Gaussian Belief Propagation Based Multiuser Detection

Danny Bickson[1] and Danny Dolev
School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem 91904, Israel
{daniel51,dolev}@cs.huji.ac.il

Ori Shental[1], Paul H. Siegel and Jack K. Wolf
Center for Magnetic Recording Research
University of California - San Diego
La Jolla, CA 92093
{oshental,psiegel,jwolf}@ucsd.edu

***Abstract*— In this work, we present a novel construction for solving the linear multiuser detection problem using the Gaussian Belief Propagation algorithm. Our algorithm yields an efficient, iterative and distributed implementation of the MMSE detector. Compared to our previous formulation, the new algorithm offers a reduction in memory requirements, the number of computational steps, and the number of messages passed. We prove that a detection method recently proposed by Montanari *et al.* is an instance of ours, and we provide new convergence results applicable to both.**

I. Introduction

Belief propagation (BP), also known as the sum-product algorithm, is a powerful and efficient tool in solving, exactly or approximately, inference problems in probabilistic graphical models. The underlying essence of estimation theory is to detect a hidden input to a channel from its observed output. The channel can be represented as a certain graphical model, while the detection of the channel input is equivalent to performing inference in the corresponding graph.

The use of BP [1] for detection purposes has been proven to be very beneficial in several applications in communications. For randomly-spread code-division multiple-access (CDMA) in the large-system limit, Kabashima has introduced a tractable BP-based multiuser detection (MUD) scheme, which exhibits near-optimal error performance for binary-input additive white Gaussian noise (BI-AWGN) channels [2]. This message-passing scheme has recently been extended to the case where the ambient noise level is unknown [3], [4]. As for sub-optimal detection, the nonlinear soft parallel interference cancellation (PIC) detector was reformulated by Tanaka and Okada as an approximate BP solution [5] to the MUD problem.

In contrast to the dense, fully-connected nature of the graphical model of the non-orthogonal CDMA channel, a one-dimensional (1-D) intersymbol interference (ISI) channel can be interpreted as a cycle-free tree graph [6]. Thus, detection in 1-D ISI channels (termed equalization) can be performed in an optimal maximum a-posteriori (MAP) manner via BP, also known in this context as the forward/backward, or BCJR, algorithm [7]. Also, Kurkoski *et al.* [8], [9] have proposed an iterative BP-like detection algorithm for 1-D ISI channels that uses a parallel message-passing schedule and achieves near-optimal performance.

For the intermediate regime of non-dense graphs but with many relatively short loops, extensions of BP to two-dimensional ISI channels have been considered by Marrow and Wolf [10], and recently Shental *et al.* [11]–[13] have demonstrated the near-optimality of a generalized version of BP for such channels. Recently, BP has been proved to asymptotically achieve optimal MAP detection for sparse linear systems with Gaussian noise [14], [15], for example, in CDMA with sparse spreading codes.

An important class of practical sub-optimal detectors is based on linear detection. This class includes, for instance, the conventional single-user matched filter (MF), decorrelator (a.k.a. zero-forcing equalizer), linear minimum mean-square error (MMSE) detector and many other detectors with widespread applicability [16], [17]. In general, linear detection can be viewed as the solution to a (deterministic) set of linear equations describing the original (probabilistic) estimation problem. Note that the mathematical operation behind linear detection extends to other tasks in communication, *e.g.*, channel precoding at the transmitter [18].

Recently, linear detection has been explicitly linked to BP [19], using a Gaussian belief propagation (GaBP) algorithm. This allows for a distributed implementation of the linear detector [20], circumventing the need of, potentially cumbersome, direct matrix inversion (via, *e.g.*, Gaussian elimination). The derived iterative framework was compared quantitatively with 'classical' iterative methods for solving systems of linear equations, such as those investigated in the context of linear implementation of CDMA demodulation [21]–[23]. GaBP is shown to yield faster convergence than these standard methods. Another important work is the BP-based MUD, recently derived and analyzed by Montanari *et al.* [24] for Gaussian input symbols.

There are several drawbacks to the linear detection technique of [19]. First, the input matrix $\mathbf{R}_{n\times n} = \mathbf{S}^T_{n\times k}\mathbf{S}_{k\times n}$ (the chip correlation matrix) needs to be computed prior to running the algorithm. This computation requires n^2k operations. In case where the matrix $\mathbf{S}$ is sparse [15], the matrix $\mathbf{R}$ might

[1]Contributed equally to this work.
Supported in part by NSF Grant No. CCR-0514859 and EVERGROW, IP 1935 of the EU Sixth Framework.

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

not no longer be sparse. Second, GaBP uses $2n^2$ memory to store the messages. For a large n this could be prohibitive.

In this paper, we propose a new construction that addresses those two drawbacks. In our improved construction, given a non-rectangular CDMA matrix $\mathbf{S}_{n\times k}$, we compute the MMSE detector $x = (\mathbf{S}^T\mathbf{S} + \Psi)^{-1}\mathbf{S}^T y$ where Ψ is the AWGN diagonal covariance matrix. We utilize the GaBP algorithm which is an efficient iterative distributed algorithm. The new construction uses only $2nk$ memory for storing the messages. When $k \ll n$ this represents significant saving relative to the $2n^2$ in our previously proposed algorithm. Furthermore, we do not explicitly compute $\mathbf{S}^T\mathbf{S}$, saving an extra n^2k overhead.

We show that Montanari's algorithm [24] is an instance of our method. By showing this, we are able to prove new convergence results for Montanari's algorithm. Montanari proves that his method converges on normalized random-spreading CDMA sequences, assuming Gaussian signaling. Using binary signaling, he conjectures convergence to the large system limit. Here, we extend Montanari's result, to show that his algorithm converges also for non-random CDMA sequences when binary signaling is used, under weaker conditions. Another advantage of our work is that we allow different noise levels per bit transmitted.

The paper is organized as follows. Section II formulates the problem of linear detection and presents the distributed GaBP-based linear detection scheme. Section III describes a novel construction for efficiently computing the MMSE detector. The relation to a factor graph construction is explored in Section IV. New convergence results for Montanari's work are presented in Section V. We conclude in Section VI. In the Appendix we further explore the relation to Montanari's work.

We shall use the following notations. The operator $\{\cdot\}^T$ stands for a vector or matrix transpose, the matrix $\mathbf{I}_N$ is a $N\times N$ identity matrix, while the symbols $\{\cdot\}_i$ and $\{\cdot\}_{ij}$ denote entries of a vector and matrix, respectively. $\mathrm{N}(i)$ is the set of graph node connected to node i.

II. Linear Detection via Belief Propagation

Consider a discrete-time channel with a real input vector $\mathbf{x} = \{x_1, \ldots, x_K\}^T$ governed by an arbitrary prior distribution, $P_{\mathbf{x}}$, and a corresponding real output vector $\mathbf{y} = \{y_1, \ldots, y_K\}^T = f\{\mathbf{x}^T\} \in \mathbb{R}^K$. Here, the function $f\{\cdot\}$ denotes the channel transformation. By definition, linear detection compels the decision rule to be

$$\hat{\mathbf{x}} = \Delta\{\mathbf{x}^*\} = \Delta\{\mathbf{A}^{-1}\mathbf{b}\}, \tag{1}$$

where $\mathbf{b} = \mathbf{y}$ is the $K \times 1$ observation vector and the $K \times K$ matrix $\mathbf{A}$ is a positive-definite symmetric matrix approximating the channel transformation. The vector $\mathbf{x}^*$ is the solution (over $\mathbb{R}$) to $\mathbf{A}\mathbf{x} = \mathbf{b}$. Estimation is completed by adjusting the (inverse) matrix-vector product to the input alphabet, dictated by $P_{\mathbf{x}}$, accomplished by using a proper clipping function $\Delta\{\cdot\}$ (*e.g.*, for binary signaling $\Delta\{\cdot\}$ is the sign function).

For example, linear channels, which appear extensively in many applications in communication and data storage systems, are characterized by the linear relation

$$\mathbf{y} = f\{\mathbf{x}\} = \mathbf{R}\mathbf{x} + \mathbf{n},$$

where $\mathbf{n}$ is a $K \times 1$ additive noise vector and $\mathbf{R} = \mathbf{S}^T\mathbf{S}$ is a positive-definite symmetric matrix, often known as the correlation matrix. The $N \times K$ matrix $\mathbf{S}$ describes the physical channel medium while the vector $\mathbf{y}$ corresponds to the output of a bank of filters matched to the physical channel $\mathbf{S}$.

Assuming linear channels with AWGN with variance σ^2 as the ambient noise, the general linear detection rule (1) can describe known linear detectors. For example [16], [17]:

- The conventional matched filter (MF) detector is obtained by taking $\mathbf{A} \triangleq \mathbf{I}_K$ and $\mathbf{b} = \mathbf{y}$. This detector is optimal, in the MAP-sense, for the case of zero cross-correlations, *i.e.*, $\mathbf{R} = \mathbf{I}_K$, as happens for orthogonal CDMA or when there is no ISI effect.
- The decorrelator (zero forcing equalizer) is achieved by substituting $\mathbf{A} \triangleq \mathbf{R}$ and $\mathbf{b} = \mathbf{y}$. It is optimal in the noiseless case.
- The linear minimum mean-square error (MMSE) detector can also be described by using $\mathbf{A} = \mathbf{R} + \sigma^2\mathbf{I}_K$. This detector is known to be optimal when the input distribution $P_{\mathbf{x}}$ is Gaussian.

In general, linear detection is suboptimal because of its deterministic underlying mechanism (*i.e.*, solving a given set of linear equations), in contrast to other estimation schemes, such as MAP or maximum likelihood, that emerge from an optimization criterion.

In [19], linear detection, in its general form (1), was implemented using an efficient message-passing algorithm. The linear detection problem was shifted from an algebraic to a probabilistic domain. Instead of solving a deterministic vector-matrix linear equation, an inference problem is solved in a graphical model describing a certain Gaussian distribution function. Given the overall channel matrix $\mathbf{R}$ and the observation vector $\mathbf{y}$, one knows how to write explicitly $p(\mathbf{x})$ and the corresponding graph $\mathcal{G}$ with edge potentials ('compatibility functions') ψ_{ij} and self-potentials ('evidence') ϕ_i. These graph potentials are determined according to the following pairwise factorization of the Gaussian distribution $p(\mathbf{x})$

$$p(\mathbf{x}) \propto \prod_{i=1}^{K} \phi_i(x_i) \prod_{\{i,j\}} \psi_{ij}(x_i, x_j),$$

resulting in $\psi_{ij}(x_i, x_j) \triangleq \exp(-x_i R_{ij} x_j)$ and $\phi_i(x_i) = \exp\left(b_i x_i - R_{ii}x_i^2/2\right)$. The set of edges $\{i, j\}$ corresponds to the set of all non-zero entries of $\mathbf{A}$ for which $i > j$. Hence, we would like to calculate the marginal densities, which must also be Gaussian, $p(x_i) \sim \mathcal{N}(\mu_i = \{\mathbf{R}^{-1}\mathbf{y}\}_i, P_i^{-1} = \{\mathbf{R}^{-1}\}_{ii})$, where μ_i and P_i are the marginal mean and inverse variance (a.k.a. precision), respectively. It is shown that the inferred mean μ is identical to the desired solution $x^* = \mathbf{R}^{-1}\mathbf{y}$. Table I lists the GaBP algorithm update rules.

TABLE I
COMPUTING $\mathbf{A}^{-1}\mathbf{b}$ VIA GABP. ONLINE MATLAB IMPLEMENTATION IS PROVIDED IN [25].

#	Stage	Operation
1.	*Initialize*	Compute $P_{ii} = A_{ii}$ and $\mu_{ii} = b_i/A_{ii}$. Set $P_{ki} = 0$ and $\mu_{ki} = 0$, $\forall k \neq i$.
2.	*Iterate*	Propagate P_{ki} and μ_{ki}, $\forall k \neq i$ such that $A_{ki} \neq 0$. Compute $P_{i\backslash j} = P_{ii} + \sum_{k\in\mathbb{N}(i)\backslash j} P_{ki}$ and $\mu_{i\backslash j} = P_{i\backslash j}^{-1}(P_{ii}\mu_{ii} + \sum_{k\in\mathbb{N}(i)\backslash j} P_{ki}\mu_{ki})$. Compute $P_{ij} = -A_{ij}P_{i\backslash j}^{-1}A_{ji}$ and $\mu_{ij} = -P_{ij}^{-1}A_{ij}\mu_{i\backslash j}$.
3.	*Check*	If P_{ij} and μ_{ij} did not converge, return to #2. Else, continue to #4.
4.	*Infer*	$P_i = P_{ii} + \sum_{k\in\mathbb{N}(i)} P_{ki}$, $\mu_i = P_i^{-1}(P_{ii}\mu_{ii} + \sum_{k\in\mathbb{N}(i)} P_{ki}\mu_{ki})$.
5.	*Decide*	$\hat{x}_i = \Delta\{\mu_i\}$

III. DISTRIBUTED ITERATIVE COMPUTATION OF THE MMSE DETECTOR

In this section, we efficiently extend the applicability of the proposed GaBP-based solver for systems with symmetric matrices [19] to systems with any square (*i.e.*, also nonsymmetric) or rectangular matrix. We first construct a new symmetric data matrix $\tilde{\mathbf{R}}$ based on an arbitrary (non-rectangular) matrix $\mathbf{S} \in \mathbb{R}^{k\times n}$

$$\tilde{\mathbf{R}} \triangleq \begin{pmatrix} \mathbf{I}_k & \mathbf{S}^T \\ \mathbf{S} & -\Psi \end{pmatrix} \in \mathbb{R}^{(k+n)\times(k+n)}. \quad (2)$$

Additionally, we define a new vector of variables $\tilde{\mathbf{x}} \triangleq \{\hat{\mathbf{x}}^T, \mathbf{z}^T\}^T \in \mathbb{R}^{(k+n)\times 1}$, where $\hat{\mathbf{x}} \in \mathbb{R}^{k\times 1}$ is the (to be shown) solution vector and $\mathbf{z} \in \mathbb{R}^{n\times 1}$ is an auxiliary hidden vector, and a new observation vector $\tilde{\mathbf{y}} \triangleq \{\mathbf{0}^T, \mathbf{y}^T\}^T \in \mathbb{R}^{(k+n)\times 1}$.

Now, we would like to show that solving the symmetric linear system $\tilde{\mathbf{R}}\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ and taking the first k entries of the corresponding solution vector $\tilde{\mathbf{x}}$ is equivalent to solving the original (not necessarily symmetric) system $\mathbf{R}\mathbf{x} = \mathbf{y}$. Note that in the new construction the matrix $\tilde{\mathbf{R}}$ is sparse again, and has only $2nk$ off-diagonal nonzero elements. When running the GaBP algorithm we have only $2nk$ messages, instead of n^2 in the previous construction.

Writing explicitly the symmetric linear system's equations, we get

$$\hat{\mathbf{x}} + \mathbf{S}^T\mathbf{z} = \mathbf{0}, \quad \mathbf{S}\hat{\mathbf{x}} - \Psi\mathbf{z} = \mathbf{y}.$$

Thus,

$$\hat{\mathbf{x}} = \Psi^{-1}\mathbf{S}^T(\mathbf{y} - \mathbf{S}\hat{\mathbf{x}}),$$

and extracting $\hat{\mathbf{x}}$ we have

$$\hat{\mathbf{x}} = (\mathbf{S}^T\mathbf{S} + \Psi)^{-1}\mathbf{S}^T\mathbf{y}.$$

Note, that when the noise level is zero, $\Psi = 0_{m\times m}$, we get the Moore-Penrose pseudoinverse solution

$$\hat{\mathbf{x}} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{y} = \mathbf{S}^\dagger\mathbf{y}.$$

IV. RELATION TO FACTOR GRAPH

In this section we give an alternate proof of the correctness of our construction. Given the inverse covariance matrix $\tilde{\mathbf{R}}$ defined in (2), and the shift vector $\tilde{\mathbf{x}}$ we can derive the matching self and edge potentials

$$\psi_{ij}(x_i, x_j) \triangleq \exp(-x_i R_{ij} x_j)$$

$$\phi_i(x_i) \triangleq \exp(-1/2 x_i R_{ii}^2 x_i - x_i y_i)$$

which is a factorization of the Gaussian system distribution

$$p(\mathbf{x}) \propto \prod_i \phi_i(x_i) \prod_{i,j} \psi_{ij}(x_i, x_j) =$$

$$= \prod_{i\leq k} \phi_i(x_i) \prod_{i>k} \phi_i(x_i) \prod_{i,j} \psi_{ij}(x_i, x_j) =$$

$$= \prod_{i\leq k} \overbrace{\exp(-\frac{1}{2}x_i^2)}^{\text{prior on x}} \prod_{i>k} \exp(-\frac{1}{2}\Psi_i x_i^2 - x_i y_i) \prod_{i,j} \exp(-x_i \overbrace{S_{ij}}^{R_{ij}} x_j)$$

Next, we show the relation of our construction to a factor graph. We will use a factor graph with k nodes to the left (the bits transmitted) and n nodes to the right (the signal received), shown in Fig. 1. Using the definition $\tilde{\mathbf{x}} \triangleq \{\hat{\mathbf{x}}^T, \mathbf{z}^T\}^T \in \mathbb{R}^{(k+n)\times 1}$ the vector $\hat{\mathbf{x}}$ represents the k input bits and the vector $\mathbf{z}$ represents the signal received. Now we can write the system probability as:

$$p(\tilde{\mathbf{x}}) \propto \int_{\hat{\mathbf{x}}} \mathcal{N}(\hat{\mathbf{x}}; 0, I)\mathcal{N}(\mathbf{z}; S\hat{\mathbf{x}}, \Psi)d\hat{\mathbf{x}}$$

It is known that the marginal distribution over $\mathbf{z}$ is:

$$= \mathcal{N}(\mathbf{z}; 0, \mathbf{S}^T\mathbf{S} + \Psi)$$

This distribution is Gaussian, with the following parameters:

$$E(\mathbf{z}|\hat{\mathbf{x}}) = (\mathbf{S}^T\mathbf{S} + \Psi)^{-1}\mathbf{S}^T\mathbf{y}$$

$$Cov(\mathbf{z}|\hat{\mathbf{x}}) = (\mathbf{S}^T\mathbf{S} + \Psi)^{-1}$$

It is interesting to note that a similar construction was used by Frey [26] in his seminal 1999 work when discussing the factor analysis learning problem. While his work is beyond the scope of this paper, it can be shown that his algorithm can be modelled using the GaBP algorithm.

V. NEW CONVERGENCE RESULTS

One of the benefits of using our new construction is that we propose a new mechanism to provide future convergence results. In the Appendix we prove that Montanari's algorithm is an instance of our algorithm, thus our convergence results apply to Montanari's algorithm as well.

We know that if the matrix $\tilde{\mathbf{R}}$ is strictly diagonally dominant, then GaBP converges and the marginal means converge

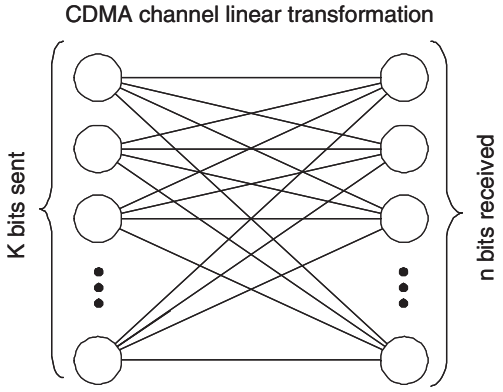

Fig. 1. Factor graph describing the linear channel

to the true means [27, Claim 4]. Noting that the matrix $\tilde{\mathbf{R}}$ is symmetric, we can determine the applicability of this condition by examining its columns. Referring to (4) we see that in the first k columns, we have the k CDMA sequences. We assume random-spreading binary CDMA sequences which are normalized to one. In other words, the absolute sum of each column is $\sqrt{n}$. In that case, the matrix $\tilde{\mathbf{R}}$ is not diagonally dominant (DD). We can add a regularization term of $\sqrt{n}+\epsilon$ to force the matrix $\tilde{\mathbf{R}}$ to be DD, but we pay in changing the problem. In the next n columns of the matrix $\tilde{\mathbf{R}}$, we have the diagonal covariance matrix Ψ with different noise levels per bit in the main diagonal, and zero elsewhere. The absolute sum of each column of S is $k/\sqrt{n}$, thus when the noise level of each bit satisfies $\Psi_i > k/\sqrt{n}$, we have a convergence guarantee. Note, that the convergence condition is a *sufficient condition*. Based on Montanari's work, we also know that in the large system limit, the algorithm converges for binary signaling, even in the absence of noise.

An area of future work is to utilize this observation to identify CDMA schemes with matrices $\mathbf{S}$ that when fitted into the matrix $\tilde{\mathbf{R}}$ are either DD, or comply to the spectral radius convergence condition of [28].

VI. Conclusion

We presented a novel distributed algorithm for computing the MMSE detector for the CDMA multiuser detection problem. Our work utilizes the Gaussian Belief Propagation algorithm while improving two existing constructions [19], [24] in this field. Although we described our algorithm in the context of multiuser detection, it has wider applicability. For example, it provides an efficient iterative method for computing the Moore-Penrose pseudoinverse, and it can also be applied to the factor analysis learning problem [26].

APPENDIX: Montanari's algorithm is an instance of our algorithm

In this section we show that Montanari's algorithm is an instance of our algorithm. Our algorithm is more general. First, we allow different noise level for each received bit, unlike his work which uses a single fixed noise for the whole system. In practice, the bits are transmitted using different frequencies, thus suffering from different noise levels. Second, the update rules in his paper are fitted only to the randoml-spreading CDMA codes, where the matrix A contains only values which are drawn uniformly from $\{-1,1\}$. Assuming binary signalling, he conjectures convergence to the large system limit. Our new convergence proof holds for any CDMA matrices provided that the absolute sum of the chip sequences is one, under weaker conditions on the noise level. Third, we propose in [19] an efficient broadcast version for saving messages in a broadcast supporting network.

The probability distribution of the factor graph used by Montanari is:

$$d\mu_y^{N,K} = \frac{1}{Z_y^{N,K}} \prod_{a=1}^{N} \exp(-\frac{1}{2}\sigma^2\omega_a^2 + jy_a\omega_a) \prod_{i=1}^{K} \exp(-\frac{1}{2}x_i^2) \cdot \prod_{i,a} \exp(-\frac{j}{\sqrt{N}} s_{ai}\omega_a x_i) d\omega$$

Extracting the self and edge potentials from the above probability distribution:

$$\psi_{ii}(\mathbf{x}_i) \triangleq \exp(-\frac{1}{2}\mathbf{x}_i^2) \propto \mathcal{N}(\mathbf{x}; 0, 1)$$

$$\psi_{aa}(\omega_a) \triangleq \exp(-\frac{1}{2}\sigma^2\omega_a^2 + j\mathbf{y}_a\omega_a) \propto \mathcal{N}(\omega_a; j\mathbf{y}_a, \sigma^2)$$

$$\psi_{ia}(\mathbf{x}_i, \omega_a) \triangleq \exp(-\frac{j}{\sqrt{N}} s_{ai}\omega_a\mathbf{x}_i) \propto \mathcal{N}(\mathbf{x}; \frac{j}{\sqrt{N}} s_{ai}, 0)$$

For convenience, Table II provides a translation between the notations used in this paper (taken from [19]) and that used by Montanari *et al.* in [24]:

TABLE II
SUMMARY OF NOTATIONS

This work [19]	Montanari *el al.* [24]	Description
P_{ij}	$\lambda_{i\to a}^{(t+1)}$	precision msg from left to right
	$\hat{\lambda}_{a\to i}^{(t+1)}$	precision msg from right to left
μ_{ij}	$\gamma_{i\to a}^{(t+1)}$	mean msg from left to right
	$\hat{\gamma}_{a\to i}^{(t+1)}$	mean msg from right to left
μ_{ii}	y_i	prior mean of left node
	0	prior mean of right node
P_{ii}	1	prior precision of left node
Ψ_i	σ^2	prior precision of right node
μ_i	$\frac{G_i}{L_i}$	posterior mean of node
P_i	L_i	posterior precision of node
A_{ij}	$\frac{-js_{ia}}{\sqrt{N}}$	covariance
A_{ji}	$\frac{-js_{ai}}{\sqrt{N}}$	covariance
	j	$j=\sqrt{-1}$

Now we derive Montanari's update rules. We start with the precision message from left to right:

$$\overbrace{\lambda_{i\to a}^{(t+1)}}^{P_{ij}} = 1 + \frac{1}{N}\Sigma_{b\neq a}\frac{s_{ib}^2}{\lambda_{b\to i}^{(t)}} =$$

$$= \overbrace{1}^{P_{ii}} + \Sigma_{b\neq a} \overbrace{\frac{1}{N}\frac{s_{ib}^2}{\lambda_{b\to i}^{(t)}}}^{P_{ki}}$$

$$= \overbrace{1}^{P_{ii}} - \Sigma_{b\neq a} \overbrace{\frac{-js_{ib}}{\sqrt{N}}}^{-A_{ij}} \overbrace{\frac{1}{\lambda_{b\to i}^{(t)}}}^{(P_{j\setminus i})^{-1}} \overbrace{\frac{-js_{ib}}{\sqrt{N}}}^{A_{ji}}.$$

By looking at Table 1, it is easy to verify that this precision update rule is equivalent to that in #2 of Table I.

Using the same logic we get the precision message from right to left:

$$\overbrace{\hat{\lambda}_{i\to a}^{(t+1)}}^{P_{ji}} = \overbrace{\sigma^2}^{P_{ii}} + \overbrace{\frac{1}{N}\Sigma_{k\neq i}\frac{s_{ka}^2}{\lambda_{k\to a}^{(t)}}}^{-A_{ij}^2 P_{j\setminus i}^{-1}}$$

The mean message from left to right is given by

$$\gamma_{i\to a}^{(t+1)} = \frac{1}{N}\Sigma_{b\neq a}\frac{s_{ib}}{\lambda_{b\to i}^{(t)}}\hat{\gamma}_{b\to i}^{(t)} =$$

$$= \overbrace{0}^{\mu_{ii}} - \Sigma_{b\neq a} \overbrace{\frac{-js_{ib}}{\sqrt{N}}}^{-A_{ij}} \overbrace{\frac{1}{\hat{\lambda}_{b\to i}^{(t)}}}^{P_{j\setminus i}^{-1}} \overbrace{\hat{\gamma}_{b\to i}^{(t)}}^{\mu_{j\setminus i}}.$$

The same calculation is done for the mean from right to left:

$$\hat{\gamma}_{i\to a}^{(t+1)} = y_a - \frac{1}{N}\Sigma_{k\neq i}\frac{s_{ka}}{\lambda_{k\to a}^{(t)}}\gamma_{k\to a}^{(t)}$$

Finally, the left nodes calculated the precision and mean by

$$G_i^{(t+1)} = \frac{1}{\sqrt{N}}\Sigma_b\frac{s_{ib}}{\lambda_{b\to i}^{(t)}}\hat{\gamma}_{b\to i}^{(t)}\ ,\ J_i = G_i^{-1}$$

$$L_i^{(t+1)} = 1 + \frac{1}{N}\Sigma_b\frac{s_{ib}^2}{\lambda_{b\to i}^{(t)}}\ ,\ \mu_i = L_i G_i^{-1}.$$

The key difference between the two constructions is that Montanari uses a directed factor graph while we use an undirected graphical model. As a consequence, our construction provides additional convergence results and simpler update rules.

REFERENCES

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

[2] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A: Math. Gen.*, vol. 36, pp. 11 111–11 121, Oct. 2003.

[3] J. P. Neirotti and D. Saad, "Improved message passing for inference in densely connected systems," *Europhys. Lett.*, vol. 71, no. 5, pp. 866–872, Sept. 2005.

[4] ——, "Efficient Bayesian inference for learning in the Ising linear perceptron and signal detection in CDMA," *Physica A*, vol. 365, pp. 203–210, Feb. 2006.

[5] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and statistical neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 700–706, Feb. 2005.

[6] F. Kschischang, B. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[7] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, no. 3, pp. 284–287, Mar. 1974.

[8] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1410–1422, June 2002.

[9] ——, "Correction to 'Joint message-passing decoding of LDPC codes and partial-response channels'," *IEEE Trans. Inform. Theory*, vol. 49, p. 2076, Aug. 2003.

[10] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional ISI channels," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Paris, France, Mar. 2003, pp. 131–134.

[11] O. Shental, N. Shental, A. J. Weiss, and Y. Weiss, "Generalized belief propagation receiver for near-optimal detection of two-dimensional channels with memory," in *Proc. IEEE Information Theory Workshop (ITW)*, San Antonio, Texas, USA, Oct. 2004, pp. 225–229.

[12] O. Shental, N. Shental, and S. Shamai (Shitz), "On the achievable information rates of two-dimensional channels with memory," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Adelaide, Australia, Sept. 2005.

[13] O. Shental, N. Shental, S. Shamai (Shitz), I. Kanter, A. J. Weiss, and Y. Weiss, "Finite-state input two-dimensional Gaussian channels with memory: Estimation and information via graphical models and statistical mechanics," *IEEE Trans. Inform. Theory*, vol. 54, pp. 1500 – 1513, April 2008.

[14] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Punta del Este, Uruguay, Mar. 2006.

[15] C. C. Wang and D. Guo, "Belief propagation is asymptotically equivalent to MAP detection for sparse linear systems," in *Proc. 44th Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sept. 2006.

[16] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[17] J. G. Proakis, *Digital Communications*, 4th ed. New York, USA: McGraw-Hill, 2000.

[18] B. R. Vojčić and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.

[19] D. Bickson, O. Shental, P. H. Siegel, J. K. Wolf, and D. Dolev, "Linear detection via belief propagation," in *Proc. 45th Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sept. 2007.

[20] O. Shental, D. Bickson, P. H. Siegel, J. K. Wolf, and D. Dolev, "Gaussian belief propagation solver for systems of linear equations," in *IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, July 2008.

[21] A. Grant and C. Schlegel, "Iterative implementations for linear multiuser detectors," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1824–1834, Oct. 2001.

[22] P. H. Tan and L. K. Rasmussen, "Linear interference cancellation in CDMA based on iterative techniques for linear equation systems," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2099–2108, Dec. 2000.

[23] A. Yener, R. D. Yates, , and S. Ulukus, "CDMA multiuser detection: A nonlinear programming approach," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1016–1024, June 2002.

[24] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multi-user detection," in *Proc. 43th Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sept. 2005.

[25] Gaussian Belief Propagation implementation in matlab [online] `http://www.cs.huji.ac.il/labs/danss/p2p/gabp/`.

[26] B. J. Frey, "Local probability propagation for factor analysis," in *Neural Information Processing Systems (NIPS)*, Denver, Colorado, Dec 1999, pp. 442–448.

[27] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.

[28] J. K. Johnson, D. M. Malioutov, and A. S. Willsky, "Walk-sum interpretation and analysis of Gaussian belief propagation," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2006, pp. 579–586.