixed points of BP \( P_B^m(\zeta) \), the pseudomarginals are also continuous. Note, however, despite each minimum varying continuously with increasing \( \zeta \), the global minimum may not (as a former local minimum becomes the global minimum).

In Theorem 4, we have shown that continuity of the Bethe free energy \( F_B(\zeta) \) and the corresponding pseudomarginals \( P_B^0(\zeta) \) implies a smooth solution path. We implicitly assumed that this further implies a well-behaved solution path. In general this is the case but there are two scenarios that might be problematic as we will discuss now:

First, it is absolutely critical that the stationary points of \( F_B \) are zero-dimensional, i.e., that every minimum is a distinct point, and that there are finitely many stationary points. Fortunately, this

\[16. \text{Moreover – according to its definition in (13) –} F_B(\tilde{P}_B^0(\zeta)) \text{is a polynomial, which guarantees that it is not just continuously differentiable but in fact an analytical function [39]. Note that this is in accordance with the fact that true phase transitions (singularities in the derivative of the free energy) can only occur in the thermodynamic limit, where (24) becomes a sum over infinitely many terms that equates to infinity.}\]

\[17. \text{This is actually not restricted to problems with a unique minimum; instead sets of minima are considered in [8] to prove that minima of constrained minimization problems vary continuously under continuous changes of the objective function (see also [38, Section 5]).}\]
is the case and the number of stationary points of $F_B$ is always finite [50]; the same is true for the fixed points of BP (for as long as the messages are properly normalized) [27].

Second, a potential issue is how SBP copes with bifurcations of the solution path. A bifurcation is a point along the solution path at which a minimum $F_B^0$ turns into a maximum with two new minima branching off; i.e., one minimum splitting up into two new minima (see Example 1 for the illustration of a typical bifurcation). In that case, SBP will follow the evolution of a single minimum and discard the other one. Which fixed point is selected mainly depends on the implementation details of BP (this is because we rely on BP to find the next fixed point after our prediction). Note that different minima are usually wide apart (see Example 1 or [22]). If, however, two minima would be close together, SBP might switch between branches because of numerical issues (e.g., because of quantization errors or a step-size too large). Since this is not a fundamental property of the solution path, we will assume sufficient numerical precision and rule out this phenomenon. Thus, the existence of bifurcations is no particular problem and SBP retains the property of a unique solution path.

Although bifurcations are not problematic from a practical perspective, it is still relevant for our theoretical analysis of SBP to understand if and where bifurcations of the solution path exist. Unfortunately, the current state of knowledge is still relatively limited in this respect. Note, however, that at least for unidirectional models it is actually possible to make precise statements regarding the existence of bifurcations.

**Lemma 5.** Let $G$ be a unidirectional model. Then, if the local potentials of all nodes are specified by $\theta_i = 0$, a bifurcation occurs. Else, if some local potentials are nonzero, no bifurcation exists along the solution path.

**Proof:** Since we consider unidirectional models there are two minima of $F_B$ at most. We denote these minima by $F_B^0$ and $F_B^1$. We will explain now, why these two minima branch off the previously unique minimum as we increase the scaling parameter $\zeta$ if and only if $\theta_i = 0$. Therefore, we denote the critical scaling parameter, for which the two minima begin to exist, by $\zeta_c$. Then, by definition, a bifurcation exists if and only if both minima coincide for $\zeta = \zeta_c$, i.e., if

$$\lim_{\zeta \rightarrow \zeta_c^-} m_i^0 = \lim_{\zeta \rightarrow \zeta_c^+} m_i^0.$$  (25)

First, remember that both fixed points are symmetric if all $\theta_i = 0$, i.e., $m_i^0 = -m_i^0$ (see Example 1 and Fig. 2). Because of this symmetry, both fixed points will bifurcate from a fixed point with zero mean $m_i = 0$. Note, that this specific case is well understood and serves as the archetype for the existence of bifurcations on the models we consider in this work [33].

Second, it remains to show that this is not the case if $\theta_i \neq 0$. Therefore, we show that both minima do not coincide for $\zeta = \zeta_c$, i.e.,

$$\lim_{\zeta \rightarrow \zeta_c^-} m_i^0 \neq \lim_{\zeta \rightarrow \zeta_c^+} m_i^0.$$  (26)

This is an immediate consequence of the fact that both fixed points are not symmetric anymore (see Section 2.5). In fact, $m_i^0 > 0$ is strictly positive, whereas $m_i^1 < 0$ is strictly negative; as a consequence, both fixed points can never coincide, which rules out the existence of bifurcations along the solution path.

The implications of Lemma 5 are illustrated in Example 1 and 2. Note how even small values for $\theta_i \neq 0$ break the symmetry between both fixed points and consequently imply the absence of bifurcations.

To summarize, Theorem 4 and Lemma 5 guarantee the existence of a smooth solution path that connects the start point $P_B^0(\zeta = 0)$ to the endpoint $P_B^1(\zeta = 1)$. This is a property of fundamental importance as, at least in principle, it allows path-tracking algorithms (as SBP) to proceed along this well-behaved solution path. The only remaining question is whether SBP is actually capable of proceeding along the solution path and if it obtains the endpoint; i.e., if SBP converges.

5.2.3 Convergence Properties of SBP

So far, we have substantiated the claim that a smooth solution path emerges from the trivial start system leading to the target system. The existence of this smooth solution path alone, however, is not sufficient to guarantee that SBP obtains the marginals of the target system. As already discussed in Section 3, SBP relies on BP to proceed along the solution path; ideally, keeping the overall amount of BP iterations low. If – at some point (for $\zeta < 1$) – the fixed point $P_B^0(\zeta)$ becomes unstable, this approach will inevitably break down. Nonetheless, SBP will proceed along the solution path for as long as it has a stable fixed point and approximates the marginals precisely at the onset of instability. That is, even if the terminal fixed point $P_B^1(\zeta = 1)$ is unstable, SBP will always converge and approximate the marginals in a limited number of iterations.\(^{18}\)

**Theorem 6** (Property 1.3). There exists some $\zeta \leq 1$ so that SBP converges to $P_B^1(\zeta) \in \mathcal{L}$ in $O(M N_{BP})$.

**Proof:** SBP increases $\zeta$ as long as BP converges in less than $N_{BP}$ iterations, and stops otherwise. Consequently, BP corrects the accuracy of the fixed point for each value $\zeta_m$ within a bounded number of iterations. The run-time of SBP is further determined by the choice of $M$, i.e., the step-size (see Section 3.1). That is, BP is run until convergence for at most $M$ times and thus, SBP converges in at most $M \cdot N_{BP}$ iterations.

SBP is consequently capable of tracking the fixed point that emerges as $\zeta$ increases and requires $M N_{BP}$ iterations at most. SBP may, however, only converge to a surrogate model for $\zeta_m < 1$ and is not guaranteed to obtain the pseudomarginals of the endpoint. One can characterize this error by computing a bound on $|F_B^0(\zeta_m) - F_B^1(\zeta = 1)|$ given the difference between $\Psi(\zeta_m)$ and $\Psi(\zeta = 1)$ (cf. [16, Theorem 16]).

SBP works particularly well for attractive models, for which every minimum of $F_B$ is also a stable fixed point of BP [50]. More precisely, if for some scaling term $\zeta_m$ the fixed point of the solution path becomes unstable, i.e., if it turns into a local maximum, two stable fixed points will branch of [50, Theorem 6]. Thus for attractive models (and also for balanced models) a smooth solution path with a stable fixed point always exists so that SBP is consequently capable of tracking the solution path until its end.

**Corollary 6.1.** Let $G$ be an attractive or unidirectional model. Then, every point along the solution path going from $\zeta = 0$ to $\zeta = 1$ is stable. Consequently, some number of steps $M$ exists for which SBP converges and obtains the marginals $P_B^0(\zeta = 1)$.

**Proof.** According to Theorem 4, a smooth solution path $c(\zeta)$ exists that goes from the start point $P_B^0(\zeta = 0)$ to the endpoint $P_B^1(\zeta = 1)$.
\( \hat{P}_B(\zeta = 1) \), where every point along \( c(\zeta) \) corresponds to a local minimum of the Bethe free energy \( F_B \). Every local minimum is also stable for attractive models [50]. Therefore, the correction by BP will converge in every step if initialized sufficiently close to the fixed point, i.e., if the step size is chosen small enough.

### 5.3 Accuracy of SBP

The convergence properties of SBP (Property 1) are of fundamental importance for understanding SBP. Yet, so far we have only been concerned if SBP converges but neglected any discussion regarding the quality of the obtained solution. Thus, we will now theoretically analyze how well the fixed point obtained by SBP approximates the marginals. Unfortunately, however, our understanding of what characterizes a fixed point that approximates the marginals well is relatively poor. Indeed, it is not even obvious how the accuracy of the approximated marginals relates to the accuracy of the approximated free energy [22].

Therefore, we will mainly focus our discussion on the simplest model class for which the solution space is well understood (i.e., unidirectional models). Note that for unidirectional models SBP is actually guaranteed to converge (see Section 2.5). Regarding the accuracy of SBP – when applied to unidirectional models – we summarize the properties as follows:

**Properties 2 (Accuracy of SBP for Unidirectional Models).** Let \( G \) be a unidirectional model, i.e., all edges are attractive (\( J_{ij} > 0 \)) and all local potentials are specified by \( \theta_i \) with the same sign. Without loss of generality we consider the case where all \( \theta_i \geq 0 \).

Then, the solution path ends at a fixed point that approximates the marginals and the free energy well. More specifically, SBP either obtains the exact marginals (if all \( \theta_i = 0 \)) or finds the global minimum of the Bethe free energy \( F_B \) that – for unidirectional models – corresponds to the fixed point with the most accurate marginals. Let us now break down the argument into the following three points:

1. For \( \theta_i = 0 \), SBP obtains the exact marginals, i.e., \( \hat{P}_B = P_B \) (see Theorem 7).
2. For \( \theta_i \neq 0 \), SBP finds the global minimum of \( F_B(\zeta) \) and gives the best approximation to the free energy \( F \) (see Theorem 9).
3. For \( \theta_i \neq 0 \), the global minimum of \( F_B(\zeta) \) also gives the best approximation of the marginals, i.e., no other minimum of \( F_B(\zeta) \) has more accurate marginals (see Theorem 10).

First, we show that SBP obtains the exact marginals in the special case where all local potentials are specified by \( \theta_i = 0 \).

**Theorem 7 (Property 2.1).** Consider a unidirectional model with \( \theta_i = 0 \). Then, the pseudomarginals obtained by SBP \( \hat{P}_B(\zeta = 1) \) are identical to the exact ones \( P_B(\zeta = 1) \).

**Proof:** For attractive models with \( \theta_i = 0 \) it is straightforward to compute the exact marginals. The exact marginals of all variables are uniformly distributed and have zero mean \( m_i(\zeta = 0) = 0 \) for all values of \( \zeta \); note that these marginals also correspond to a stationary point \( F_B(\zeta) \) [33]. For sufficiently small values of \( J_{ij} \) this stationary point is a (unique) minimum but for large values of \( J_{ij} \) it turns into a maximum (see [30, pp.385] and Example 1).

19. Note that identical results can be obtained in a straightforward manner for \( \theta_i < 0 \) because of symmetry properties.

Even though the fixed point with \( m_i(\zeta = 0) = 0 \) is unstable, SBP will still converge to this fixed point. This is rather surprising, as – in general – BP can neither converge to an unstable fixed point nor would it remain there because of quantization errors. There are two reasons why SBP obtains the potentially unstable fixed point in the special case of models with \( \theta_i = 0 \) nonetheless. First, note that the (unique) fixed point of the initial mode is identical to the exact solution; hence all variables have zero mean at the start point, i.e., \( m_i^*(\zeta = 0) = 0 \). Second, all messages \( \mu_{ij}(x_j) \in \mu_i^*(\zeta = 0) \) take the same value of \( \mu_{ij}^*(x_j) = 1/2 \), since \( \theta_i = 0 \). The extrapolated messages will stay the same for every iteration of SBP because – as discussed above – they already constitute a valid (albeit possibly unstable) fixed point (see Example 1). Note, however, that these fixed point messages can be represented without quantization error in binary arithmetic. Thus SBP will remain on the fixed point corresponding to the exact marginals.

The above statement arguably holds only for very restricted models and is thus of limited practical relevance. Indeed, not only does standard BP obtain the same fixed point if initialized with identical messages but the exact marginals are also known and can be computed analytically if all \( \theta_i = 0 \), obviating the need for approximate inference methods.

Our next results will discuss the accuracy of SBP for the more general class of unidirectional models both with respect to approximating the marginals and the Free energy (or partition function respectively).

Therefore, we first derive how the scaling term \( \zeta \) affects the marginals along the solution path \( c(\zeta) \) by generalizing Griffiths’ inequality [13] to the fixed points of BP. In particular, this analysis reveals that the means \( m_i^* \) are monotonically increasing with \( \zeta \).

**Lemma 8 (Monotonicity of the Means).** Consider two attractive models \( U_m \) and \( U_n \) on the same graph \( G \), associated scaling parameters \( \zeta_m < \zeta_n \), and local potentials specified by \( \theta_i > 0 \). Note that we are effectively considering two nearly equivalent models with \( U_n \) having strictly larger couplings than \( U_m \). Now, let us consider a BP fixed point of \( U_m \) with positive means \( m_i^*(\zeta_m) > 0 \). Then, the corresponding fixed point of \( U_n \) has larger means \( m_i^*(\zeta_n) > m_i^*(\zeta_m) \) and correlations \( \gamma_{ij}^*(\zeta_n) \geq \gamma_{ij}^*(\zeta_m) \).

**Proof:** The main ingredient of this proof is to show how the ratio between the messages for their respective states \( \bar{\mu}_{ij} = \frac{\mu_{ij}(X_j = +1)}{\mu_{ij}(X_j = -1)} \) increases monotonically with \( \zeta \). Then, the fact that the means and the correlations increase monotonically follows directly from the definition of the pseudomarginals in terms of the messages (see (8) and (9)).

First, we show that the message ratio \( \bar{\mu}_{ij} \) increases monotonically: therefore, let us denote the messages on both models by \( \mu_{(m)ij}(x_j) \) and \( \mu_{(n)ij}(x_j) \) respectively. Furthermore, without loss of generality, we assume that all couplings of \( U_n \) are \( \varepsilon \)-larger than those of \( U_m \). Then, for every \( \varepsilon \) there exists \( \delta > 0 \) so that whenever \( 0 < \bar{\mu}_{(m)ij} - \bar{\mu}_{(m)ij} < \delta \), we have \( 0 < J_{(m)ij} - J_{(m)ij} < \varepsilon \). Note that by assumption \( m_i \in (0, 1] \) so that

\[
\mu_{ij}(X_j = +1) \geq \mu_{ij}(X_j = -1). \tag{27}
\]

First, we show that for all \((i, j) \in E\)

\[
\frac{\mu_{(m)ij}(X_j = +1)}{\mu_{(m)ij}(X_j = -1)} < \frac{\mu_{(n)ij}(X_j = +1)}{\mu_{(n)ij}(X_j = -1)}. \tag{28}
\]

Therefore, consider the update rules of (6) for both states
\[ \mu_{(n)ij}^{n+1}(X_j = 1) \propto e^{J_{(m)ij} + \theta_{(m)ij} + \epsilon} \prod_{X_k \in N(i) \setminus j} \mu_{(n)ki}^n(X_i = +1) \\
+ e^{-J_{(m)ij} - \theta_{(m)ij} - \epsilon} \prod_{X_k \in N(i) \setminus j} \mu_{(n)ki}^n(X_i = -1), \quad (29) \]

\[ \mu_{(n)ij}^{n+1}(X_j = -1) \propto e^{-J_{(m)ij} + \theta_{(m)ij} - \epsilon} \prod_{X_k \in N(i) \setminus j} \mu_{(n)ki}^n(X_i = +1) \\
+ e^{J_{(m)ij} - \theta_{(m)ij} + \epsilon} \prod_{X_k \in N(i) \setminus j} \mu_{(n)ki}^n(X_i = -1). \quad (30) \]

In (29) the larger product is multiplied by \( e^\epsilon \) and the smaller product is divided by \( e^{-\epsilon} \). For (30) it is exactly the other way round so that the ratio between the messages increases which proofs (28).

We shall denote the imposed difference \( \delta \in \mathbb{R}_+ \) on the messages by

\[ \mu_{(n)ij}^\circ(X_j = +1) = \mu_{(m)ij}^\circ(X_j = +1) + \delta, \]

\[ \mu_{(n)ij}^\circ(X_j = -1) = \mu_{(m)ij}^\circ(X_j = -1) - \delta. \]

Second, we show that \( m_\epsilon^0(\chi_m) > m_\epsilon^0(\chi_m) \), which is an immediate consequence of plugging (31) and (32) into (8) (see Example 1 that illustrates how all fixed points, i.e., minima of the Bethe free energy do not necessarily extend to the accuracy of the partition function best).

Finally, it remains to show that \( 0 \leq \chi_{ij}(\chi_m) \leq \chi_{ij}(\chi_n) \) without loss of generality we assume that all variables have equal degree \( d \) and constant local fields \( \theta_i = \theta \). First we show that (i) holds, i.e., \( \chi = \chi_{ij,0} \) is positive. Let us express the marginals by (9) and denote the messages by \( \mu = \mu_{(m)ij}(X_j = 1) \). It follows that \( \mu_{(m)ij}(X_j = -1) = (1 - \mu) \) and that

\[ \chi = e^{J + 2\theta} \mu^{2d} + e^{J - 2\theta} (1 - \mu)^{2d} - 2e^{-J} \mu^d (1 - \mu)^d. \]

Let us further represent the messages by \( \mu = 1/2 + \beta \) with \( \beta \in [0, 1/2] \). It follows that

\[ \chi \geq \left( \frac{1}{2} + \beta \right)^{2d} + \left( \frac{1}{2} - \beta \right)^{2d} - 2 \left( \frac{1}{2} + \beta \right)^d \left( \frac{1}{2} - \beta \right)^d \]

\[ = \left( \frac{1}{2} - \beta \right)^d - \left( \frac{1}{2} + \beta \right)^d \]

\[ \geq 0, \]

(35)

where (a) follows from neglecting all exponential terms and thus upper bounding the positive term and lower bounding the negative term (with equality if and only if \( J = 0 \) and \( \theta = 0 \)) and (b) is a direct consequence of the square in (34). Now let us show that (ii) holds, i.e., \( \chi \) increases monotonically, by taking the derivative of (33), so that

\[ \frac{\partial}{\partial \mu} \chi = 2d \left( e^{J + 2\theta} \mu^{2d - 1} - e^{J - 2\theta} (1 - \mu)^{2d - 1} \right) \\
+ 2d \epsilon e^{-J} (\mu^d (1 - \mu)^{d - 1} - \mu^{d - 1} (1 - \mu)^d) \]

\[ \geq 2d \epsilon e^{-J} (\mu^d (1 - \mu)^{d - 1} - \mu^{d - 1} (1 - \mu)^d) \]

\[ \geq 0, \]

(38)

where (a) follows from neglecting the, strictly positive, first term in (36), and (b) is a direct consequence from (27). \( \square \)

**Theorem 9 (Property 2).** Consider a unidirectional model, i.e., an attractive model with \( \theta_i > 0 \). Then the minimum of the Bethe free energy along the solution path \( F_B^* (\zeta) \) is always the global minimum, i.e., \( F_B^* (\zeta) = \min F_B (\zeta) \). Moreover, the Bethe free energy of this stationary point decreases monotonically with \( \zeta \).

**Proof:** A unique start point \( \tilde{F}_B^* (\zeta = 0) \) exists by Theorem 3. For \( \theta_i > 0 \) this fixed point has only positive means \( m_i^s (\zeta = 0) > 0 \) and correlations \( \chi_{ij}^s (\zeta = 0) > 0 \).

Consequently, let us show that \( F_B^* (\zeta = 1) \) correspond to the endpoint of the solution path \( c(\zeta) \) that emerges from the origin; then, it immediately follows that the error with respect to the endpoint \( F_B^* (\zeta = 1) \) decreases along the solution path: i.e., consider two arbitrary values \( m, n \in [0, 1] \) such that \( n > m \), then \( |F_B^* (\zeta = n) - F_B^* (\zeta = 1)| \geq |F_B^* (\zeta = m) - F_B^* (\zeta = 1)| \).

It remains to show that SBP obtains the fixed point \( F_B^* (\zeta) \) that is the global minimum of the Bethe free energy for a given scaling parameter \( \zeta \). Therefore, we must consider the fact that unidirectional models have two fixed points at most, one of which has positive means (see Lemma 2). As discussed above, SBP proceeds along a solution path belonging to the consistent fixed point with \( m_i^s (\zeta) > 0 \). The second fixed point \( t \), if it exists, has negative means \( m_i^t (\zeta) < 0 \). Given that SBP obtains the consistent fixed point, it follows from Lemma 2 that SBP obtains the global minimum, i.e., \( F_B^* \geq F_B^* = F_B^* \).

**Theorem 9** makes it explicit that – for unidirectional models – SBP favors the fixed point that defines the global minimum of the Bethe free energy. This is in accordance with the observation that the average energy is linear in the pseudomarginals (see (11)), which further implies that varying the local potentials essentially tilts the energy landscape [cf. (37)]. Note that this behavior is also illustrated in Example 1 and 2 accordingly: i.e., as opposed to Fig. 2, local potentials with \( \theta_i > 0 \) tilt the energy landscape in Fig. 3 so that the consistent fixed point constitutes the global minimum of the Bethe free energy.

Although finding the global minimum of the Bethe free energy is desirable, one is ultimately interested in approximating the exact free energy as good as possible. Fortunately, both properties coincide for unidirectional models.

**Corollary 9.1 (Accuracy of the Bethe Approximation).** Consider a unidirectional model. Then there is no other fixed point of BP \( \tilde{P}_B^k \) that approximates the free energy \( F^* \) better than the one obtained by SBP, i.e.,

\[ \tilde{P}_B^k (\zeta) = \arg \min \{ F_B^* (\zeta) - F^* \}. \]

**Proof:** This is a direct consequence of SBP obtaining the global minimum of the Bethe free energy (see Theorem 9) and the fact that the Bethe free energy upper bounds the free energy for attractive models [cf. (40)].

This implies that the fixed point obtained by SBP minimizes the approximation error of the free energy \( |F_B^* (\zeta) - F^*| \). Moreover – because of the direct relationship between the Bethe free energy and the Bethe approximation of the partition function – this implies that the fixed point obtained by SBP also approximates the partition function best.

Note, however, that statements about the approximation quality of the free energy do not necessarily extend to the accuracy of the Bethe free energy.
the pseudomarginals [22]. Nonetheless, at least for unidirectional models, this is the case and SBP obtains the fixed point with the most accurate marginals.

**Theorem 10 (Accuracy of Marginals).** Consider a unidirectional model and assume that the RSB assumption holds (see Conjecture 1). Then there is no other fixed point of BP $\hat{P}_B$ that approximates the marginals better than the one obtained by SBP, i.e.,

$$\hat{P}_B(\zeta) = \arg\min_{P_B(\zeta) \in \mathcal{L}} |\hat{P}_B(\zeta) - P_B(\zeta)|.$$

*Proof:* Note that we are only considering unidirectional models here. For these models, although not formally verified, the RSB assumption is accepted to hold. Thus, if multiple fixed points do exist this allows us to express the exact marginals as a convex combination of all fixed points according to Conjecture 1.

Moreover, remember that the Bethe free energy of unidirectional models has two minima at most (see Section 2.5). This – and the fact that the convex combination of both fixed points yields the exact marginals by assumption – implies that the fixed point obtained by SBP (i.e., the one that minimizes the Bethe free energy) approximates the marginals best.

To conclude, the fixed point along the solution path is always stable for unidirectional models. Thus, SBP follows the solution path to its endpoint. In doing so, SBP not only decreases the Bethe free energy and obtains the global minimum but it also provides the fixed point with the most accurate marginals.

## 6 Conclusion

In this paper, we introduced an iterative algorithm to perform approximate inference: self-guided belief propagation (SBP) is a simple and robust method that gradually accounts for the pairwise potentials and guides itself towards a unique, stable, and accurate solution.

We provide a comprehensive theoretical analysis in order to validate the underlying assumptions: in particular, we showed that:

(i) a smooth solution path exists and originates from a unique start point that is obtained by neglecting the pairwise potentials; (ii) this solution path is well-behaved and can be tracked by SBP; and (iii) for unidirectional models, the solution of SBP approximates the exact solution well and corresponds to the global optimum of the Bethe approximation.

The theoretical analysis does not fully answer for which models SBP is expected to perform particularly well. Although SBP obtains the most accurate fixed point for unidirectional models, the marginals will not be approximated well for models with strong couplings, for the simple reason that accurate fixed points do not exist at all. We expect that SBP is particularly advantageous for all models that have a large amount of fixed points with varying accuracy. Thus SBP will perform well for dense models with random couplings (particularly if they are centered around zero) for which the marginals are largely affected by the local potentials. Not only do we expect that SBP approximates the marginals well in that case but it is actually essential for estimating marginals as BP will often fail to converge.

In order to verify SBP empirically, we applied SBP to a range of models, not only attractive but also general ones. In our experiments, SBP consistently improves the accuracy in comparison with BP (with and without damping). Besides, SBP approximates the exact marginals well on graphical models for which BP does not converge at all.

**Acknowledgment**

The authors greatly appreciate the valuable discussions with Bernhard C. Geiger, KNOW-Center, and acknowledge his help in making the formal definition of SBP much more concise. Moreover, we thank Alexander Ihler, University of California Irvine, for the encouraging discussion during UAI’17 that was pivotal for actually pursuing the – at this time – rather unpolished idea. We further appreciate the help of Florian Kulmer in implementing and empirically evaluating our algorithm and are particularly indebted to Harald Leisenberger for carefully reading the manuscript.

This work was supported by the Graz University of Technology LEAD project “Dependable Internet of Things in Adverse Environments”. Adrian Weller acknowledges support from the David MacKay Newton research fellowship at Darwin College, The Alan Turing Institute under EPSRC grant EP/N510129/1 & U/B/000074, and the Leverhulme Trust via the Leverhulme Centre for the Future of Intelligence (CFI).

**References**

[1] E. Allgower and K. Georg. *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics, 2003.

[2] D. Belanger and A. Young. The random field Ising model. *Journal of magnetism and magnetic materials*, 100(1-3):272–291, 1991.

[3] A. Braunstein, F. Kayhan, G. Montorsi, and R. Zecchina. Encoding for the Blackwell channel with reinforced belief propagation. In *IEEE International Symposium on Information Theory*, 2007.

[4] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

[5] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin. Counting independent sets using the Bethe approximation. *SIAM Journal on Discrete Mathematics*, 25(2):1012–1034, 2011.

[6] A. Coja-Oghlan and W. Perkins. Bethe states of random factor graphs. *Communications in Mathematical Physics*, 366(1):173–201, 2019.

[7] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.

[8] G. B. Dantzig, J. Folkman, and N. Shapiro. On the continuity of the minimum set of a continuous function. *Journal of Mathematical Analysis and Applications*, 17:519–548, 1967.

[9] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, 20(2):565–592, 2010.

[10] G. Elidan, I. McGraw, and D. Koller. Residual belief propagation: Informal scheduling for asynchronous message passing. In *Proceedings of UAI*, 2006.

[11] A. Globerson and T. Jaakkola. Convergent propagation algorithms via oriented trees. In *Proceedings of UAI*, 2007.

[12] L. A. Goldberg, M. Jerrum, and M. Paterson. The computational complexity of two-state spin systems. *Random Structures & Algorithms*, 23(2):133–154, 2003.

[13] R. B. Griffiths. Correlations in Ising ferromagnets. ii. external magnetic fields. *Journal of Mathematical Physics*, 8(3):484–489, 1967.

[14] F. Harary et al. On the notion of balance of a signed graph. *The Michigan Mathematical Journal*, 2(2):143–146, 1953.

[15] T. Hazan and A. Shashua. Convergent message-passing algorithms for inference over general graphs with convex free energies. In *Proceedings of UAI*, 2008.

[16] A. Ihler, J. Fisher, and A. Willsky. Loopy belief propagation: convergence and effects of message errors. In *Journal of Machine Learning Research*, pages 905–936, 2005.

[17] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22(5):1087–1116, 1993.

[18] J. K. Johnson, D. Oyen, M. Chertkov, and P. Netrapalli. Learning planar Ising models. *The Journal of Machine Learning Research*, 17(1):7539–7564, 2016.

[19] M. Jordan. Graphical models. *Statistical Science*, pages 140–155, 2004.

[20] C. Knoll, D. Mehta, T. Chen, and F. Perko. Fixed Points of Belief Propagation–An Analysis via Polynomial Homotopy Continuation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017.
C. Knoll and F. Pernkopf. Loopy Belief Propagation—Local Stability Analysis for Non-Vanishing Fields. In Proceedings of UAI, 2017.

C. Knoll and F. Pernkopf. Belief propagation: Accurate marginals or accurate partition function – where is the difference? In Proceedings of UAI, 2019.

C. Knoll, M. Rath, S. Tschiatschek, and F. Pernkopf. Message Scheduling Methods for Belief Propagation. In Machine Learning and Knowledge Discovery in Databases, pages 295–310. Springer, 2015.

F. Koehler. Fast convergence of belief propagation to global optima: Beyond correlation decay. In Proceedings of NeurIPS, 2019.

D. Koller and N. Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT press, 2009.

F. Kschischang, B. Frey, and H. Loeliger. Factor graphs and the sum–product algorithm. IEEE Transaction on Information Theory, 47(2), 2001.

V. Martin, J.-M. Lasgouttes, and C. Furtlehner. The role of normalization in the belief propagation algorithm. preprint arXiv:1101.4170, 2011.

T. Meltzer, A. Globerson, and Y. Weiss. Convergent message passing algorithms: a unifying view. In Proceedings of UAI, 2009.

O. Meshi, A. Jaimovich, A. Globerson, and N. Friedman. Convexifying the Bethe free energy. In Proceedings of UAI, 2009.

M. Mezard and A. Montanari. Information, Physics, and Computation. Oxford Univ. Press, 2009.

M. Mezard, G. Parisi, and M. Virasoro. Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications, volume 9. World Scientific Publishing Co Inc, 1987.

J. Mooij and H. Kappen. On the properties of the Bethe approximation and loopy belief propagation on binary networks. In Journal of Machine Learning Research, volume 11, pages 1101–1138, 2010.

J. M. Mooij and H. J. Kappen. Sufficient conditions for convergence of the sum–product algorithm. IEEE Trans. on Information Theory, 53(12):4422–4437, 2007.

K. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: an empirical study. In Proceedings of UAI, 1999.

F. Pernkopf, R. Pecharz, and S. Tschiatschek. Introduction to Probabilistic Graphical Models. Academic Press’ Library in Signal Processing, 2014.

G. Perugini and F. Ricci-Tersenghi. Improved belief propagation algorithm finds many Bethe states in the random-field Ising model on random graphs. Physical Review E, 97(1):012152, 2018.

X. Pirkow, Y. Ahmadian, and K. D. Miller. Learning unbelievable probabilities. In Proceedings of NIPS, 2011.

R. T. Rockafellar and R. J.-B. Wets. Variational Analysis, volume 317. Springer Science & Business Media, 2009.

W. Rudin. Principles of mathematical analysis, volume 3. McGraw-hill New York, 1964.

N. Ruozzi. Beyond log-supermodularity: lower bounds and the Bethe partition function. In Proceedings of UAI, 2013.

A. Saade, F. Krzakala, and L. Zdeborová. Spectral bounds for the Ising ferromagnet on an arbitrary given graph. Journal of Statistical Mechanics: Theory and Experiment, 2017(5):053403, 2017.

J. Shin. Complexity of Bethe Approximation. In Proceedings of AISTATS, pages 1037–1045, 2012.

A. Sommese and C. Wampler. The Numerical Solution of Systems of Polynomials Arising in Engineering and Science, volume 99. World Scientific, 2005.

D. Sontag and T. S. Jaakkola. New outer bounds on the marginal polytope. In Proceedings of NIPS, 2008.

C. Srinivasas, S. Ravanbaksh, and B. Frey. Survey Propagation beyond Constraint Satisfaction Problems. In Proceedings of AISTATS, pages 286–295, 2016.

R. A. S. Willsky, E. B. Sudderth, and M. J. Wainwright. Loop series and Bethe variational bounds in attractive graphical models. In NIPS, 2008.

J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. IEEE Trans. on Information Theory, 51(7):2282–2312, 2005.

A. Yuille and A. Rangarajan. The concave-convex procedure. Neural Computation, 15(4):915–936, 2003.

L. Zdeborová and F. Krzakala. Generalization of the cavity method for adiabatic evolution of gibbs states. Physical Review B, 81(22), 2010.

L. Zdeborová and F. Krzakala. Statistical physics of inference: Thresholds and algorithms. Advances in Physics, 65(5):453–555, 2016.