Merging Belief Propagation and the Mean Field Approximation: A Free Energy Approach

Riegler, Erwin; Kirkelund, Gunvor Elisabeth; Manchón, Carles Navarro; Badiu, Mihai-Alin; Fleury, Bernard Henri

Published in:
I E E E Transactions on Information Theory

DOI (link to publication from Publisher):
10.1109/TIT.2012.2218573

Publication date:
2013

Document Version
Early version, also known as pre-print

Merging Belief Propagation and the Mean Field Approximation: A Free Energy Approach

Erwin Riegler*, Gunvor Elisabeth Kirkelund†, Carles Navarro Manchón†, Mihai-Alin Badiu†‡, Bernard Henri Fleury†
* Vienna University of Technology (VUT), Austria
E-mail: erwin.riegler@nt.tuwien.ac.at
† Aalborg University (AAU), Denmark
E-mail: gunvor@es.aau.dk, cnm@es.aau.dk, in_mba@es.aau.dk, bfl@es.aau.dk
‡ Technical University of Cluj-Napoca, Romania
E-mail: Mihai.Badiu@com.utcluj.ro

***Abstract*—We present a joint message passing approach that combines belief propagation and the mean field approximation. Our analysis is based on the region-based free energy approximation method proposed by Yedidia et al. We show that the message passing fixed-point equations obtained with this combination correspond to stationary points of a constrained region-based free energy approximation. Moreover, we present a convergent implementation of these message passing fixed-point equations provided that the underlying factor graph fulfills certain technical conditions. In addition, we show how to include hard constraints in the part of the factor graph corresponding to belief propagation. Finally, we demonstrate an application of our method to iterative channel estimation and decoding in an OFDM system.**

I. Introduction

Variational techniques have been used for decades in quantum and statistical physics, where they are referred to as the *mean field* (MF) approximation [2]. Later, they found their way to the area of machine learning or statistical inference, see, e.g., [3]–[6]. The basic idea of variational inference is to derive the statistics of "hidden" random variables given the knowledge of "visible" random variables of a certain probability density function (pdf). In the MF approximation, this pdf is approximated by some "simpler," e.g., (fully) factorized pdf and minimizing the Kullback-Leibler divergence between the approximating and the true pdf, which can be done in an iterative, i.e., message passing like way. Apart from being fully factorized, the approximating pdf typically fulfills additional constraints that allow for messages with a simple structure, which can be updated in a simple way. For example, additional exponential conjugacy constraints result in messages propagating along the edges of the underlying Bayesian network that are described by a small number of parameters [5]. Variational inference methods were recently applied in [7] to the *channel state estimation/interference cancellation part* of a class of MIMO-OFDM receivers that iterate between detection, channel estimation, and decoding.

An approach different from the MF approximation is *belief propagation* (BP) [8]. Roughly speaking, with BP one tries to find *local* approximations, which are—exactly or approximately—the marginals of a certain pdf[1]. This can also be done in an iterative way, where messages are passed along the edges of a factor graph [10]. A typical application of BP is *decoding* of turbo or low density parity check (LDPC) codes. Based on the excellent performance of BP, a lot of variations have been derived in order to improve the performance of this algorithm even further. For example, minimizing an upper bound on the log partition function of a pdf leads to the powerful tree reweighted BP algorithm [11]. An offspring of this idea is the recently developed uniformly tree reweighted BP algorithm [12]. Another example is [13] where methods from information geometry are used to compute correction terms for the beliefs obtained by loopy BP. An alternative approach for turbo decoding that uses projections (that are dual in the sense of [14, Ch. 3] to the one used in [13]) on constraint subsets can be found in [15]. A combination of the approaches used in [13] and in [15] can be found in [16].

Both methods, BP and the MF approximation, have their own virtues and disadvantages. For example, the MF approximation

- \+ always admits a convergent implementation;
- \+ has simple message passing update rules, in particular for conjugate-exponential models;
- – is not compatible with hard constraints,

and BP

- \+ yields a good approximation of the marginal distributions if the factor graph has no short cycles;
- \+ is compatible with hard constraints like, e.g., code constraints;
- – may have a high complexity, especially when applied to probabilistic models involving both, discrete and continuous random variables.

This work was supported by the WWTF grant ICT10-066, the FWF grant S10603-N13 within the National Research Network SISE, Renesas Mobile Corporation, the 4GMCT cooperative research project funded by Intel Mobile Communications, Agilent Technologies, Aalborg University and the Danish National Advanced Technology Foundation, and the project SIDOC under contract no. POSDRU/88/1.5/S/60078. The research of Mihai-Alin Badiu was carried out while he visited Aalborg University. The results of this paper have been presented partially in [1].

[1]Following the notation used in [9], we use the name BP also for loopy BP.

Hence, it is of great benefit to apply BP and the MF approximation on the same factor graph in such a combination that their respective virtues can be exploited while circumventing their drawbacks. To this end a *unified message passing algorithm* is needed that allows for combining both approaches.

The fixed-point equations of both—BP and the MF approximation—can be obtained by minimizing an approximation of the Kullback-Leibler divergence, called region-based free energy approximation. This approach differs from other methods, see, e.g., [17][2], because the starting point for the derivation of the corresponding message passing fixed-point equations is the same objective function for both, BP and the MF approximation. The main technical result of this work is Theorem 2, where we show that the message passing fixed-point equations for such a combination of BP and the MF approximation correspond to stationary points of one single constrained region-based free energy approximation and provide a clear rule stating how to couple the messages propagating in the BP and MF part. In fact, based on the factor graph corresponding to a factorization of a probability mass function (pmf) and a choice for a separation of this factorization into BP and MF factors, Theorem 2 gives the message passing fixed-point equations for the factor graph representing the whole factorization of the pmf. One example of an application of Theorem 2 is joint channel estimation, interference cancellation, and decoding. Typically these tasks are considered separately and the coupling between them is described in a heuristic way. As an example of this problematic, there has been a debate in the research community on whether a posteriori probabilities (APP) or extrinsic values should be fed back from the decoder to the rest of the receiver components; several authors coincide in proposing the use of extrinsic values for MIMO detection [18]–[20] while using APP values for channel estimation [19], [20], but no thorough justification for this choice is given apart from its superior performance shown by simulation results. Despite having a clear rule to update the messages for the whole factor graph representing a factorization of a pmf, an additional advantage is the fact that solutions of fixed-point equations for the messages are related to the stationary points of the corresponding constrained region-based free energy approximation. This correspondence is important because it yields an interpretation of the computed beliefs for arbitrary factor graphs similar to the case of solely BP, where solutions of the message passing fixed-point equations do in general not correspond to the true marginals if the factor graph has cycles but always correspond to stationary points of the constrained Bethe free energy [9]. Moreover, this observation allows us to present a systematic way of updating the messages, namely, Algorithm 1, that is guaranteed to converge provided that the factor graph representing the factorization of the pmf fulfills certain technical conditions.

The paper is organized as follows. In the remainder of this section we fix our notation. Section II is devoted to the introduction of the region-based free energy approximations proposed by [9] and to recall how BP, the MF approximation, and the EM algorithm [21] can be obtained by this method. Since the MF approximation is typically used for parameter estimation, we briefly show how to extend it to the case of continuous random variables using an approach presented already in [22, pp. 36–38] that avoids complicated methods from variational calculus. Section III is the main part of this work. There we state our main result, namely, Theorem 2, and show how the message passing fixed-point equations of a combination of BP and the MF approximation can be related to the stationary points of the corresponding constrained region-based free energy approximation. We then (i) prove Lemma 2, which generalizes Theorem 2 to the case where the factors of the pmf in the BP part are no longer restricted to be strictly positive real-valued functions, and (ii) present Algorithm 1 that is a convergent implementation of the message passing update equations presented in Theorem 2 provided that the factor graph representing the factorization of the pmf fulfills certain technical conditions. As a byproduct, (i) gives insights for solely BP (which is a special case of the combination of BP and the MF approximation) with hard constraints, where only conjectures are formulated in [9]. In Section IV we apply Algorithm 1 to joint channel estimation and decoding in an OFDM system. More advanced receiver architectures together with numerical simulations and a comparison with other state of the art receivers can be found in [23]. Finally, we conclude in Section V and present an outlook for further research directions.

A. Notation

Capital calligraphic letters $\mathcal{A}, \mathcal{I}, \mathcal{N}$ denote finite sets. The cardinality of a set $\mathcal{I}$ is denoted by $|\mathcal{I}|$. If $i \in \mathcal{I}$ we write $\mathcal{I} \setminus i$ for $\mathcal{I} \setminus \{i\}$. We use the convention that $\prod_{\emptyset}(\dots) \triangleq 1$ where $\emptyset$ denotes the empty set. For any finite set $\mathcal{I}$, $I_{\mathcal{I}}$ denotes the indicator function on $\mathcal{I}$, i.e., $\mathrm{I}_{\mathcal{I}}(i) = 1$ if $i \in \mathcal{I}$ and $\mathrm{I}_{\mathcal{I}}(i) = 0$ else. We denote by capital letters X discrete random variables with a finite number of realizations and pmf p_X. For a random variable X we use the convention that x is a representative for all possible realizations of X, i.e., x serves as a running variable, and denote a particular realization by $\bar{x}$. For example, $\sum_{x}(\dots)$ runs through all possible realizations x of X and for two functions f and g depending on all realizations x of X, $f(x) = g(x)$ means that $f(\bar{x}) = g(\bar{x})$ for each particular realization $\bar{x}$ of X. If F is a functional of a pmf p_X of a random variable X and g is a function depending on all realizations x of X, then $\frac{\partial F}{\partial p(x)} = g(x)$ means that $\frac{\partial F}{\partial p(\bar{x})} = g(\bar{x})$ is well defined and holds for each particular realization $\bar{x}$ of X. We write $\mathbf{x} = (x_i \mid i \in \mathcal{I})^{\mathrm{T}}$ for the realizations of the vector of random variables $\mathbf{X} = (X_i \mid i \in \mathcal{I})^{\mathrm{T}}$. If $i \in \mathcal{I}$ then $\sum_{\mathbf{x} \setminus x_i}(\dots)$ runs through all possible realizations of $\mathbf{X}$ but X_i. For any nonnegative real valued function f with argument $\mathbf{x} = (x_i \mid i \in \mathcal{I})^{\mathrm{T}}$ and $i \in \mathcal{I}$, $f\,|_{\bar{x}_i}$ denotes f with fixed argument $x_i = \bar{x}_i$. If a function f is identically zero we write $f \equiv 0$ and $f \not\equiv 0$ means that it is not identically zero. For two real valued functions f and g with the same domain and argument x, we write $f(x) \propto g(x)$ if $f = cg$ for some

[2] An information geometric interpretation of the different objective functions used in [17] can be found in [14, Ch. 2].

real positive constant $c \in \mathbb{R}_+$. We use the convention that $0\ln(0) = 0$, $a\ln(\frac{a}{0}) = \infty$ if $a > 0$, and $0\ln(\frac{0}{0}) = 0$ [24, p. 31]. For $x \in \mathbb{R}$, $\delta(x) = 1$ if $x = 0$ and zero else. Matrices $\mathbf{\Lambda} \in \mathbb{C}^{m\times n}$ are denoted by capital boldface Greek letters. The superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ stand for transposition and Hermitian transposition, respectively. For a matrix $\mathbf{\Lambda} \in \mathbb{C}^{m\times n}$, the entry in the ith row and jth column is denoted by $\lambda_{i,j} = [\mathbf{\Lambda}]_{i,j}$. For two vectors $\mathbf{x} = (x_i \mid i \in \mathcal{I})^{\mathrm{T}}$ and $\mathbf{y} = (y_i \mid i \in \mathcal{I})^{\mathrm{T}}$, $\mathbf{x} \odot \mathbf{y} = (x_i y_i \mid i \in \mathcal{I})^{\mathrm{T}}$ denotes the Hadamard product of $\mathbf{x}$ and $\mathbf{y}$. Finally, $\mathrm{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma})$ stands for the pdf of a jointly proper complex Gaussian random vector $\mathbf{X} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$.

II. Known results

A. Region-based free energy approximations [9]

Let $p_{\mathbf{X}}$ be a certain positive pmf of a vector $\mathbf{X}$ of random variables X_i $(i \in \mathcal{I})$ that factorizes as

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{a\in\mathcal{A}} f_a(\mathbf{x}_a) \tag{1}$$

where $\mathbf{x} \triangleq (x_i \mid i \in \mathcal{I})^{\mathrm{T}}$ and $\mathbf{x}_a \triangleq (x_i \mid i \in \mathcal{N}(a))^{\mathrm{T}}$ with $\mathcal{N}(a) \subseteq \mathcal{I}$ for all $a \in \mathcal{A}$. Without loss of generality we assume that $\mathcal{A} \cap \mathcal{I} = \emptyset$, which can always be achieved by renaming indices.[3] Since $p_{\mathbf{X}}$ is a strictly positive pmf, we can assume without loss of generality that all the factors f_a of $p_{\mathbf{X}}$ in (1) are real-valued positive functions. Later in Section III, we shall show how to relax the positivity constraint for some of these factors. The factorization in (1) can be visualized in a *factor graph* [10][4]. In a factor graph, $\mathcal{N}(a)$ is the set of all variable nodes connected to a factor node $a \in \mathcal{A}$ and $\mathcal{N}(i)$ represents the set of all factor nodes connected to a variable node $i \in \mathcal{I}$. An example of a factor graph is depicted in Figure 1.

A *region* $R \triangleq (\mathcal{I}_R, \mathcal{A}_R)$ consists of subsets of indices $\mathcal{I}_R \subseteq \mathcal{I}$ and $\mathcal{A}_R \subseteq \mathcal{A}$ with the restriction that $a \in \mathcal{A}_R$ implies that $\mathcal{N}(a) \subseteq \mathcal{I}_R$. To each region R we associate a *counting number* $c_R \in \mathbb{Z}$. A set $\mathcal{R} \triangleq \{(R, c_R)\}$ of regions and associated counting numbers is called *valid* if

$$\sum_{(R,c_R)\in\mathcal{R}} c_R \,\mathrm{I}_{\mathcal{A}_R}(a) = \sum_{(R,c_R)\in\mathcal{R}} c_R \,\mathrm{I}_{\mathcal{I}_R}(i) = 1, \quad \forall\, a \in \mathcal{A}, i \in \mathcal{I}.$$

[3]For example, we can write

$$\mathcal{I} = \{1, 2, \dots, |\mathcal{I}|\}$$
$$\mathcal{A} = \{\overline{1}, \overline{2}, \dots, \overline{|\mathcal{A}|}\}.$$

This implies that any function that is defined pointwise on $\mathcal{A}$ and $\mathcal{I}$ is well defined. For example, if in addition to the definition of the sets $\mathcal{N}(a)$ $(a \in \mathcal{A})$ we set $\mathcal{N}(i) \triangleq \{a \in \mathcal{A} \mid i \in \mathcal{N}(a)\}$ for all $i \in \mathcal{I}$, the function

$$\begin{aligned}\mathcal{N} : \mathcal{I} \cup \mathcal{A} &\to \Pi(\mathcal{I} \cup \mathcal{A})\\ a &\mapsto \mathcal{N}(a), \quad \forall\, a \in \mathcal{A}\\ i &\mapsto \mathcal{N}(i), \quad \forall\, i \in \mathcal{I}\end{aligned}$$

with $\Pi(\mathcal{I} \cup \mathcal{A})$ denoting the collection of all subsets of $\mathcal{I} \cup \mathcal{A}$ is well defined because $i \neq a\ \forall i \in \mathcal{I}, a \in \mathcal{A}$.

[4]Throughout the paper we work with Tanner factor graphs as opposed to Forney factor graphs.

For a positive function b approximating $p_{\mathbf{X}}$, we define the *variational free energy* [9][5]

$$\begin{aligned} F(b) &\triangleq \sum_{\mathbf{x}} b(\mathbf{x}) \ln \frac{b(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} \\ &= \underbrace{\sum_{\mathbf{x}} b(\mathbf{x}) \ln b(\mathbf{x})}_{\triangleq -H(b)} - \underbrace{\sum_{\mathbf{x}} b(\mathbf{x}) \ln p_{\mathbf{X}}(\mathbf{x})}_{\triangleq -U(b)}. \end{aligned} \tag{2}$$

In (2), $H(b)$ denotes the entropy [24, p. 5] of b and $U(b)$ is called average energy of b. Note that $F(b)$ is the Kullback-Leibler divergence [24, p. 19] between b and $p_{\mathbf{X}}$, i.e., $F(b) = D(b \,||\, p_{\mathbf{X}})$. For a set $\mathcal{R}$ of regions and associated counting numbers, the *region-based free energy approximation* is defined as [9] $F_{\mathcal{R}} \triangleq U_{\mathcal{R}} - H_{\mathcal{R}}$ with

$$\begin{aligned} U_{\mathcal{R}} &\triangleq - \sum_{(R,c_R)\in\mathcal{R}} c_R \sum_{a\in\mathcal{A}_R} \sum_{\mathbf{x}_R} b_R(\mathbf{x}_R) \ln f_a(\mathbf{x}_a) \\ H_{\mathcal{R}} &\triangleq - \sum_{(R,c_R)\in\mathcal{R}} c_R \sum_{\mathbf{x}_R} b_R(\mathbf{x}_R) \ln b_R(\mathbf{x}_R). \end{aligned}$$

Here, each b_R is defined locally on a region R. Instead of minimizing F with respect to b, we minimize $F_{\mathcal{R}}$ with respect to all b_R $((R, c_R) \in \mathcal{R})$ where the b_R have to fulfill certain constraints. The quantities b_R are called *beliefs*. We give two examples of valid sets of regions and associated counting numbers.

Example 2.1: The trivial example $\mathcal{R}_{\text{MF}} \triangleq \{((\mathcal{I}, \mathcal{A}), 1)\}$. It leads to the MF fixed-point equations, as will be shown in Subsection II-C.

Example 2.2: We define two types of regions:

1) *large regions:* $R_a \triangleq (\mathcal{N}(a), \{a\})$ with $c_{R_a} = 1\ \forall\, a \in \mathcal{A}$;
2) *small regions:* $R_i \triangleq (\{i\}, \emptyset)$ with $c_{R_i} = 1 - |\mathcal{N}(i)|\ \forall\, i \in \mathcal{I}$.

Note that this definition is well defined due to our assumption that $\mathcal{A} \cap \mathcal{I} = \emptyset$. The region-based free energy approximation corresponding to the valid set of regions and associated counting numbers

$$\mathcal{R}_{\text{BP}} \triangleq \{(R_i, c_{R_i}) \mid i \in \mathcal{I}\} \cup \{(R_a, c_{R_a}) \mid a \in \mathcal{A}\}$$

is called the *Bethe free energy* [9], [25]. It leads to the BP fixed-point equations, as will be shown in Subsection II-B. The Bethe free energy is equal to the variational free energy when the factor graph has no cycles [9].

B. BP fixed-point equations

The fixed-point equations for BP can be obtained from the Bethe free energy by imposing additional marginalization constraints and computing the stationary points of the corresponding Lagrangian function [9], [26]. The Bethe free energy

[5]If p is not normalized to one, the definition of the variational free energy contains an additional normalization constant, called Helmholtz free energy [9, pp. 4–5].

reads

$$F_{\mathrm{BP}} = \sum_{a\in\mathcal{A}}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln\frac{b_a(\mathbf{x}_a)}{f_a(\mathbf{x}_a)} - \sum_{i\in\mathcal{I}}(|\mathcal{N}(i)|-1)\sum_{x_i} b_i(x_i)\ln b_i(x_i) \tag{3}$$

with $b_a \triangleq b_{R_a}\ \forall\ a\in\mathcal{A}$, $b_i \triangleq b_{R_i}\ \forall\ i\in\mathcal{I}$, and $F_{\mathrm{BP}} \triangleq F_{\mathcal{R}_{\mathrm{BP}}}$. The normalization constraints for the beliefs b_i $(i\in\mathcal{I})$ and the marginalization constraints for the beliefs b_a $(a\in\mathcal{A})$ can be included in the Lagrangian [27, Sec. 3.1.3]

$$L_{\mathrm{BP}} \triangleq F_{\mathrm{BP}} - \sum_{a\in\mathcal{A}}\sum_{i\in\mathcal{N}(a)}\sum_{x_i}\lambda_{a,i}(x_i)\Big(b_i(x_i) - \sum_{\mathbf{x}_a\setminus x_i} b_a(\mathbf{x}_a)\Big) - \sum_{a\in\mathcal{A}}\gamma_a\Big(\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) - 1\Big). \tag{4}$$

The stationary points of the Lagrangian in (4) are then related to the BP fixed-point equations by the following theorem.

Theorem 1: [9, Th. 2] Stationary points of the Lagrangian in (4) must be BP fixed-points with positive beliefs fulfilling

$$\begin{cases} b_a(\mathbf{x}_a) = z_a\, f_a(\mathbf{x}_a)\displaystyle\prod_{i\in\mathcal{N}(a)} n_{i\to a}(x_i), & \forall\ a\in\mathcal{A}\\ b_i(x_i) = \displaystyle\prod_{a\in\mathcal{N}(i)} m_{a\to i}(x_i), & \forall\ i\in\mathcal{I}\end{cases} \tag{5}$$

with

$$\begin{cases} m_{a\to i}(x_i) = z_a\displaystyle\sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a)\prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j)\\ n_{i\to a}(x_i) = \displaystyle\prod_{c\in\mathcal{N}(i)\setminus a} m_{c\to i}(x_i)\end{cases} \tag{6}$$

for all $a\in\mathcal{A}, i\in\mathcal{N}(a)$ and vice versa. Here, z_a $(a\in\mathcal{A})$ are positive constants that ensure that the beliefs b_a $(a\in\mathcal{A})$ are normalized to one.

Often, the following alternative system of fixed-point equations is solved instead of (6).

$$\begin{cases} \widetilde{m}_{a\to i}(x_i) = \omega_{a,i}\displaystyle\sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a)\prod_{j\in\mathcal{N}(a)\setminus i} \widetilde{n}_{j\to a}(x_j)\\ \widetilde{n}_{i\to a}(x_i) = \displaystyle\prod_{c\in\mathcal{N}(i)\setminus a} \widetilde{m}_{c\to i}(x_i)\end{cases} \tag{7}$$

for all $a\in\mathcal{A}, i\in\mathcal{N}(a)$ where $\omega_{a,i}$ $(a\in\mathcal{A}, i\in\mathcal{N}(a))$ are arbitrary positive constants. The reason for this is that for a fixed scheduling the messages computed in (6) differ from the messages computed in (7) only by positive constants, which drop out when the beliefs are normalized. See also [9, Eq. (68) and Eq. (69)], where the " $\propto$ " symbol is used in the update equations indicating that the normalization constants are irrelevant. A solution of (7) can be obtained, e.g., by updating corresponding likelihood ratios of the messages in (6) or by updating the messages according to (6) but ignoring the normalization constants z_a $(a\in\mathcal{A})$. The algorithm converges if the normalized beliefs do not change any more. Therefore, a rescaling of the messages is irrelevant and a solution of (7) is obtained. However, we note that rescaling a solution of (7) has not necessarily to be a solution of (6). Hence, the beliefs obtained by solving (7) need not be stationary points of the Lagrangian in (4). To the best of our knowledge, this elementary insight is not published yet in the literature and we state a necessary and sufficient condition when a solution of (7) can be rescaled to a solution of (6) in the following lemma.

Lemma 1: Suppose that $\{\widetilde{m}_{a\to i}(x_i), \widetilde{n}_{i\to a}(x_i)\}$ $(a\in\mathcal{A}, i\in\mathcal{N}(a))$ is a solution of (7) and set

$$\tilde{z}_a \triangleq \frac{1}{\sum_{\mathbf{x}_a} f_a(\mathbf{x}_a)\prod_{i\in\mathcal{N}(a)}\widetilde{n}_{i\to a}(x_i)}, \quad \forall a\in\mathcal{A}. \tag{8}$$

Then this solution can be rescaled to a solution of (6) if and only if there exist positive constants g_i $(i\in\mathcal{I})$ such that

$$\omega_{a,i} = g_i\tilde{z}_a, \quad \forall\ a\in\mathcal{A}, i\in\mathcal{N}(a). \tag{9}$$

Proof: See Appendix A. ∎

Remark 2.1: Note that for factor graphs that have a tree-structure the messages obtained by running the forward-backward algorithm [10] always fulfill (9) because we have $\omega_{a,i} = 1$ $(a\in\mathcal{A}, i\in\mathcal{N}(a))$ and $\tilde{z}_a = 1$ $(a\in\mathcal{A})$ in this case.

C. Fixed point equations for the MF approximation

A message passing interpretation of the MF approximation was derived in [5], [28]. In this section, we briefly show how the corresponding fixed-point equations can be obtained by the free energy approach. To this end we use $\mathcal{R}_{\mathrm{MF}}$ from Example 2.1 together with the factorization constraint[6]

$$b(\mathbf{x}) = \prod_{i\in\mathcal{I}} b_i(x_i). \tag{10}$$

Plugging (10) into the expression for the region-based free energy approximation corresponding to the trivial approximation $\mathcal{R}_{\mathrm{MF}}$ we get

$$F_{\mathrm{MF}} = \sum_{i\in\mathcal{I}}\sum_{x_i} b_i(x_i)\ln b_i(x_i) - \sum_{a\in\mathcal{A}}\sum_{\mathbf{x}_a}\prod_{i\in\mathcal{N}(a)} b_i(x_i)\ln f_a(\mathbf{x}_a) \tag{11}$$

with $F_{\mathrm{MF}} \triangleq F_{\mathcal{R}_{\mathrm{MF}}}$. Assuming that all the beliefs b_i $(i\in\mathcal{I})$ have to fulfill a normalization constraint, the stationary points of the corresponding Lagrangian for the MF approximation can easily be evaluated to be

$$b_i(x_i) = z_i\exp\left(\sum_{a\in\mathcal{N}(i)}\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j)\ln f_a(\mathbf{x}_a)\right) \tag{12}$$

for all $i\in\mathcal{I}$ where the positive constants z_i $(i\in\mathcal{I})$ are such that b_i is normalized to one for all $i\in\mathcal{I}$.[7]

For the MF approximation there always exists a convergent algorithm that computes beliefs b_i $(i\in\mathcal{I})$ solving (12) by simply using (12) as an iterative update equation for the beliefs. At each step the objective function, i.e., the Lagrangian corresponding to the region-based free energy approximation

[6]For binary random variables with pmf in an exponential family it was shown in [29] that this gives a good approximation whenever the truncation of the Plefka expansion does not introduce a significant error.

[7]The Lagrange multiplier [27, p. 283] for each belief b_i $(i\in\mathcal{I})$ corresponding to the normalization constraint can be absorbed into the positive constant z_i $(i\in\mathcal{I})$.

of the MF approximation (11), cannot increase and the algorithm is guaranteed to converge. Note that in order to derive a particular update b_i $(i \in \mathcal{I})$ we need all previous updates b_j with $j \in \bigcup_{a\in\mathcal{N}(i)} \mathcal{N}(a) \setminus i$.

By setting $n_{i\to a}(x_i) \triangleq b_i(x_i) \ \forall\ i \in \mathcal{I}, a \in \mathcal{N}(i)$ the fixed-point equations in (12) are transformed into the message passing fixed-point equations

$$\begin{cases} n_{i\to a}(x_i) = z_i \prod_{a\in\mathcal{N}(i)} m_{a\to i}(x_i) \\ m_{a\to i}(x_i) = \exp\left(\sum_{\mathbf{x}_a\setminus x_i} \prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j) \ln f_a(\mathbf{x}_a) \right) \end{cases} \tag{13}$$

for all $a \in \mathcal{A}, i \in \mathcal{N}(a)$. The MF approximation can be extended to the case where $p_{\mathbf{X}}$ is a pdf, which is shown in Appendix B. Formally, each sum over x_i $(i \in \mathcal{I})$ in (12) and (13) has to be replaced by a Lebesgue integral whenever the corresponding random variable X_i is continuous.

D. EM algorithm

Message passing interpretations of the EM algorithm [21] were derived in [30], [31]. It can be shown that the EM algorithm is a special instance of the MF approximation [32, Sec. 2.3.1], which we briefly summarize in the following. Suppose that we apply the MF approximation to $p_{\mathbf{X}}$ in (1) as described before. In addition, we assume that for all $i \in \mathcal{E} \subseteq \mathcal{I}$ the beliefs b_i fulfill the constraints that $b_i(x_i) = \delta(x_i - \tilde{x}_i)$. Using the fact that $0\ln(0) = 0$, we can rewrite F_{MF} in (11) as

$$\begin{aligned} F_{\text{MF}} = &\sum_{i\in\mathcal{I}\setminus\mathcal{E}} \sum_{x_i} b_i(x_i) \ln b_i(x_i) \\ &- \sum_{a\in\mathcal{A}} \sum_{\mathbf{x}_a} \prod_{i\in\mathcal{N}(a)} b_i(x_i) \ln f_a(\mathbf{x}_a). \end{aligned} \tag{14}$$

For all $i \in \mathcal{I} \setminus \mathcal{E}$ the stationary points of F_{MF} in (14) have the same analytical expression as the one obtained in (12). For $i \in \mathcal{E}$, minimizing F_{MF} in (14) with respect to $\tilde{x}_i$ yields

$$\tilde{x}_i = \underset{x_i}{\operatorname{argmin}}(F_{\text{MF}}) \tag{15}$$

$$= \underset{x_i}{\operatorname{argmax}} \left(\prod_{a\in\mathcal{N}(i)} \exp\left(\sum_{\mathbf{x}_a\setminus x_i} \prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j) \ln f_a(\mathbf{x}_a) \right) \right). \tag{16}$$

Setting $n_{i\to a}(x_i) \triangleq b_i(x_i) \ \forall\ i \in \mathcal{I}, a \in \mathcal{N}(i)$, we get the message passing update equations defined in (13) *except* that we have to replace the messages $n_{i\to a}(x_i)$ for all $i \in \mathcal{E}$ and $a \in \mathcal{N}(i)$ by

$$\begin{aligned} n_{i\to a}(x_i) = &\delta(x_i - \tilde{x}_i) \\ &\text{with } \tilde{x}_i = \underset{x_i}{\operatorname{argmax}} \left(\prod_{a\in\mathcal{N}(i)} m_{a\to i}(x_i) \right) \end{aligned} \tag{17}$$

for all $i \in \mathcal{E}, a \in \mathcal{N}(a)$.

III. Combined BP / MF approximation fixed-point equations

Let

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{a\in\mathcal{A}_{\text{MF}}} f_a(\mathbf{x}_a) \prod_{b\in\mathcal{A}_{\text{BP}}} f_b(\mathbf{x}_b) \tag{18}$$

be a partially factorized pmf with $\mathcal{A}_{\text{MF}} \cap \mathcal{A}_{\text{BP}} = \emptyset$ and $\mathcal{A} \triangleq \mathcal{A}_{\text{MF}} \cup \mathcal{A}_{\text{BP}}$. As before, we have $\mathbf{x} \triangleq (x_i \mid i \in \mathcal{I})$, $\mathbf{x}_a \triangleq (x_i \mid i \in \mathcal{N}(a))^{\text{T}}$ with $\mathcal{N}(a) \subseteq \mathcal{I}$ for all $a \in \mathcal{A}$, and $\mathcal{N}(i) \triangleq \{a \in \mathcal{A} \mid i \in \mathcal{N}(a)\}$ for all $i \in \mathcal{I}$. We refer to the factor graph representing the factorization $\prod_{a\in\mathcal{A}_{\text{BP}}} f_a(\mathbf{x}_a)$ in (18) as "BP part" and to the factor graph representing the factorization $\prod_{a\in\mathcal{A}_{\text{MF}}} f_a(\mathbf{x}_a)$ in (18) as "MF part". Furthermore, we set

$$\mathcal{I}_{\text{MF}} \triangleq \bigcup_{a\in\mathcal{A}_{\text{MF}}} \mathcal{N}(a) \qquad \mathcal{I}_{\text{BP}} \triangleq \bigcup_{a\in\mathcal{A}_{\text{BP}}} \mathcal{N}(a)$$

and

$$\mathcal{N}_{\text{MF}}(i) \triangleq \mathcal{A}_{\text{MF}} \cap \mathcal{N}(i) \qquad \mathcal{N}_{\text{BP}}(i) \triangleq \mathcal{A}_{\text{BP}} \cap \mathcal{N}(i).$$

Next, we define the following regions and counting numbers:

1) one MF region $R_{\text{MF}} \triangleq (\mathcal{I}_{\text{MF}}, \mathcal{A}_{\text{MF}})$ with $c_{R_{\text{MF}}} = 1$;
2) small regions $R_i \triangleq (\{i\}, \emptyset)$ with $c_{R_i} = 1 - |\mathcal{N}_{\text{BP}}(i)| - \text{I}_{\mathcal{I}_{\text{MF}}}(i)$ for all $i \in \mathcal{I}_{\text{BP}}$;
3) large regions $R_a \triangleq (\mathcal{N}(a), \{a\})$ with $c_{R_a} = 1$ for all $a \in \mathcal{A}_{\text{BP}}$.

This yields the valid set of regions and associated counting numbers

$$\begin{aligned} \mathcal{R}_{\text{BP, MF}} \triangleq &\{(R_i, c_{R_i}) \mid i \in \mathcal{I}_{\text{BP}}\} \cup \{(R_a, c_{R_a}) \mid a \in \mathcal{A}_{\text{BP}}\} \\ &\cup \{(R_{\text{MF}}, c_{R_{\text{MF}}})\}. \end{aligned} \tag{19}$$

The additional terms $\text{I}_{\mathcal{I}_{\text{MF}}}(i)$ in the counting numbers of the small regions R_i $(i \in \mathcal{I})$ defined in 2) compared to the counting numbers of the small regions for the Bethe approximation (see Example 2.2) guarantee that $\mathcal{R}_{\text{BP, MF}}$ is indeed a valid set of regions and associated counting numbers.

The valid set of regions and associated counting numbers in (19) gives the region-based free energy approximation

$$\begin{aligned} F_{\text{BP, MF}} = &\sum_{a\in\mathcal{A}_{\text{BP}}} \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) \ln \frac{b_a(\mathbf{x}_a)}{f_a(\mathbf{x}_a)} \\ &- \sum_{a\in\mathcal{A}_{\text{MF}}} \sum_{\mathbf{x}_a} \prod_{i\in\mathcal{N}(a)} b_i(x_i) \ln f_a(\mathbf{x}_a) \\ &- \sum_{i\in\mathcal{I}} (|\mathcal{N}_{\text{BP}}(i)| - 1) \sum_{x_i} b_i(x_i) \ln b_i(x_i) \end{aligned} \tag{20}$$

with $F_{\text{BP, MF}} \triangleq F_{\mathcal{R}_{\text{BP, MF}}}$. In (20), we have already plugged in the factorization constraint

$$b_{\text{MF}}(\mathbf{x}_{\text{MF}}) = \prod_{i\in\mathcal{I}_{\text{MF}}} b_i(x_i)$$

with $\mathbf{x}_{\text{MF}} \triangleq (x_i \mid i \in \mathcal{I}_{\text{MF}})^{\text{T}}$ and $b_{\text{MF}} \triangleq b_{R_{\text{MF}}}$. The beliefs b_i $(i \in \mathcal{I})$ and b_a $(a \in \mathcal{A}_{\text{BP}})$ have to fulfill the normalization constraints

$$\begin{aligned} \sum_{x_i} b_i(x_i) = 1, &\quad \forall\ i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}} \\ \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) = 1, &\quad \forall\ a \in \mathcal{A}_{\text{BP}} \end{aligned} \tag{21}$$

and the marginalization constraints

$$b_i(x_i) = \sum_{\mathbf{x}_a \setminus x_i} b_a(\mathbf{x}_a), \quad \forall\, a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a). \tag{22}$$

Remark 3.1: Note that there is no need to introduce normalization constraints for the beliefs b_i $(i \in \mathcal{I}_{\text{BP}})$. If $a \in \mathcal{N}_{\text{BP}}(i)$, then it follows from the marginalization and normalization constraint for the belief b_a that

$$\begin{aligned} 1 &= \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) \\ &= \sum_{x_i} \Big(\sum_{\mathbf{x}_a \setminus x_i} b_a(\mathbf{x}_a) \Big) \\ &= \sum_{x_i} b_i(x_i). \end{aligned}$$

We will show in Lemma 2 that the region-based free energy approximation in (20) fulfilling the constraints (21) and (22) is a finite quantity, i.e., that $-\infty < F_{\text{BP, MF}} < \infty$.

The constraints (21) and (22) can be included in the Lagrangian [27, Sec. 3.1.3]

$$\begin{aligned} L_{\text{BP, MF}} \triangleq\; & F_{\text{BP, MF}} \\ & - \sum_{a \in \mathcal{A}_{\text{BP}}} \sum_{i \in \mathcal{N}(a)} \sum_{x_i} \lambda_{a,i}(x_i) \Big(b_i(x_i) - \sum_{\mathbf{x}_a \setminus x_i} b_a(\mathbf{x}_a) \Big) \\ & - \sum_{i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}} \gamma_i \Big(\sum_{x_i} b_i(x_i) - 1 \Big) \\ & - \sum_{a \in \mathcal{A}_{\text{BP}}} \gamma_a \Big(\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) - 1 \Big). \end{aligned} \tag{23}$$

The stationary points of the Lagrangian $L_{\text{BP, MF}}$ in (23) are then obtained by setting the derivatives of $L_{\text{BP, MF}}$ with respect to the beliefs and the Lagrange multipliers equal to zero. The following theorem relates the stationary points of the Lagrangian $L_{\text{BP, MF}}$ to solutions of fixed-point equations for the beliefs.

Theorem 2: Stationary points of the Lagrangian in (23) in the combined BP–MF approach must be fixed-points with positive beliefs fulfilling

$$\begin{cases} b_a(\mathbf{x}_a) = z_a\, f_a(\mathbf{x}_a) \prod\limits_{i \in \mathcal{N}(a)} n_{i \to a}(x_i), \quad \forall\, a \in \mathcal{A}_{\text{BP}} \\ b_i(x_i) = z_i \prod\limits_{a \in \mathcal{N}_{\text{BP}}(i)} m^{\text{BP}}_{a \to i}(x_i) \prod\limits_{a \in \mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{a \to i}(x_i), \quad \forall\, i \in \mathcal{I} \end{cases} \tag{24}$$

with

$$\begin{cases} n_{i \to a}(x_i) = z_i \prod\limits_{c \in \mathcal{N}_{\text{BP}}(i) \setminus a} m^{\text{BP}}_{c \to i}(x_i) \prod\limits_{c \in \mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{c \to i}(x_i), \\ \qquad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \\ m^{\text{BP}}_{a \to i}(x_i) = z_a \sum\limits_{\mathbf{x}_a \setminus x_i} f_a(\mathbf{x}_a) \prod\limits_{j \in \mathcal{N}(a) \setminus i} n_{j \to a}(x_j), \\ \qquad \forall\, a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a) \\ m^{\text{MF}}_{a \to i}(x_i) = \exp\Bigg(\sum\limits_{\mathbf{x}_a \setminus x_i} \prod\limits_{j \in \mathcal{N}(a) \setminus i} n_{j \to a}(x_j) \ln f_a(\mathbf{x}_a) \Bigg), \\ \qquad \forall\, a \in \mathcal{A}_{\text{MF}}, i \in \mathcal{N}(a) \end{cases} \tag{25}$$

and vice versa. Here, z_i $(i \in \mathcal{I})$ and z_a $(a \in \mathcal{A}_{\text{BP}})$ are positive constants that ensure that the beliefs b_i $(i \in \mathcal{I})$ and b_a $(a \in \mathcal{A})$ are normalized to one with $z_i = 1\ \forall\, i \in \mathcal{I}_{\text{BP}}$.

Proof: See Appendix C. ∎

Remark 3.2: Note that for each $i \in \mathcal{I} \setminus \mathcal{I}_{\text{BP}}$ Theorem 2 can be generalized to the case where X_i is a continuous random variable following the derivation presented in Appendix B. Formally, each sum over x_i with $i \in \mathcal{I} \setminus \mathcal{I}_{\text{BP}}$ in the third identity in (25) has to be replaced by a Lebesgue integral whenever the corresponding random variable X_i is continuous.

Remark 3.3: Note that Theorem 2 clearly states whether "extrinsic" values or "APPs" should be passed. In fact, the first equation in (25) implies that each message $n_{i \to a}(x_i)$ $(a \in \mathcal{A}, i \in \mathcal{I})$ is an "extrinsic" value when $a \in \mathcal{A}_{\text{BP}}$ and an "APP" when $a \in \mathcal{A}_{\text{MF}}$.

A. Hard constraints for BP

Some suggestions on how to generalize Theorem 1 ([9, Th. 2]) to hard constraints, i.e., to the case where the factors of the pmf $p_{\mathbf{X}}$ are not restricted to be strictly positive real-valued functions, can be found in [9, Sec. VI.D]. An example of hard constraints are deterministic functions like, e.g., code constraints. However, the statements formulated there are only conjectures and based on the assumption that we can always compute the derivative of the Lagrange function with respect to the beliefs. This is not always possible because

$$\frac{\partial F_{\text{BP}}}{\partial b_a(\mathbf{x}_a)} \to \infty \text{ as } f(\mathbf{x}_a) \to 0$$

with F_{BP} from (3). In the sequel, we show how to generalize Theorem 2 to the case where $f_a \geq 0\ \forall\, a \in \mathcal{A}_{\text{BP}}$ based on the simple observation that we are interested in solutions where the region-based free energy approximation is not plus infinity (recall that we want to minimize this quantity). As a byproduct, this also yields an extension of Theorem 1 ([9, Th. 2]) to hard constraints by simply setting $\mathcal{A}_{\text{MF}} = \emptyset$.

Lemma 2: Suppose that

$$f_a \geq 0, \quad \forall\, a \in \mathcal{A}_{\text{BP}}, \tag{26}$$

$$f_a > 0, \quad \forall\, a \in \mathcal{A}_{\text{MF}}, \tag{27}$$

and $p_{\mathbf{X}}\,|_{\bar{x}_i} \not\equiv 0$ for all $i \in \mathcal{I}$ and each realization $\bar{x}_i$ of X_i.[8] Furthermore, we assume that b_i $(i \in \mathcal{I})$ and b_a $(a \in \mathcal{A}_{\text{BP}})$ fulfill the constraints (21) and (22). Then

1) $F_{\text{BP,MF}} > -\infty$;
2) The condition
$$b_a(\bar{\mathbf{x}}_a) = 0, \quad \forall\, \bar{\mathbf{x}}_a \text{ with } a \in \mathcal{A}_{\text{BP}}, f_a(\bar{\mathbf{x}}_a) = 0 \tag{28}$$
is necessary and sufficient for $F_{\text{BP,MF}} < \infty$;
3) If (28) is fulfilled, the remaining stationary points $b_i(x_i)$ $(i \in \mathcal{I})$ and $b_a(\mathbf{x}_a)$ excluding all $\bar{\mathbf{x}}_a$ from (28) $(a \in \mathcal{A}_{\text{BP}})$ of the Lagrangian in (23) are positive beliefs fulfilling (24) and (25) excluding all $\bar{\mathbf{x}}_a$ from (28) and vice versa.
4) Moreover, (24) and (25) hold for all realizations $\bar{\mathbf{x}}_a$ (including all $\bar{\mathbf{x}}_a$ from (28)) and, therefore, (24) contains (28) as a special case.

[8] If $p_{\mathbf{X}}\,|_{\bar{x}_i} \equiv 0$ then we can simply remove this relization $\bar{x}_i$ of X_i.

Proof: See Appendix D. ∎

Remark 3.4: At first sight it seems to be a contradiction to the marginalization constraints (22) that (28) holds and all the beliefs b_i ($i \in \mathcal{I}_{\text{BP}}$) are strictly positive functions. To illustrate that this is indeed the case, let $i \in \mathcal{I}_{\text{BP}}$, $a \in \mathcal{N}_{\text{BP}}(i)$, and fix one realization $\bar{x}_i$ of X_i. Since $p_{\mathbf{X}} \mid_{\bar{x}_i} \not\equiv 0$ we also have $f_a \mid_{\bar{x}_i} \not\equiv 0$. This implies that $f_a(\bar{\mathbf{x}}_a) \neq 0$ for at least one realization $\bar{\mathbf{x}}_a = (\bar{x}_j \mid j \in \mathcal{N}(a))^{\text{T}}$ with $i \in \mathcal{N}(a)$ and, therefore, $b_a(\bar{\mathbf{x}}_a) \neq 0$. The marginalization constraints (22) together with the fact that the belief b_a must be a nonnegative function then implies that we have indeed $b_i(\bar{x}_i) > 0$.

B. Convergence and main algorithm

If the BP part has no cycles and

$$|\mathcal{N}(a) \cap \mathcal{I}_{\text{BP}}| \leq 1, \quad \forall a \in \mathcal{A}_{\text{MF}}, \tag{29}$$

then there exists a convergent implementation of the combined message passing equations in (25). In fact, we can iterate between updating the beliefs b_i with $i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$ and the forward backward algorithm in the BP part, as outlined in the following Algorithm.

Algorithm 1: If the BP part has no cycle and (29) is fulfilled, the following implementation of the fixed-point equations in (25) is guaranteed to converge.

1) Initialize b_i for all $i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$ and send the corresponding messages $n_{i\to a}(x_i) = b_i(x_i)$ to all factor nodes $a \in \mathcal{N}_{\text{MF}}(i)$.
2) Use all messages $m^{\text{MF}}_{a\to i}(x_i)$ with $i \in \mathcal{I}_{\text{BP}} \cap \mathcal{I}_{\text{MF}}$ and $a \in \mathcal{N}_{\text{MF}}(i)$ as fixed input for the BP part and run the forward/backward algorithm [10]. The fact that the resulting beliefs b_i with $i \in \mathcal{I}_{\text{BP}}$ cannot increase the region-based free energy approximation in (20) is proved in Appendix E.
3) For each $i \in \mathcal{I}_{\text{MF}} \cap \mathcal{I}_{\text{BP}}$ and $a \in \mathcal{N}_{\text{MF}}(i)$ the message $n_{i\to a}(x_i)$ is now available and can be used for further updates in the MF part.
4) For each $i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$ successively recompute the message $n_{i\to a}(x_i)$ and send it to all $a \in \mathcal{N}_{\text{MF}}(i)$. Note that for all indices $i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$
$$\frac{\partial^2 F_{\text{BP, MF}}}{\partial b_i(x_i)^2} = \frac{1}{b_i(x_i)} > 0,$$
which implies that for each index $i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$ we are solving a convex optimization problem. Therefore, the region-based free energy approximation in (20) cannot increase.
5) Proceed as described in 2).

Remark 3.5: If the factor graph representing the BP part has cycles then Algorithm 1 can be modified by running loopy BP in step 2). However, in this case the algorithm is not guaranteed to converge.

IV. Application to iterative channel estimation and decoding

In this section, we present an example where we show how to compute the updates of the messages in (25) based on Algorithm 1. We choose a simple channel model where the updates of the messages are simple enough in order to avoid overstressed notation. A class of more complex MIMO-OFDM receiver architectures together with numerical simulations can be found in [23]. In our example, we will use BP for modulation and decoding and the MF approximation for estimating the parameters of the a posteriori distribution of the channel gains. This splitting is convenient because BP works well with hard constraints and the MF approximation yields very simple message passing update equations due to the fact that the MF part in our example is a conjugate-exponential model [5]. Applying BP to all factor nodes would be intractable because the complexity is too high, cf. the discussion in Subsection IV-C.

Specifically, we consider an OFDM system with $M+N$ active subcarriers. We denote by $\mathcal{D} \subset [1:M+N]$ and $\mathcal{P} \subset [1:M+N]$ the sets of subcarrier indices for the data and pilot symbols, respectively with $|\mathcal{P}| = M$, $|\mathcal{D}| = N$, and $\mathcal{P} \cap \mathcal{D} = \emptyset$. After removing the cyclic prefix we get the following input-output relationship in the frequency domain:

$$\begin{aligned} \mathbf{Y}_{\text{D}} &= \mathbf{H}_{\text{D}} \odot \mathbf{X}_{\text{D}} + \mathbf{Z}_{\text{D}} \\ \mathbf{Y}_{\text{P}} &= \mathbf{H}_{\text{P}} \odot \mathbf{x}_{\text{P}} + \mathbf{Z}_{\text{P}} \end{aligned} \tag{30}$$

where $\mathbf{X}_{\text{D}} \triangleq (X_i \mid i \in \mathcal{D})^{\text{T}}$ is the random vector corresponding to the transmitted data symbols, $\mathbf{x}_{\text{P}} \triangleq (x_i \mid i \in \mathcal{P})^{\text{T}}$ is the vector containing the transmitted pilot symbols, and $\mathbf{H}_{\text{D}} \triangleq (H_i \mid i \in \mathcal{D})^{\text{T}}$ and $\mathbf{H}_{\text{P}} \triangleq (H_i \mid i \in \mathcal{P})^{\text{T}}$ are random vectors representing the multiplicative action of the channel while $\mathbf{Z}_{\text{D}} \triangleq (Z_i \mid i \in \mathcal{D})^{\text{T}}$ and $\mathbf{Z}_{\text{P}} \triangleq (Z_i \mid i \in \mathcal{P})^{\text{T}}$ are random vectors representing additive noise with $p_{\mathbf{Z}}(\mathbf{z}) = \text{CN}(\mathbf{z}; \mathbf{0}, \gamma^{-1}\mathbf{I}_{M+N})$ and $\mathbf{Z} \triangleq (Z_i \mid i \in \mathcal{D} \cup \mathcal{P})^{\text{T}}$. Note that (30) is very general and can also be used to model, e.g., a time-varying frequency-flat channel. In the transmitter, a random vector $\mathbf{U} = (U_i \mid i \in [1:K])$ representing the information bits is encoded and interleaved using a rate $R = K/LN$ encoder and a random interleaver, respectively into the random vector

$$\mathbf{C} = \left(\mathbf{C}^{(1)\text{T}}, \ldots, \mathbf{C}^{(N)\text{T}}\right)^{\text{T}}$$

with length LN representing the coded and interleaved bits. Each random subvector $\mathbf{C}^{(n)}$ with length L is then mapped, i.e., modulated, to X_{i_n} with $i_n \in \mathcal{D}$ ($n \in [1:N]$).

Setting $\mathbf{Y} \triangleq (Y_i \mid i \in \mathcal{D} \cup \mathcal{P})^{\text{T}}$ and $\mathbf{H} \triangleq (H_i \mid i \in \mathcal{D} \cup \mathcal{P})^{\text{T}}$, the pdf $p_{\mathbf{Y},\mathbf{X}_{\text{D}},\mathbf{H},\mathbf{C},\mathbf{U}}$ admits the factorization

$$\begin{aligned} & p_{\mathbf{Y},\mathbf{X}_{\text{D}},\mathbf{H},\mathbf{C},\mathbf{U}}(\mathbf{y}, \mathbf{x}_{\text{D}}, \mathbf{h}, \mathbf{c}, \mathbf{u}) \\ &= p_{\mathbf{Y}|\mathbf{X}_{\text{D}},\mathbf{H}}(\mathbf{y}|\mathbf{x}_{\text{D}}, \mathbf{h})\, p_{\mathbf{H}}(\mathbf{h})\, p_{\mathbf{X}_{\text{D}}|\mathbf{C}}(\mathbf{x}_{\text{D}}|\mathbf{c})\, p_{\mathbf{C}|\mathbf{U}}(\mathbf{c}|\mathbf{u})\, p_{\mathbf{U}}(\mathbf{u}) \\ &= \prod_{i\in\mathcal{D}} p_{Y_i|X_i,H_i}(y_i|x_i, h_i) \prod_{j\in\mathcal{P}} p_{Y_j|H_j}(y_j|h_j) \times p_{\mathbf{H}}(\mathbf{h}) \\ &\quad \times \prod_{n\in[1:N]} p_{X_{i_n}|\mathbf{C}^{(n)}}\left(x_{i_n}|\mathbf{c}^{(n)}\right) \times p_{\mathbf{C}|\mathbf{U}}(\mathbf{c}|\mathbf{u}) \\ &\quad \times \prod_{k\in[1:K]} p_{U_k}(u_k) \end{aligned} \tag{31}$$

where we used the fact that $\mathbf{H}$ is independent on $\mathbf{X}_{\text{D}}$, $\mathbf{C}$, and $\mathbf{U}$ and $\mathbf{Y}$ is independent on $\mathbf{C}$ and $\mathbf{U}$ conditioned on $\mathbf{X}_{\text{D}}$.

Note that

$$p_{Y_i|X_i,H_i}(y_i|x_i,h_i)(y_i) = \frac{\gamma}{\pi}\exp(-\gamma|y_i - h_i x_i|^2) = \mathrm{CN}(y_i; h_i x_i, 1/\gamma), \quad \forall i \in \mathcal{D} \quad (32)$$

$$p_{Y_i|H_i}(y_i|h_i)(y_i) = \frac{\gamma}{\pi}\exp(-\gamma|y_i - h_i x_i|^2) = \mathrm{CN}(y_i; h_i x_i, 1/\gamma), \quad \forall i \in \mathcal{P}. \quad (33)$$

We choose for the prior distribution of $\mathbf{H}$

$$p_{\mathbf{H}}(\mathbf{h}) = \mathrm{CN}(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{H}}^{\mathrm{P}}, {\boldsymbol{\Lambda}_{\mathbf{H}}^{\mathrm{P}}}^{-1}).$$

Now define

$$\mathcal{I} \triangleq \{X_i \mid i \in \mathcal{D}\} \cup \{\mathbf{H}\} \cup \{C_1^{(1)}, \ldots, C_L^{(N)}\} \cup \{U_1, \ldots, U_K\} \quad (34)$$

$$\mathcal{A} \triangleq \{p_{Y_i|X_i,H_i} \mid i \in \mathcal{D}\} \cup \{p_{Y_i|H_i} \mid i \in \mathcal{P}\} \cup \{p_{\mathbf{H}}\} \cup \{p_{X_{i_n}|\mathbf{C}^{(n)}} \mid n \in [1:N]\} \cup \{p_{\mathbf{C}|\mathbf{U}}\} \cup \{p_{U_k} \mid k \in [1:K]\} \quad (35)$$

and set $f_a \triangleq a$ for all $a \in \mathcal{A}$. For example, we have $f_{p_{\mathbf{H}}}(\mathbf{h}) = p_{\mathbf{H}}(\mathbf{h})$. We choose a splitting of $\mathcal{A}$ into $\mathcal{A}_{\mathrm{BP}}$ and $\mathcal{A}_{\mathrm{MF}}$ with

$$\mathcal{A}_{\mathrm{BP}} \triangleq \{p_{X_{i_n}|\mathbf{C}^{(n)}} \mid n \in [1:N]\} \cup \{p_{\mathbf{C}|\mathbf{U}}\} \cup \{p_{U_k} \mid k \in [1:K]\} \quad (36)$$

$$\mathcal{A}_{\mathrm{MF}} \triangleq \{p_{Y_i|X_i,H_i} \mid i \in \mathcal{D}\} \cup \{p_{Y_i|H_i} \mid i \in \mathcal{P}\} \cup \{p_{\mathbf{H}}\}.$$

With this selection

$$\mathcal{I}_{\mathrm{BP}} = \{X_i \mid i \in \mathcal{D}\} \cup \{C_1^{(1)}, \ldots, C_L^{(N)}\} \cup \{U_1, \ldots, U_K\} \quad (37)$$

$$\mathcal{I}_{\mathrm{MF}} = \{X_i \mid i \in \mathcal{D}\} \cup \{\mathbf{H}\},$$

which implies that $\mathcal{I}_{\mathrm{BP}} \cap \mathcal{I}_{\mathrm{MF}} = \{X_i \mid i \in \mathcal{D}\}$. The factor graph corresponding to the factorization in (31) with the splitting of $\mathcal{A}$ into $\mathcal{A}_{\mathrm{MF}}$ and $\mathcal{A}_{\mathrm{BP}}$ as in (36) is depicted in Figure 1.

We now show how to apply the variant of Algorithm 1 referred to in Remark 3.5 to the factor graph depicted in Figure 1. Note that (29) is fulfilled in this example; however, cycles occur in the BP part of the factor graph due to the combination of high-order modulation and (convolutional) coding (see Table I).

Algorithm 2: 1) Initialize

$$b_{\mathbf{H}}(\mathbf{h}) = \mathrm{CN}(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{H}}, \boldsymbol{\Lambda}_{\mathbf{H}}^{-1})$$

and set

$$n_{\mathbf{H}\to p_{Y_i|X_i,H_i}}(\mathbf{h}) = b_{\mathbf{H}}(\mathbf{h}), \quad \forall\, i \in \mathcal{D}.$$

2) Using the particular form of the distributions $p_{Y_i|X_i,H_i}$ $(i \in \mathcal{D})$ in (32) and $p_{Y_i|H_i}$ $(i \in \mathcal{P})$ in (33), compute

$$m^{\mathrm{MF}}_{p_{Y_i|X_i,H_i}\to X_i}(x_i) \propto \exp\left(-\gamma \int \mathrm{d}\mathbf{h}\, n_{\mathbf{H}\to p_{Y_i|X_i,H_i}}(\mathbf{h})|y_i - h_i x_i|^2\right) \propto \exp\left(-\gamma(\sigma_{H_i}^2 + |\mu_{H_i}|^2)\left|x_i - \frac{y_i \mu_{H_i}^*}{\sigma_{H_i}^2 + |\mu_{H_i}|^2}\right|^2\right) \propto \mathrm{CN}\left(x_i; \frac{y_i \mu_{H_i}^*}{\sigma_{H_i}^2 + |\mu_{H_i}|^2}, \frac{1}{\gamma(\sigma_{H_i}^2 + |\mu_{H_i}|^2)}\right), \quad \forall\, i \in \mathcal{D}$$

with $\sigma_{H_i}^2 \triangleq [\boldsymbol{\Lambda}_{\mathbf{H}}^{-1}]_{i,i}$ $(i \in \mathcal{D})$.

3) Use the messages $m^{\mathrm{MF}}_{p_{Y_i|X_i,H_i}\to X_i}(x_i)$ $(i \in \mathcal{D})$ as fixed input for the BP part and run BP.

4) After running BP in the BP part, compute the messages $n_{X_i\to p_{Y_i|X_i,H_i}}(x_i)$ $(i \in \mathcal{D})$ and update the messages in the MF part. Namely, after setting

$$\mu_{X_i} \triangleq \sum_{x_i} n_{X_i\to p_{Y_i|X_i,H_i}}(x_i) x_i$$

$$\sigma_{X_i}^2 \triangleq \sum_{x_i} n_{X_i\to p_{Y_i|X_i,H_i}}(x_i) |x_i - \mu_{X_i}|^2$$

for all $i \in \mathcal{D}$ compute the messages

$$m^{\mathrm{MF}}_{p_{Y_i|X_i,H_i}\to \mathbf{H}}(h_i) \propto \exp\left(-\gamma \sum_{x_i} n_{X_i\to p_{Y_i|X_i,H_i}}(x_i)|y_i - h_i x_i|^2\right) \propto \exp\left(-\gamma(\sigma_{X_i}^2 + |\mu_{X_i}|^2)\left|h_i - \frac{y_i \mu_{X_i}^*}{\sigma_{X_i}^2 + |\mu_{X_i}|^2}\right|^2\right) \propto \mathrm{CN}\left(h_i; \frac{y_i \mu_{X_i}^*}{\sigma_{X_i}^2 + |\mu_{X_i}|^2}, \frac{1}{\gamma(\sigma_{X_i}^2 + |\mu_{X_i}|^2)}\right), \quad \forall\, i \in \mathcal{D}$$

$$m^{\mathrm{MF}}_{p_{Y_i|H_i}\to \mathbf{H}}(h_i) \propto \exp(-\gamma|y_i - h_i x_i|^2) \propto \mathrm{CN}\left(h_i; \frac{y_i x_i^*}{|x_i|^2}, \frac{1}{\gamma|x_i|^2)}\right), \quad \forall\, i \in \mathcal{P}$$

$$m^{\mathrm{MF}}_{p_{\mathbf{H}}\to \mathbf{H}}(\mathbf{h}) = \mathrm{CN}(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{H}}^{\mathrm{P}}, {\boldsymbol{\Lambda}_{\mathbf{H}}^{\mathrm{P}}}^{-1})$$

and

$$n_{\mathbf{H}\to p_{Y_i|X_i,H_i}}(\mathbf{h}) = z_{\mathbf{H}} \prod_{i\in\mathcal{D}} m^{\mathrm{MF}}_{p_{Y_i|X_i,H_i}\to \mathbf{H}}(h_i) \prod_{j\in\mathcal{P}} m^{\mathrm{MF}}_{p_{Y_i|H_i}\to \mathbf{H}}(h_j) m^{\mathrm{MF}}_{p_{\mathbf{H}}\to \mathbf{H}}(\mathbf{h}) = \frac{\det(\boldsymbol{\Lambda}_{\mathbf{H}})}{\pi^{M+N}} \exp\Big(-(\mathbf{h} - \mu_{\mathbf{H}})^{\mathrm{H}} \boldsymbol{\Lambda}_{\mathbf{H}} (\mathbf{h} - \mu_{\mathbf{H}})\Big) = \mathrm{CN}(\mathbf{h}; \mu_{\mathbf{H}}, \boldsymbol{\Lambda}_{\mathbf{H}}^{-1}), \quad \forall\, i \in \mathcal{D}.$$

Here, we used Lemma 3 in Appendix F to get the updated parameters

$$\boldsymbol{\mu}_{\mathbf{H}} = \boldsymbol{\Lambda}_{\mathbf{H}}^{-1}(\boldsymbol{\Lambda}_{\mathbf{H}}^{\mathrm{P}} \boldsymbol{\mu}_{\mathbf{H}}^{\mathrm{P}} + \widetilde{\boldsymbol{\Lambda}}_{\mathbf{H}} \widetilde{\boldsymbol{\mu}}_{\mathbf{H}}) \qquad \boldsymbol{\Lambda}_{\mathbf{H}} = \boldsymbol{\Lambda}_{\mathbf{H}}^{\mathrm{P}} + \widetilde{\boldsymbol{\Lambda}}_{\mathbf{H}} \quad (38)$$

with

$$\widetilde{\lambda}_{\mathbf{H}_{ij}} = \begin{cases} \gamma(\sigma_{X_i}^2 + |\mu_{X_i}|^2) & \text{if } i = j \in \mathcal{D} \\ \gamma|x_i|^2 & \text{if } i = j \in \mathcal{P} \\ 0 & \text{else} \end{cases}$$

and

$$\widetilde{\lambda}_{\mathbf{H}_{ii}} \widetilde{\mu}_{\mathbf{H}_i} = \begin{cases} \gamma y_i \mu_{X_i}^* & \text{if } i \in \mathcal{D} \\ \gamma y_i x_i^* & \text{if } i \in \mathcal{P}. \end{cases}$$

The update for the belief $b_{\mathbf{H}}$ is

$$b_{\mathbf{H}}(\mathbf{h}) = n_{\mathbf{H}\to p_{Y_i|X_i,H_i}}(\mathbf{h}),$$

i.e., $b_{\mathbf{H}}(\mathbf{h}) = \mathrm{CN}(\mathbf{h}; \mu_{\mathbf{H}}, \boldsymbol{\Lambda}_{\mathbf{H}}^{-1})$.

5) Proceed as described in 2).

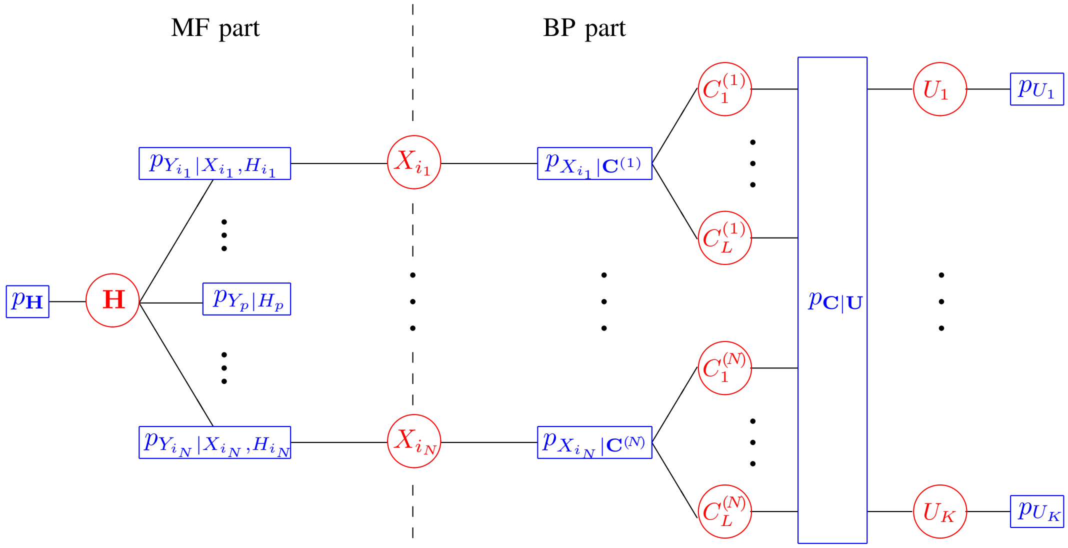

Fig. 1. Factor graph corresponding to the factorization of the pdf in (31) with $\mathcal{D} = \{i_1, \ldots, i_N\}$ and $p \in \mathcal{P}$. The splitting of the factor graph into BP and MF parts is chosen in such a way that utilizes most of the advantages of BP and the MF approximation.

A. *"Extrinsic" values versus "APP"*

In consideration of Remark 3.3 it is instructive to analyze the messages coming from the variable nodes $\mathcal{I}_{\text{BP}} \cap \mathcal{I}_{\text{MF}} = \{X_1, \ldots, X_N\}$, which are contained in the BP and MF parts of the factor graph depicted in Figure 1. Whether a message passing from a variable node to a factor node is an "extrinsic" value or an "APP" depends on whether the corresponding factor node is in the BP or the MF part. Thus, the messages

$$\begin{aligned} &n_{X_{i_n} \to p_{X_{i_n}|\mathbf{C}^{(n)}}}(x_{i_n}) \\ &= m^{\text{MF}}_{p_{Y_{i_n}|X_{i_n},H_{i_n}} \to X_{i_n}}(x_{i_n}), \quad \forall\, n \in [1:N], \end{aligned}$$

which are passed into the BP part, are "extrinsic" values, whereas the messages

$$\begin{aligned} &n_{X_{i_n} \to p_{Y_{i_n}|X_{i_n},H_{i_n}}}(x_{i_n}) \\ &= m^{\text{BP}}_{p_{X_{i_n}|\mathbf{C}^{(n)}} \to X_{i_n}}(x_{i_n})\; m^{\text{MF}}_{p_{Y_{i_n}|X_{i_n},H_{i_n}} \to X_{i_n}}(x_{i_n}), \\ &\quad \forall\, n \in [1:N], \end{aligned}$$

which are passed into the MF part, are "APPs". Note that this result is aligned with the strategies proposed in [19], [20] where "APPs" are used for channel estimation and "extrinsic values" for detection.

B. *Level of MF approximation*

Note that there is an ambiguity in the choice of variable nodes in the MF part. This ambiguity reflects the "level of the MF approximation" and results in a family of different algorithms. For example, instead of choosing $\mathbf{H}$ as a single random variable, we could have chosen H_i $(i \in [1:M+N])$ to be separate variable nodes in the factor graph. In this case we make the assumption that the random variables H_i $(i \in [1:M+N])$ are independent and the set of indices $\mathcal{I}$ in (34) has to be replaced by

$$\begin{aligned} \mathcal{I} \triangleq &\{X_i \mid i \in \mathcal{D}\} \cup \{H_i \mid i \in \mathcal{D} \cup \mathcal{P}\} \\ &\cup \{C_1^{(1)}, \ldots, C_L^{(N)}\} \cup \{U_1, \ldots, U_K\}. \end{aligned}$$

Since this is an additional approximation, the performance of the receiver is a nonincreasing function of the level of MF approximation. However, it is possible that the complexity reduces by applying an additional MF approximation. See [23, Sec. IV-B] for further discussions on this ambiguity for a class of MIMO-OFDM receivers.

C. *Comparison with BP combined with Gaussian approximation*

The example makes evident how the complexity of the message-passing algorithm can be simplified by exploiting the conjugate-exponential property of the MF part, which leads to simple update equations of the belief $b_{\mathbf{H}}$. In fact, at each iteration in the algorithm we only have to update the parameters of a Gaussian distribution (38). In comparison let us consider an alternative split of $\mathcal{A}$ by moving the factor nodes $p_{Y_i|X_i,H_i}$ $(i \in \mathcal{D})$ in (32) and $p_{Y_i|H_i}$ $(i \in \mathcal{P})$ in (33) to the BP part. This is equivalent to applying BP to the whole factor graph in Figure 1 because $m^{\text{MF}}_{p_{\mathbf{H}} \to \mathbf{H}} = m^{\text{BP}}_{p_{\mathbf{H}} \to \mathbf{H}}$. Doing so, each message $m^{\text{BP}}_{p_{Y_i|X_i,H_i} \to \mathbf{H}}(h_i)$ $(i \in \mathcal{D})$ does no longer admit a closed form expression in terms of the mean and the variance of the random variable X_i and becomes a mixture of Gaussian pdfs with 2^L components; in consequence, each message $n_{\mathbf{H} \to p_{Y_i|X_i,H_i}}(\mathbf{h})$ $(i \in \mathcal{D})$ becomes a sum of $2^{L(N-1)}$ terms. To keep the complexity of computing these messages tractable one has to rely on additional approximations.

As suggested in [33], [34], we can approximate each message $m^{\text{BP}}_{p_{Y_i|X_i,H_i} \to \mathbf{H}}(h_i)$ $(i \in \mathcal{D})$ by a Gaussian pdf. BP combined with this approximation is comparable in terms of complexity to Algorithm 2, since the computations of the updates of the messages are equally complex. However, Algorithm 2 clearly outperforms this alternative, as can be seen in Figure 2. It can also be noticed that the performance of Algorithm 2 is close to the case with perfect channel state information (CSI) at the receiver, even with a low density of pilots, i.e., such that the spacing between any two consecutive pilots (Δ_P) approximately equals the coherence bandwidth[9] (W_{coh}) of the channel or twice of it.

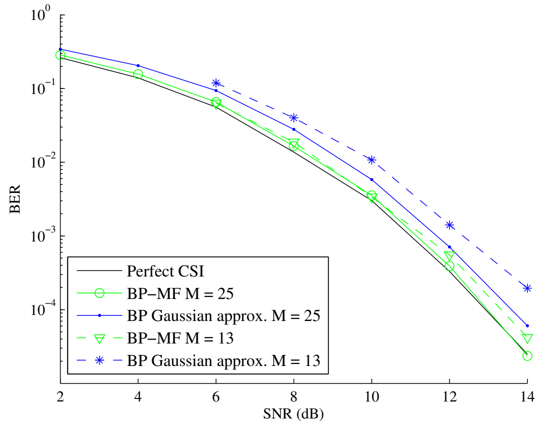

Fig. 2. Bit error rate (BER) as a function of signal-to-noise ratio (SNR) for Algorithm 2 (BP–MF), BP combined with Gaussian approximation as described in Subsection IV-C, and perfect CSI at the receiver. Pilot spacing $\Delta_P \approx W_{\text{coh}}$ $(M = 25)$ and $\Delta_P \approx 2W_{\text{coh}}$ $(M = 13)$.

TABLE I
PARAMETERS OF THE OFDM SYSTEM.

Number of subcarriers	$M + N = 300$
Number of evenly spaced pilots	$M \in \{13, 25\}$
Modulation scheme for pilot symbols	QPSK
Modulation scheme for data symbols	16 QAM $(L = 4)$
Convolutional channel code	$R = 1/3$ $(133, 171, 165)_8$
Multipath channel model	3 GPP ETU
Subcarrier spacing	15 kHz
Coherence bandwidth	$W_{\text{coh}} \approx 200$ kHz

D. Estimation of noise precision

Algorithm 2 can be easily extended to the case where the noise precision γ is a realization of a random variable Γ. In fact, since $\ln p_{Y_i|X_i,H_i,\Gamma}$ $(i \in \mathcal{D})$ and $\ln p_{Y_i|H_i,\Gamma}$ $(i \in \mathcal{P})$ are linear in γ, we can replace any dependence on γ in the existing messages in Algorithm 2 by the expected value of Γ and get simple expressions for the additional messages using a Gamma prior distribution for Γ, reflecting the powerfulness of exploiting the conjugate-exponential model property in the MF part for parameter estimation. See [23, Sec. IV-A] for further details on the explicit form of the additional messages.

V. CONCLUSION AND OUTLOOK

We showed that the message passing fixed-point equations of a combination of BP and the MF approximation correspond to stationary points of one single constrained region-based free energy approximation. These stationary points are in one-to-one correspondence to solutions of a coupled system of message passing fixed-point equations. For an arbitrary factor graph and a choice of a splitting of the factor nodes into a set of MF and BP factor nodes, our result gives immediately the corresponding message passing fixed-point equations and yields an interpretation of the computed beliefs as stationary points. Moreover, we presented an algorithm for updating the messages that is guaranteed to converge provided that the factor graph fulfills certain technical conditions. We also showed how to extend the MF part in the factor graph to continuous random variables and to include hard constraints in the BP part of the factor graph. Finally, we illustrated the computation of the messages of our algorithm in a simple example. This example demonstrates the efficiency of the combined scheme in models in which BP messages are computationally intractable. The proposed algorithm performs significantly better than the commonly used approach of using BP combined with a Gaussian approximation of computationally demanding messages.

An interesting extension of our result would be to generalize the BP part to contain also continuous random variables. The results in [35] provide a promising approach. Indeed, they could be used to generalize the Lagrange multiplier for the marginalization constraints to the continuous case. However, these methods are based on the assumption that the objective function is Fréchet differentiable [36, p. 172]. In general a region-base free energy approximation is neither Fréchet differentiable nor Gateaux differentiable, at least not without any modification of the definitions used in standard text books [36, pp. 171–172][10]. An extension to continuous random variables in the BP part would allow to apply a combination of BP with the MF approximation, e.g., for sensor self-localization, where both methods are used [37], [38]. Another interesting extension could be to generalize the region-based free energy approximation such that the messages in the BP part are equivalent to the messages passed in tree reweighted BP or to include second order correction terms in the MF approximation that are similar to the Onsager reaction term [29].

VI. ACKNOWLEDGMENT

The authors wish to thank Prof. Ralf R. Müller for his comments on a previous draft of this paper.

APPENDIX

A. Proof of Lemma 1

Suppose that $\{\widetilde{m}_{a\to i}(x_i), \widetilde{n}_{i\to a}(x_i)\}$ $(a \in \mathcal{A}, i \in \mathcal{N}(a))$ is a solution of (7) and set

$$\begin{aligned}\widetilde{m}_{a\to i}(x_i) &= \kappa_{a,i} m_{a\to i}(x_i), \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a)\\ \widetilde{n}_{i\to a}(x_i) &= \tau_{a,i} n_{i\to a}(x_i), \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a)\end{aligned} \tag{39}$$

with $\kappa_{a,i}, \tau_{a,i} > 0$ $(a \in \mathcal{A}, i \in \mathcal{N}(a))$. Plugging (39) into (7) we obtain the following fixed-point equations for the messages

[9] Calculated as the reciprocal of the maximum excess delay.

[10] For a positive real-valued function b, $b + \Delta b$ might fail to be a positive real-valued function for arbitrary perturbations Δb with sufficiently small norm $\|\Delta b\|$.

$\{m_{a\to i}(x_i), n_{i\to a}(x_i)\}$ $(a \in \mathcal{A}, i \in \mathcal{N}(a))$.

$$\begin{cases} \kappa_{a,i} m_{a\to i}(x_i) \\ \quad = \omega_{a,i} \Big(\prod_{j\in\mathcal{N}(a)\setminus i} \tau_{a,j} \Big) \sum_{\mathbf{x}_a \setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j) \\ \tau_{a,i} n_{i\to a}(x_i) \\ \quad = \Big(\prod_{c\in\mathcal{N}(i)\setminus a} \kappa_{c,i} \Big) \prod_{c\in\mathcal{N}(i)\setminus a} m_{c\to i}(x_i) \end{cases} \tag{40}$$

for all $a \in \mathcal{A}, i \in \mathcal{N}(a)$. Now (40) is equivalent to (6) if and only if

$$\tau_{a,i} = \prod_{c\in\mathcal{N}(i)\setminus a} \kappa_{c,i}, \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \tag{41}$$

$$z_a = \frac{\omega_{a,i} \prod_{j\in\mathcal{N}(a)\setminus i} \tau_{a,j}}{\kappa_{a,i}}, \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \tag{42}$$

where the positive constants z_a $(a \in \mathcal{A})$ are such that the beliefs b_a $(a \in \mathcal{A})$ in (5) are normalized to one. This normalization of the beliefs b_a $(a \in \mathcal{A})$ in (5) gives

$$\begin{aligned} \frac{1}{z_a} &= \sum_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)} n_{j\to a}(x_j) \\ &= \frac{\sum_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)} \widetilde{n}_{j\to a}(x_j)}{\prod_{j\in\mathcal{N}(a)} \tau_{a,j}} \\ &= \frac{1}{\tilde{z}_a \prod_{j\in\mathcal{N}(a)} \tau_{a,j}}, \quad \forall\, a \in \mathcal{A} \end{aligned} \tag{43}$$

where we used (39) in the second step and (8) in the last step. Combining (41), (42), and (43) we obtain

$$\begin{aligned} \frac{1}{\tilde{z}_a} &= \frac{\kappa_{a,i}\tau_{a,i}}{\omega_{a,i}} \\ &= \frac{g_i}{\omega_{a,i}}, \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \end{aligned}$$

with

$$g_i \triangleq \prod_{c\in\mathcal{N}(i)} \kappa_{c,i}, \quad \forall\, i \in \mathcal{I}.$$

Now suppose that (9) is fulfilled. Setting

$$\begin{aligned} \kappa_{a,i} &= g_i^{\frac{1}{|\mathcal{N}(i)|}}, \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \\ \tau_{a,i} &= g_i^{1-\frac{1}{|\mathcal{N}(i)|}} \quad \forall\, a \in \mathcal{A}, i \in \mathcal{N}(a) \end{aligned}$$

and reversing all the steps finishes the proof.

B. Extension of the MF approximation to continuous random variables

Suppose that $p_{\mathbf{X}}$ is a pdf for the vector of random variables $\mathbf{X}$. In this appendix, we assume that all integrals in the region-based free energy approximation are Lebesgue integrals and have finite values, which can be verified by inspection of the factors f_a $(a \in \mathcal{A})$ and the analytic expressions of the computed beliefs b_i $(i \in \mathcal{I})$. An example where the MF approximation is applied to continuous random variables and combined with BP is discussed in Section IV.

For each $i \in \mathcal{I}$ we can rewrite F_{MF} in (11) as

$$\begin{aligned} F_{\text{MF}} = D(b_i \,||\, a_i) &+ \sum_{j\in\mathcal{I}\setminus i} \int b_j(x_j) \ln b_j(x_j)\, \mathrm{d}x_j \\ &- \sum_{a\in\mathcal{A}\setminus\mathcal{N}(i)} \int \ln f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)} b_j(x_j)\, \mathrm{d}x_j \end{aligned} \tag{44}$$

with

$$a_i(x_i) \triangleq \exp\Big(\sum_{a\in\mathcal{N}(i)} \int \ln f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j)\, \mathrm{d}x_j \Big), \quad \forall\, i \in \mathcal{I}.$$

It follows from [22, pp. 36–38] that $D(b_i \,||\, a_i)$ is minimized subject to $\int b_i(x_i)\, \mathrm{d}x_i = 1$ if and only if

$$b_i(x_i) = \frac{a_i(x_i)}{\int a_i(x_i)\, \mathrm{d}x_i}. \tag{45}$$

Formally, b_i in (45) differs from b_i in (12) by replacing sums with Lebesgue integrals.

C. Proof of Theorem 2

The proof of Theorem 2 is based on the ideas of the proof of [9, Th. 2]. However, we will see that we get a significant simplification by augmenting it with some of the arguments originally used in [11] for Markov random fields and adopted to factor graphs in [12]. In particular, we shall make use of the following observation. Recall the expression for $F_{\text{BP, MF}}$ in (20)

$$\begin{aligned} F_{\text{BP, MF}} = &\sum_{a\in\mathcal{A}_{\text{BP}}} \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) \ln \frac{b_a(\mathbf{x}_a)}{f_a(\mathbf{x}_a)} \\ &- \sum_{a\in\mathcal{A}_{\text{MF}}} \sum_{\mathbf{x}_a} \prod_{i\in\mathcal{N}(a)} b_i(x_i) \ln f_a(\mathbf{x}_a) \\ &- \sum_{i\in\mathcal{I}} (|\mathcal{N}_{\text{BP}}(i)| - 1) \sum_{x_i} b_i(x_i) \ln b_i(x_i), \end{aligned} \tag{46}$$

the marginalization constraints

$$b_i(x_i) = \sum_{\mathbf{x}_a \setminus x_i} b_a(\mathbf{x}_a), \quad \forall\, a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a), \tag{47}$$

and the normalization constraints

$$\begin{aligned} &\sum_{x_i} b_i(x_i) = 1, \quad \forall\, i \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}} \\ &\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) = 1, \quad \forall\, a \in \mathcal{A}_{\text{BP}}. \end{aligned} \tag{48}$$

Using the marginalization constraints (47), we see that

$$\begin{aligned}
&\sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln\prod_{i\in\mathcal{N}(a)} b_i(x_i)\\
&=\sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{\mathbf{x}_a}\sum_{i\in\mathcal{N}(a)} b_a(\mathbf{x}_a)\ln b_i(x_i)\\
&=\sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{i\in\mathcal{N}(a)}\sum_{x_i} b_i(x_i)\ln b_i(x_i)\\
&=\sum_{i\in\mathcal{I}_{\text{BP}}}\sum_{a\in\mathcal{N}_{\text{BP}}(i)}\sum_{x_i} b_i(x_i)\ln b_i(x_i)\\
&=\sum_{i\in\mathcal{I}_{\text{BP}}}|\mathcal{N}_{\text{BP}}(i)|\sum_{x_i} b_i(x_i)\ln b_i(x_i).
\end{aligned} \tag{49}$$

Combining (49) with (46), we further get

$$\begin{aligned}
F_{\text{BP, MF}} = &-\sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln f_a(\mathbf{x}_a)\\
&-\sum_{a\in\mathcal{A}_{\text{MF}}}\sum_{\mathbf{x}_a}\prod_{i\in\mathcal{N}(a)} b_i(x_i)\ln f_a(\mathbf{x}_a)\\
&+\sum_{i\in\mathcal{I}}\sum_{x_i} b_i(x_i)\ln b_i(x_i)\\
&+\sum_{a\in\mathcal{A}_{\text{BP}}} I_a
\end{aligned} \tag{50}$$

with the mutual information [24, p. 19]

$$I_a \triangleq \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln\frac{b_a(\mathbf{x}_a)}{\prod_{i\in\mathcal{N}(a)} b_i(x_i)},\quad \forall\ a\in\mathcal{A}_{\text{BP}}. \tag{51}$$

Next, we shall compute the stationary points of the Lagrangian

$$\begin{aligned}
L_{\text{BP, MF}} = &F_{\text{BP, MF}}\\
&-\sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{i\in\mathcal{N}(a)}\sum_{x_i}\lambda_{a,i}(x_i)\Big(b_i(x_i)-\sum_{\mathbf{x}_a\setminus x_i} b_a(\mathbf{x}_a)\Big)\\
&-\sum_{i\in\mathcal{I}_{\text{MF}}\setminus\mathcal{I}_{\text{BP}}}\gamma_i\Big(\sum_{x_i} b_i(x_i)-1\Big)\\
&-\sum_{a\in\mathcal{A}_{\text{BP}}}\gamma_a\Big(\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)-1\Big)
\end{aligned} \tag{52}$$

using the expression for $F_{\text{BP, MF}}$ in (50). The particular form of $F_{\text{BP, MF}}$ in (50) is convenient because the marginalization constraints in (47) imply that for all $i\in\mathcal{I}$ and $a\in\mathcal{A}_{\text{BP}}$ we have $\frac{\partial I_a}{\partial b_i(x_i)} = -\mathrm{I}_{\mathcal{N}_{\text{BP}}(i)}(a)$. Setting the derivative of $L_{\text{BP, MF}}$ in (52) with respect to $b_i(x_i)$ and $b_a(\mathbf{x}_a)$ equal to zero for all $i\in\mathcal{I}$ and $a\in\mathcal{A}_{\text{BP}}$, we get the following fixed-point equations for the stationary points:

$$\begin{aligned}
\ln b_i(x_i) = &\sum_{a\in\mathcal{N}_{\text{BP}}(i)}\lambda_{a,i}(x_i)\\
&+\sum_{a\in\mathcal{N}_{\text{MF}}(i)}\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j)\ln f_a(\mathbf{x}_a)\\
&+|\mathcal{N}_{\text{BP}}(i)| + \mathrm{I}_{\mathcal{I}_{\text{MF}}\setminus\mathcal{I}_{\text{BP}}}(i)\gamma_i - 1,\quad \forall\ i\in\mathcal{I}\\
\ln b_a(\mathbf{x}_a) = &\ln f_a(\mathbf{x}_a) - \sum_{i\in\mathcal{N}(a)}\lambda_{a,i}(x_i) + \ln\Big(\prod_{i\in\mathcal{N}(a)} b_i(x_i)\Big)\\
&+\gamma_a - 1,\quad \forall\ a\in\mathcal{A}_{\text{BP}}.
\end{aligned} \tag{53}$$

Setting

$$\begin{aligned}
m^{\text{BP}}_{a\to i}(x_i) &\triangleq \exp\Big(\lambda_{a,i}(x_i) + 1 - \frac{1}{|\mathcal{N}_{\text{BP}}(i)|}\Big),\\
&\qquad \forall\ a\in\mathcal{A}_{\text{BP}}, i\in\mathcal{N}(a)\\
m^{\text{MF}}_{a\to i}(x_i) &\triangleq \exp\Bigg(\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j)\ln f_a(\mathbf{x}_a)\Bigg),\\
&\qquad \forall\ a\in\mathcal{A}_{\text{MF}}, i\in\mathcal{N}(a),
\end{aligned} \tag{54}$$

we can rewrite (53) as

$$\begin{aligned}
b_i(x_i) &= z_i\prod_{a\in\mathcal{N}_{\text{BP}}(i)} m^{\text{BP}}_{a\to i}(x_i)\prod_{a\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{a\to i}(x_i),\quad \forall\ i\in\mathcal{I}\\
b_a(\mathbf{x}_a) &= z_a\, f_a(\mathbf{x}_a)\prod_{i\in\mathcal{N}(a)}\frac{b_i(x_i)}{m^{\text{BP}}_{a\to i}(x_i)},\quad \forall\ a\in\mathcal{A}_{\text{BP}}
\end{aligned} \tag{55}$$

with

$$z_i \triangleq \exp(\mathrm{I}_{\mathcal{I}_{\text{MF}}\setminus\mathcal{I}_{\text{BP}}}(i)\gamma_i),\quad \forall\ i\in\mathcal{I} \tag{56}$$

$$z_a \triangleq \exp\left(\gamma_a - 1 + \sum_{i\in\mathcal{N}(a)}\Big(1-\frac{1}{|\mathcal{N}_{\text{BP}}(i)|}\Big)\right),\quad \forall\ a\in\mathcal{A}_{\text{BP}}. \tag{57}$$

Finally, we define

$$\begin{aligned}
n_{i\to a}(x_i) \triangleq z_i &\prod_{c\in\mathcal{N}_{\text{BP}}(i)\setminus\{a\}} m^{\text{BP}}_{c\to i}(x_i)\prod_{c\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{c\to i}(x_i),\\
&\forall\ a\in\mathcal{A}, i\in\mathcal{N}(a).
\end{aligned} \tag{58}$$

Plugging the expression for $n_{i\to a}(x_i)$ in (58) into the second line in (55), we find that

$$\begin{aligned}
b_i(x_i) &= z_i\prod_{a\in\mathcal{N}_{\text{BP}}(i)} m^{\text{BP}}_{a\to i}(x_i)\prod_{a\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{a\to i}(x_i),\quad \forall\ i\in\mathcal{I}\\
b_a(\mathbf{x}_a) &= z_a\, f_a(\mathbf{x}_a)\prod_{i\in\mathcal{N}(a)} n_{i\to a}(x_i),\quad \forall\ a\in\mathcal{A}_{\text{BP}}.
\end{aligned} \tag{59}$$

Using the marginalization constraints in (47) in combination with (59) and noting that $z_i = 1$ for all $i\in\mathcal{I}_{\text{BP}}$ we further find that

$$\begin{aligned}
&n_{i\to a}(x_i)m^{\text{BP}}_{a\to i}(x_i)\\
&=\prod_{a\in\mathcal{N}_{\text{BP}}(i)} m^{\text{BP}}_{a\to i}(x_i)\prod_{a\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{a\to i}(x_i)\\
&= b_i(x_i)\\
&=\sum_{\mathbf{x}_a\setminus x_i} b_a(\mathbf{x}_a)\\
&= z_a\sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a)\prod_{j\in\mathcal{N}(a)} n_{j\to a}(x_j),\quad \forall\ a\in\mathcal{A}_{\text{BP}}, i\in\mathcal{N}(a).
\end{aligned} \tag{60}$$

Dividing both sides of (60) by $n_{i\to a}(x_i)$ gives

$$\begin{aligned}
m^{\text{BP}}_{a\to i}(x_i) = &z_a\sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a)\prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j)\\
&\forall\ a\in\mathcal{A}_{\text{BP}}, i\in\mathcal{N}(a).
\end{aligned} \tag{61}$$

Noting that $n_{j\to a}(x_j) = b_j(x_j)$ for all $a \in \mathcal{A}_{\text{MF}}$ and $j \in \mathcal{N}(a)$, we can write the messages $m^{\text{MF}}_{a\to i}(x_i)$ in (54) as

$$m^{\text{MF}}_{a\to i}(x_i) = \exp\Bigg(\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j)\ln f_a(\mathbf{x}_a)\Bigg), \quad \forall\ a \in \mathcal{A}_{\text{MF}}, i \in \mathcal{N}(a). \tag{62}$$

Now (58), (61), and (62) are equivalent to (25) and (59) is equivalent to (24). This completes the proof that stationary points of the Lagrangian in (23) must be fixed-points with positive beliefs fulfilling (24). Since all the steps are reversible, this also completes the proof of Theorem C.

D. Proof of Lemma 2

We rewrite $F_{\text{BP,MF}}$ in (20) as $F_{\text{BP,MF}} = F_1 + F_2 + F_3$ with

$$\begin{aligned}
F_1 &\triangleq \sum_{a\in\mathcal{A}_{\text{BP}}} D(b_a\ ||\ f_a)\\
F_2 &\triangleq \sum_{a\in\mathcal{A}_{\text{MF}}} D\Big(\prod_{i\in\mathcal{N}(a)} b_i\ ||\ f_a\Big)\\
F_3 &\triangleq -\sum_{i\in\mathcal{I}}(|\mathcal{N}_{\text{BP}}(i)| + |\mathcal{N}_{\text{MF}}(i)| - 1)\sum_{x_i} b_i(x_i)\ln b_i(x_i)
\end{aligned}$$

and set

$$0 < k_a \triangleq \sum_{\mathbf{x}_a} f_a(\mathbf{x}_a), \quad \forall\ a \in \mathcal{A}.$$

Then

$$\begin{aligned}
F_1 &= \sum_{a\in\mathcal{A}_{\text{BP}}} D(b_a\,||\,f_a/k_a) - \sum_{a\in\mathcal{A}_{\text{BP}}}\ln k_a\\
&\geq -\sum_{a\in\mathcal{A}_{\text{BP}}}\ln(k_a)\\
&> -\infty\\
F_2 &= \sum_{a\in\mathcal{A}_{\text{MF}}} D\Big(\prod_{i\in\mathcal{N}(a)} b_i\ ||\ f_a/k_a\Big) - \sum_{a\in\mathcal{A}_{\text{MF}}}\ln k_a\\
&\geq -\sum_{a\in\mathcal{A}_{\text{MF}}}\ln k_a\\
&> -\infty\\
F_3 &\geq 0.
\end{aligned}$$

This proves 1). Now $F_3 < \infty$, (27) implies that $F_2 < \infty$, and (26) implies that $F_1 < \infty$ if and only if (28) if fulfilled, which proves 2).

Suppose that we have fixed all $b_a(\bar{\mathbf{x}}_a)$ ($a \in \mathcal{A}_{\text{BP}}$) from (28). Then the analysis for the remaining $b_i(x_i)$ ($i \in \mathcal{I}$) and $b_a(\mathbf{x}_a)$ excluding all $\bar{\mathbf{x}}_a$ from (28) ($a \in \mathcal{A}_{\text{BP}}$) is the same as in the proof of Theorem 2 and the resulting fixed-point equations are identical to (24) and (25) excluding all $\bar{\mathbf{x}}_a$ from (28) and vice versa, which proves 3). We can reintroduce the realizations $\bar{\mathbf{x}}_a$ with $f_a(\bar{\mathbf{x}}_a) = 0$ ($a \in \mathcal{A}_{\text{BP}}$) from (28) in (25) because they do not contribute to the message passing update equations, as can be seen immediately from the definition of the messages $m^{\text{BP}}_{a\to i}(x_i)$ ($a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a)$) in (25). The same argument implies that (28) is a special case of the first equation in (24), which proves 4) and, therefore, finishes the proof of Lemma 2.

E. Proof of convergence

In order to finish the proof of convergence for the algorithm presented in Subsection III-B, we need to show that running the forward/backward algorithm in the BP part in step 2) of Algorithm 1 cannot increase the region-based free energy approximation $F_{\text{BP,MF}}$ in (20). To this end we analyze the factorization

$$p(\mathbf{x}_{\text{BP}}) \propto \prod_{a\in\mathcal{A}_{\text{BP}}} f_a(\mathbf{x}_a)\prod_{i\in\mathcal{I}_{\text{BP}}\cap\mathcal{I}_{\text{MF}}}\ \prod_{b\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{b\to i}(x_i) \tag{63}$$

with $\mathbf{x}_{\text{BP}} \triangleq (x_i\ |\ i \in \mathcal{I}_{\text{BP}})^{\text{T}}$. The factorization in (63) is the product of the factorization of the BP part in (18) and the incoming messages from the MF part. The Bethe free energy (3) corresponding to the factorization in (63) is

$$\begin{aligned}
F_{\text{BP}} &= \sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln\frac{b_a(\mathbf{x}_a)}{f_a(\mathbf{x}_a)}\\
&\quad + \sum_{i\in\mathcal{I}_{\text{BP}}\cap\mathcal{I}_{\text{MF}}}\ \sum_{a\in\mathcal{N}_{\text{MF}}(i)}\sum_{x_i} b_i(x_i)\ln\frac{b_i(x_i)}{m^{\text{MF}}_{a\to i}(x_i)}\\
&\quad - \sum_{i\in\mathcal{I}_{\text{BP}}}(|\mathcal{N}_{\text{BP}}(i)| + |\mathcal{N}_{\text{MF}}(i)| - 1)\sum_{x_i} b_i(x_i)\ln b_i(x_i)\\
&= \sum_{a\in\mathcal{A}_{\text{BP}}}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln\frac{b_a(\mathbf{x}_a)}{f_a(\mathbf{x}_a)}\\
&\quad - \sum_{i\in\mathcal{I}_{\text{BP}}\cap\mathcal{I}_{\text{MF}}}\ \sum_{a\in\mathcal{N}_{\text{MF}}(i)}\sum_{x_i} b_i(x_i)\ln m^{\text{MF}}_{a\to i}(x_i)\\
&\quad - \sum_{i\in\mathcal{I}_{\text{BP}}}(|\mathcal{N}_{\text{BP}}(i)| - 1)\sum_{x_i} b_i(x_i)\ln b_i(x_i).
\end{aligned} \tag{64}$$

We now show that minimizing F_{BP} in (64) is equivalent to minimizing $F_{\text{BP,MF}}$ in (20) with respect to b_a and b_i for all $a \in \mathcal{A}_{\text{BP}}$ and $i \in \mathcal{I}_{\text{BP}}$. Obvioulsy,

$$\frac{\partial F_{\text{BP,MF}}}{\partial b_i(x_i)} = \frac{\partial F_{\text{BP}}}{\partial b_i(x_i)}, \quad \forall i \in \mathcal{I}_{\text{BP}}\setminus\mathcal{I}_{\text{MF}}$$

and

$$\frac{\partial F_{\text{BP,MF}}}{\partial b_a(\mathbf{x}_a)} = \frac{\partial F_{\text{BP}}}{\partial b_a(\mathbf{x}_a)}, \quad \forall a \in \mathcal{A}_{\text{BP}}.$$

This follows from the fact that $F_{\text{BP,MF}}$ differs from F_{BP} by terms that depend only on b_i with $i \in \mathcal{I}_{\text{MF}}$. Now suppose that $i \in \mathcal{I}_{\text{BP}}\cap\mathcal{I}_{\text{MF}}$. In this case, we find that

$$\begin{aligned}
\frac{\partial F_{\text{BP,MF}}}{\partial b_i(x_i)} &= (1 - |\mathcal{N}_{\text{BP}}(i)|)(\ln b_i(x_i) + 1)\\
&\quad - \sum_{a\in\mathcal{N}_{\text{MF}}(i)}\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} b_j(x_j)\ln f_a(\mathbf{x}_a)
\end{aligned} \tag{65}$$

and

$$\frac{\partial F_{\text{BP}}}{\partial b_i(x_i)} = (1 - |\mathcal{N}_{\text{BP}}(i)|)(\ln b_i(x_i) + 1) - \sum_{a\in\mathcal{N}_{\text{MF}}(i)}\ln m^{\text{MF}}_{a\to i}(x_i). \tag{66}$$

From (25) we see that

$$m^{\text{MF}}_{a\to i}(x_i) = \exp\Bigg(\sum_{\mathbf{x}_a\setminus x_i}\prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j)\ln f_a(\mathbf{x}_a)\Bigg), \quad \forall\ a \in \mathcal{N}_{\text{MF}}(i). \tag{67}$$

Note that, according to step 2) in Algorithm 1, the messages $m^{\text{MF}}_{a\to i}(x_i)$ in (67) are *fixed inputs* for the BP part. Therefore, we are not allowed to plug the expressions for the messages $m^{\text{MF}}_{a\to i}(x_i)$ in (67) into (66) in general. However, since $a \in \mathcal{A}_{\text{MF}}$ and $i \in \mathcal{I}_{\text{BP}} \cap \mathcal{I}_{\text{MF}}$, condition (29) implies that $\mathcal{N}(a) \setminus i \subseteq \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$ and guarantees that

$$n_{j\to a}(x_j) = b_j(x_j) \tag{68}$$

is constant in step 2) of Algorithm 1 for all $j \in \mathcal{N}(a) \setminus i \subseteq \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$. Therefore, we are indeed allowed to plug the expressions of the messages $m^{\text{MF}}_{a\to i}(x_i)$ in (67) into (66) and finally see that also

$$\frac{\partial F_{\text{BP,MF}}}{\partial b_i(x_i)} = \frac{\partial F_{\text{BP}}}{\partial b_i(x_i)}, \quad \forall i \in \mathcal{I}_{\text{BP}} \cap \mathcal{I}_{\text{MF}}.$$

Hence, minimizing F_{BP} in (64) is equivalent to minimizing $F_{\text{BP,MF}}$ in (20).

By assumption, the factor graph in the BP part has tree structure. Therefore, [9, Prop. 3] implies that

1) $F_{\text{BP}} \geq 0$;
2) $F_{\text{BP}} = 0$ if and only if the beliefs $\{b_i, b_a\}$ in (64) are the marginals of the factorization in (63).

Hence, for b_j fixed with $j \in \mathcal{I}_{\text{MF}} \setminus \mathcal{I}_{\text{BP}}$, we see that $F_{\text{BP,MF}}$ in (20) is minimized by the marginals of the factorization in (63).

It remains to show that running the forward/backword algorithm in the BP part as described in step 2) in Algorithm 1 indeed computes the marginals of the factorization in (63). Applying Theorem 1 to the factorization in (63) yields the message passing fixed-point equations

$$\begin{cases} n_{i\to a}(x_i) = \displaystyle\prod_{c\in\mathcal{N}_{\text{BP}}(i)\setminus a} m^{\text{BP}}_{c\to i}(x_i) \prod_{c\in\mathcal{N}_{\text{MF}}(i)} m^{\text{MF}}_{c\to i}(x_i), \\ \qquad\qquad \forall\, a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a) \\ m^{\text{BP}}_{a\to i}(x_i) = z_a \displaystyle\sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(x_j), \\ \qquad\qquad \forall\, a \in \mathcal{A}_{\text{BP}}, i \in \mathcal{N}(a). \end{cases} \tag{69}$$

The message passing fixed-point equations in (69) are the same as the message passing fixed-point equations for the BP part in (25) with fixed-input messages $m^{\text{MF}}_{a\to i}(x_i)$ for all $i \in \mathcal{I}_{\text{BP}} \cap \mathcal{I}_{\text{MF}}$ and $a \in \mathcal{N}_{\text{MF}}(i)$. Hence, running the forward/backward algorithm in the BP part indeed computes the marginals of the factorization in (63) and Algorithm 1 is guaranteed to converge.

F. Product of Gaussian distributions

Lemma 3: Let

$$p_i(\mathbf{x}) = \text{CN}(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i^{-1}), \quad \forall\, i \in [1:N].$$

Then

$$\prod_{i\in[1:N]} p_i(\mathbf{x}) \propto \text{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$

with

$$\boldsymbol{\mu} \triangleq \sum_{i\in[1:N]} \boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i\boldsymbol{\mu}_i$$
$$\boldsymbol{\Lambda} \triangleq \sum_{i\in[1:N]} \boldsymbol{\Lambda}_i.$$

Proof: Follows from direct computation. ■

References

[1] E. Riegler, G. E. Kirkelund, C. N. Manchón, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," in *IEEE Symp. Turbo Codes (ISTC 2010)*, Brest, France, Sep. 2010, pp. 1–5.
[2] P. Parisi, *Statistical Field Theory*. Oxford, UK: Perseus, 1988.
[3] Z. Ghahramani and M. J. Beal, *Graphical Models and Variational Methods*, D. Saad and M. Opper, Eds. Massachusetts, NY: MIT Press, 2000.
[4] M. I. Jordan, Z. Ghahramani, T. S. Jaakola, and L. K. Saul, "An introduction to variational methods for graphical models," *J. Mach. Learning*, vol. 37, no. 2, pp. 183–233, 1999.
[5] J. Winn and C. Bishop, "Variational message passing," *J. Mach. Learning*, vol. 6, no. 6, pp. 661–694, 2005.
[6] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.
[7] G. E. Kirkelund, C. N. Manchón, L. P. B. Christensen, E. Riegler, and B. H. Fleury, "Variational message-passing for joint channel estimation and decoding in MIMO-OFDM," in *Proc. IEEE Globecom (GLOBECOM 2010)*, Miami, FL, Dec. 2010, pp. 1–5.
[8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufman, 1998.
[9] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Th.*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
[10] F. R. Kschischang, J. F. Brendan, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Th.*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[11] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "A new class of upper bounds on the log partition function," *IEEE Trans. Inf. Th.*, vol. 51, no. 7, pp. 2313–2335, 2005.
[12] H. Wymeersch, F. Penna, and V. Savić, "Uniformly reweighted belief propagation: A factor graph approach," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2011)*, St. Petersburg, Russia, July–Aug. 2011, pp. 1–5.
[13] S. Ikeda, T. Tanaka, and S.-I. Amari, "Information geometry of turbo and low-density parity-check codes," *IEEE Trans. Inf. Th.*, vol. 50, no. 6, pp. 1097–1114, June 2004.
[14] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*. Oxford, UK: Oxford Univ. Press, 2000.
[15] M. Moher and T. A. Gulliver, "Cross-entropy and iterative decoding," *IEEE Trans. Inf. Th.*, vol. 44, no. 7, pp. 3097–3104, Nov. 1998.
[16] J. M. Walsh, C. R. Johnson, and P. A. Regalia, "A refined information geometric interpretation of turbo decoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process. (ICASSP 2005)*, Philadelthia, PA, Mar. 2005, pp. 481–484.
[17] T. Minka, "Divergence measures and message passing," *Microsoft Research, Tech. Rep. MSR-TR-2005-173*, 2005.
[18] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Comm.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
[19] J. Wehinger and C. F. Mecklenbrauker, "Iterative CDMA multiuser receiver with soft decision-directed channel estimation," *IEEE Trans. Sig. Process.*, vol. 54, no. 10, pp. 3922–3934, Oct. 2006.
[20] S. P. Rossi and R. R. Müller, "Joint twofold-iterative channel estimation and multiuser detection for MIMO-OFDM systems," *IEEE Trans. Wireless Comm.*, vol. 7, no. 11, pp. 4719–4729, Nov. 2008.
[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
[22] S. Kullback, *Information Theory and Statistics*. New York (NY): Wiley, 1978.
[23] C. N. Manchón, G. E. Kirkelund, E. Riegler, L. Christensen, and B. H. Fleury, "Receiver architectures for MIMO-OFDM based on a combined VMP-SP algorithm," submitted to IEEE Trans. Inf. Th., 2011, arXiv:1111.5848 [stat.ML].

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.
[25] H. A. Bethe, "Statistical theory of superlattices," in *Proc. Roy. Soc. London*, ser. A, vol. 150, no. 871, 1935, pp. 552–575.
[26] T. Heskes, "Stable fixed points of loopy belief propagation are minima of the bethe free energy," in *Adv. in Neural Inf. Process. Systems*. MIT Press, 2003, pp. 343–350.
[27] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 2003.
[28] J. Dauwels, "On variational message passing on factor graphs," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2007)*, Nice, France, June 2007, pp. 2546–2550.
[29] T. Tanaka, "Information geometry of mean-field approximation," *Neur. Comput.*, pp. 1951–1968, 2000.
[30] A. W. Eckford, "The factor graph EM algorithm: Applications for LDPC codes," in *Proc. IEEE SPAWC (SPAWC 2005)*, New York, NY, June 2005, pp. 910–914.
[31] J. Dauwels, S. Korl, and H.-A. Loeliger, "Expectation maximization as message passing," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2005)*, Adelaide, Australia, Sep. 2005, pp. 1–4.
[32] B. Hu, "A variational bayesian framework divergence minimization and its application in cdma receivers," Ph.D. dissertation, July 2010.
[33] Z. Shi, T. Wo, P. Höher, and G. Auer, "Graph-based soft iterative receiver for higher-order modulation," in *Proc. 12th IEEE Int. Conf. on Comm. Tech. (ICCT)*, Nanjing, China, Nov. 2010, pp. 825–828.
[34] X. Xu and R. Mathar, "Low complexity joint channel estimation and detection for LDPC coded MIMO-OFDM systems," in *Proc. IEEE 73rd Veh. Tech. Conf. (VTC Spring)*, Budapest, Hungary, May 2011, pp. 1–5.
[35] B. D. Craven, "A generalization of lagrange multipliers," *Bull. Austral. Math. Soc.*, vol. 3, pp. 353–362, 1970.
[36] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY: Wiley, 1997.
[37] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, Feb. 2009.
[38] C. Pedersen, T. Pedersen, and B. H. Fleury, "A variational message passing algorithm for sensor self-localization in wireless networks," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2011)*, St. Petersburg, Russia, Jul.–Aug. 2011, pp. 1–5.