Linear Detection via Belief Propagation

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

Ori Shental*, Paul H. Siegel and Jack K. Wolf
Center for Magnetic Recording Research
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093, USA
Email: {oshental,psiegel,jwolf@ucsd.edu}

***Abstract*— In this paper, the paradigm of linear detection is reformulated as a Gaussian belief propagation (GaBP) scheme, without resorting to direct matrix inversion. The derived iterative framework allows for a distributive message-passing implementation of this important class of sub-optimal tractable estimators. The properties of GaBP-based linear detection are addressed, while its faster convergence, in comparison with conventional iterative solution methods, is demonstrated experimentally.**

I. Introduction

Belief propagation (BP, a.k.a. sum-product algorithm) message-passing 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. In this contribution, the analogy between message-passing inference and estimation is further strengthened, unveiling a surprising link among disciplines.

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 cancelation (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.

*Contributed equally to this work.

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].

In spite of its significant role in estimation theory, linear detection has never been explicitly linked to BP, in constrast to optimal MAP detection and several sub-optimal nonlinear detection techniques. In this paper, we reformulate the general linear detector as a Gaussian belief propagation (GaBP) algorithm. This message-passing framework is not limited to the large-system limit and is suitable for channels with arbitrary prior input distribution. Revealing this missing link allows for a distributed implementation of the linear detector, circumventing the necessity of, potentially cumbersome, direct matrix inversion (via, *e.g.*, Gaussian elimination).

The exactness and convergence properties of the GaBP-based linear detector are addressed using the seminal work of Weiss and Freeman [19] and related recent developments [20], [21]. The derived iterative framework is 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 [22]–[24]. GaBP is shown to yield faster convergence than these standard methods. The BP-based MUD, recently derived and analyzed by Montanari *et al.* [25] for Gaussian input symbols, is an instance of our framework. Finally, GaBP convergence is further accelerated by adopting the methods of Aitken and Steffensen [26]. Although the current contribution is oriented

to estimation-theoretic problems and applications from the field of communications, we would like to emphasize its pertinence to the more universal problems of efficient distributed matrix inversion, solution of systems of linear equations and determinant calculation.

The paper is organized as follows. Section II formulates the problem of linear detection. In Section III we derive the distributive GaBP-based linear detection scheme, while its superior convergence rate, w.r.t. some conventional iterative methods, is discussed in the experimental study of Section IV. 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.

II. Linear Detection

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 $\mathbf{y} = \{y_1, \ldots, y_K\}^T = f\{\mathbf{x}^T\} \in \mathbb{R}^K$.[1] 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

$$\mathbf{y} = f\{\mathbf{x}\} = \mathbf{R}\mathbf{x} + \mathbf{n}, \tag{2}$$

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 $\mathbf{S}$.

Due to the vast applicability of linear channels, in Section IV we focus in our experimental study on such channels, although our paradigm is not limited to this case. 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 the following section we implement the linear detection operation, in its general form (1), in an efficient message-passing fashion.

III. GaBP and Linear Detection

As stated in the previous section, our aim is to find $\mathbf{x}^*$, a solution to the linear equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, *i.e.*, $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$, without actually inverting the non-singular matrix $\mathbf{A}$. Another way of solving this set of linear equations $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0}$ is to represent it using a quadratic form $q(\mathbf{x}) \triangleq \mathbf{x}^T\mathbf{A}\mathbf{x}/2 - \mathbf{b}^T\mathbf{x}$. As the matrix $\mathbf{A}$ is symmetric[2] (*e.g.*, $\mathbf{A} = \mathbf{S}^T\mathbf{S}$), the derivative of the quadratic form With respect to the vector $\mathbf{x}$ is given by $\mathbf{q}'(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$.

Thus, equating $\mathbf{q}'(\mathbf{x}) = \mathbf{0}$ gives the global minimum $\mathbf{x}^*$ of this convex function, which is nothing but the desired solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Now, one can define the following *jointly Gaussian* distribution

$$p(\mathbf{x}) \triangleq \mathcal{Z}^{-1} \exp\big(-q(\mathbf{x})\big) = \mathcal{Z}^{-1} \exp\left(-\mathbf{x}^T\mathbf{A}\mathbf{x}/2 + \mathbf{b}^T\mathbf{x}\right), \tag{3}$$

where $\mathcal{Z}$ is a distribution normalization factor. Defining the vector $\mu \triangleq \mathbf{A}^{-1}\mathbf{b}$, one gets the form

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{Z}^{-1} \exp\left(\mu^T\mathbf{A}\mu/2\right) \\ &\times \exp\left(-\mathbf{x}^T\mathbf{A}\mathbf{x}/2 + \mu^T\mathbf{A}\mathbf{x} - \mu^T\mathbf{A}\mu/2\right) \\ &= \zeta^{-1} \exp\big(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{A}(\mathbf{x}-\mu)\big) \\ &= \mathcal{N}(\mu, \mathbf{A}^{-1}), \end{aligned} \tag{4}$$

where the new normalization factor $\zeta \triangleq \mathcal{Z} \exp\left(-\mu^T\mathbf{A}\mu/2\right)$. To summarize to this point, the target solution $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$ is equal to $\mu \triangleq \mathbf{A}^{-1}\mathbf{b}$, which is the mean vector of the distribution $p(\mathbf{x})$ (3), as defined above.

The formulation above allows us to shift the linear detection problem from an algebraic to a probabilistic domain. Instead of solving a deterministic vector-matrix linear equation, we now solve an inference problem in a graphical model describing a certain Gaussian distribution function. Given the overall channel matrix $\mathbf{A}$ and the observation vector $\mathbf{b}$, one knows how to write explicitly $p(\mathbf{x})$ (3) and the corresponding graph $\mathcal{G}$ with edge potentials (compatibility functions) ψ_{ij} and self-potentials ('evidence') ϕ_i. These graph potentials are

[1] An extension to the complex domain is straightforward.

[2] For a non-symmetric matrix $\mathbf{A}$ an approximation of the solution $\mathbf{x}^*$ is inferred.

#	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 $P_{i\setminus j} = P_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} P_{ki}$ and $\mu_{i\setminus j} = P_{i\setminus j}^{-1}(P_{ii}\mu_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} P_{ki}\mu_{ki})$. Compute $P_{ij} = -A_{ij}P_{i\setminus j}^{-1}A_{ji}$ and $\mu_{ij} = -P_{ij}^{-1}A_{ij}\mu_{i\setminus j}$.
3.	*Check*	If P_{ij} and μ_{ij} did not converge, return to #2. Else, continue to #4.
4.	*Infer*	$P_i = P_{ii} + \sum_{k\in\mathcal{N}(i)} P_{ki}$, $\mu_i = P_i^{-1}(P_{ii}\mu_{ii} + \sum_{k\in\mathcal{N}(i)} P_{ki}\mu_{ki})$.
5.	*Decide*	$\hat{x}_i = \Delta\{\mu_i\}$

TABLE I
COMPUTING $\mathbf{A}^{-1}\mathbf{b}$ VIA GABP.

determined according to the following pairwise factorization of the Gaussian distribution $p(\mathbf{x})$ (3)

$$p(\mathbf{x}) \propto \prod_{i=1}^{K} \phi_i(x_i) \prod_{\{i,j\}} \psi_{ij}(x_i, x_j), \tag{5}$$

resulting in $\psi_{ij}(x_i, x_j) \triangleq \exp(-x_i A_{ij} x_j)$ and $\phi_i(x_i) = \exp\left(b_i x_i - A_{ii}x_i^2/2\right)$. 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, $p(x_i) \sim \mathcal{N}(\mu_i = \{\mathbf{A}^{-1}\mathbf{b}\}_i, P_i^{-1} = \{\mathbf{A}^{-1}\}_{ii})$, where μ_i and P_i are the marginal mean and inverse variance (a.k.a. precision), respectively. Recall that, according to our previous argumentation, the inferred mean μ_i is identical to the desired solution x_i^*.

The move to the probabilistic domain calls for the utilization of BP as an efficient inference engine. The sum-product rule of BP for *continuous* variables, required in our case, is given by [19]

$$m_{ij}(x_j) = \alpha \int_{x_i} \psi_{ij}(x_i, x_j)\phi_i(x_i) \prod_{k\in\mathcal{N}(i)\setminus j} m_{ki}(x_i)dx_i, \tag{6}$$

where $m_{ij}(x_j)$ is the message sent from node i to node j over their shared edge on the graph, scalar α is a normalization constant and the set $\mathcal{N}(i)\setminus j$ denotes all the nodes neighboring x_i, except x_j. The marginals are computed according to the product rule [19]

$$p(x_i) = \alpha\phi_i(x_i) \prod_{k\in\mathcal{N}(i)} m_{ki}(x_i). \tag{7}$$

GaBP is a special case of continuous BP where the underlying distribution is Gaussian. Next, we derive the GaBP update rules by substituting Gaussian distributions in the continuous BP equations.

According to the right hand side of the sum-product rule (6), node i needs to calculate the product of all incoming messages, except for the message coming from node j. Recall that since $p(\mathbf{x})$ is jointly Gaussian, the self-potentials $\phi_i(x_i)$ and the messages $m_{ki}(x_i)$ are Gaussians as well. The product of Gaussians of the *same* variable is also a Gaussian. Consider the Gaussians defined by $\mathcal{N}(\mu_1, P_1^{-1})$ and $\mathcal{N}(\mu_2, P_2^{-1})$. Their product is also a Gaussian $\mathcal{N}(\mu, P^{-1})$ with

$$\mu = P^{-1}(P_1\mu_1 + P_2\mu_2), \tag{8}$$
$$P^{-1} = (P_1 + P_2)^{-1}. \tag{9}$$

As the terms in the product of the incoming messages and the self-potential are all describing the same variable, x_i, we can use this property to demonstrate that $\phi_i(x_i)\prod_{k\in\mathcal{N}(i)\setminus j} m_{ki}(x_i)$ is proportional to a $\mathcal{N}(\mu_{i\setminus j}, P_{i\setminus j}^{-1})$ distribution. Therefore, the update rule for the inverse variance is given by (over-braces denote the origin of these terms)

$$P_{i\setminus j} = \overbrace{P_{ii}}^{\phi_i(x_i)} + \sum_{x_k\in\mathcal{N}(i)\setminus j} \overbrace{P_{ki}}^{m_{ki}(x_i)}, \tag{10}$$

where $P_{ii} \triangleq A_{ii}$ is the inverse variance associated with node i, via $\phi_i(x_i)$, and P_{ki} are the inverse variances of the messages $m_{ki}(x_i)$. Equivalently, we can calculate the mean

$$\mu_{i\setminus j} = P_{i\setminus j}^{-1}\Big(\overbrace{P_{ii}\mu_{ii}}^{\phi_i(x_i)} + \sum_{k\in\mathcal{N}(i)\setminus j} \overbrace{P_{ki}\mu_{ki}}^{m_{ki}(x_i)}\Big), \tag{11}$$

where $\mu_{ii} \triangleq b_i/A_{ii}$.

Now, we calculate the remaining terms of the message $m_{ij}(x_j)$, including the integration over x_i. After some algebraic manipulations, we use the Gaussian integral $\int_{-\infty}^{\infty} \exp(-ax^2 - bx)dx = \sqrt{\pi/a}\exp(b^2/4a)$, to show that $m_{ij}(x_j)$ is a normal distribution with mean and precision given by

$$\mu_{ij} = -P_{ij}^{-1}A_{ij}\mu_{i\setminus j}, \tag{12}$$
$$P_{ij} = -A_{ij}P_{i\setminus j}^{-1}A_{ji}. \tag{13}$$

These two scalars are the propagating messages in the GaBP scheme. Finally, the computation of the product rule (7) is similar to our previous calculations (10)-(11), but with no incoming messages excluded. The GaBP-based implementation of the linear detection operation is summarized in Table 1.

The algorithm in Table 1 can be easily distributed. Each node i receives as an input the i'th row (or column) of the matrix $\mathbf{A}$ and the scalar b_i. In each iteration, a message

containing two reals, μ_{ij} and P_{ij}, is sent to every neighboring node through their mutual edge, corresponding to a non-zero A_{ij} entry. For a dense matrix $\mathbf{A}$, each of the K nodes sends a unique message to every other node on the fully-connected graph, which results in a total of K^2 messages per iteration round.

The number of messages passed on the graph can be reduced significantly, down to only K messages per round.[3] Instead of sending a message composed of the pair μ_{ij} and P_{ij}, a node can broadcast the aggregated sums

$$\tilde{P}_i = P_{ii} + \sum_{k \in \mathcal{N}(i)} P_{ki}, \quad (14)$$

$$\tilde{\mu}_i = P_i^{-1}(P_{ii}\mu_{ii} + \sum_{k \in \mathcal{N}(i)} P_{ki}\mu_{ki}). \quad (15)$$

Now, each node locally retrieves the $P_{i\setminus j}$ (10) and $\mu_{i\setminus j}$ (11) from the sums by means of a subtraction

$$P_{i\setminus j} = \tilde{P}_i - P_{ji}, \quad (16)$$

$$\mu_{i\setminus j} = \tilde{\mu}_i - P_{i\setminus j}^{-1} P_{ji}\mu_{ji}. \quad (17)$$

The rest of the algorithm remains the same.

IV. Results and Discussion

If it converges, GaBP is known to result in exact inference [19]. In contrast to conventional iterative methods for the solution of systems of linear equations, for GaBP, determining the exact region of convergence and convergence rate remain open research problems. All that is known is a sufficient (but not necessary) condition [20], [21] stating that GaBP converges when the spectral radius[4] satisfies $\rho(|\mathbf{I}_K - \mathbf{A}|) < 1$. A stricter sufficient condition [19], actually proved earlier, determines that the matrix $\mathbf{A}$ must be diagonally dominant (*i.e.*, $|A_{ii}| > \sum_{j \neq i} |A_{ij}|, \forall i$) in order for GaBP to converge.

As these conditions are not necessary, one can find examples of channels for which the condition does not hold, yet GaBP still converges perfectly to the linear detection solution. For instance, in the case of Gaussian input signaling, *i.e.*, $P_{\mathbf{x}}$ is a normal distribution, for which linear detection becomes optimal, it can be easily shown that the proposed GaBP scheme boils down to the BP-based MUD scheme, recently introduced by Montanari *et al.* [25]. Their BP scheme, tailored for Gaussian signaling, has been proven to converge to the MMSE (and optimal) solution for any arbitrarily loaded (*i.e.*, $\rho(|\mathbf{I}_K - \mathbf{R}|) \lesseqgtr 1$) randomly-spread CDMA system.[5] Thus, Gaussian-input AWGN-CDMA is an example where the proposed GaBP scheme converges to the MAP decision for any $N \times K$ random spreading matrix $\mathbf{S}$.

In the following experimental study, we show that, when it converges, GaBP is substantially faster than alternative iterative methods. We examined two system setups of binary signaling synchronous CDMA with cross-correlation matrices

$$\mathbf{R}_3 = \frac{1}{7}\begin{pmatrix} 7 & -1 & 3 \\ -1 & 7 & -5 \\ 3 & -5 & 7 \end{pmatrix} \quad (18)$$

and

$$\mathbf{R}_4 = \frac{1}{7}\begin{pmatrix} 7 & -1 & 3 & 3 \\ -1 & 7 & 3 & -1 \\ 3 & 3 & 7 & -1 \\ 3 & -1 & -1 & 7 \end{pmatrix} \quad (19)$$

for $K = 3$ and $K = 4$ users, respectively. These correlation profiles are created by using Gold spreading sequences of length $N = 7$. These particular settings were taken from the simulation setup of Yener *et al.* [24]. Note that $\mathbf{R}_3$ and $\mathbf{R}_4$ are not diagonally dominant, but their spectral radii are less than unity, namely $\rho(|\mathbf{I}_3 - \mathbf{R}_3|) = 0.9008 < 1$ and $\rho(|\mathbf{I}_4 - \mathbf{R}_4|) = 0.8747 < 1$, respectively. Hereafter, the iterative methods being compared, including GaBP, implement a decorrelator ($\mathbf{R}^{-1}\mathbf{y}$) detector in a noiseless channel.

Table II compares the proposed GaBP algorithm with standard iterative solution methods [28] (using random initial guesses), previously employed for CDMA MUD. Multiuser detectors [22], [23] based on the algorithms of Jacobi,

Algorithm	$\mathbf{R}_3$	$\mathbf{R}_4$
Jacobi	136	50
GS	27	32
GaBP	**23**	**24**
SOR	18	20
Serial GaBP	**16**	**13**

TABLE II
Convergence rate (in iterations) of GaBP vs. standard methods (in the two simulated CDMA systems).

Algorithm	$\mathbf{R}_3$	$\mathbf{R}_4$
Jacobi+Aitkens	46	33
Jacobi+Steffensen[6]	51	–
GaBP+Steffensen	**13**	**13**
Serial GaBP+Steffensen	**9**	**7**

TABLE III
Comparison under acceleration methods.

[3] By using a similar construction to Bickson *et al.* [27].

[4] $\rho(\mathbf{B}) \triangleq \max_{1 \leq i \leq s}(|\lambda_i|)$, where $\lambda_1, \ldots \lambda_s$ are the eigenvalues of a matrix $\mathbf{B}$.

[5] For non-Gaussian signaling, *e.g.* binary, this BP-based detector is conjectured to converge only in the large-system limit, *i.e.*, as $K, N \to \infty$ [25].

[6] For $K = 4$, the Jacobi algorithm employing Steffensen acceleration converged only in part of the simulation rounds, so the number of required iterations is marked by '$-$'.

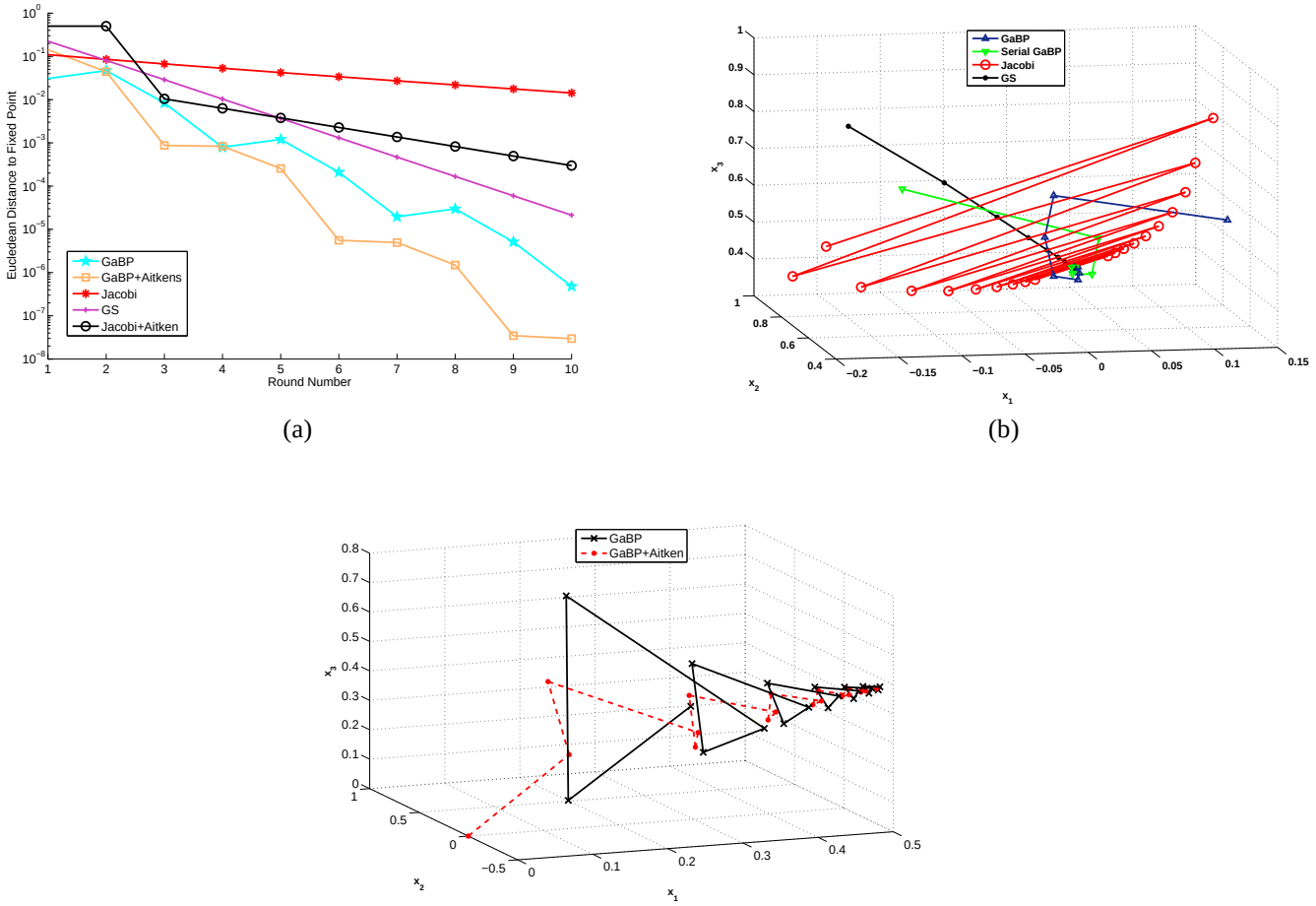

Fig. 1. (a) Convergence rate: Euclidean distance as a function of iteration rounds for GaBP and other algorithms. (b) Convergence visualization for GaBP vs. other algorithms. (c) Convergence visualization of GaBP vs. GaBP using Aitken's acceleration method.

Gauss-Seidel (GS) and (optimally weighted) successive over-relaxation (SOR)[7] were investigated. The table lists the convergence rates for the two Gold code-based CDMA settings. Convergence is identified and declared when the differences in all the iterated values are less than 10^{-6}. Clearly, GaBP yields faster convergence speed on both examined systems in comparison with the renowned Jacobi and GS algorithms. The best convergence rate, with respect to the conventional iterative methods, including SOR, is achieved for serial GaBP, *i.e.*, the proposed scheme with serial, rather than parallel (flooding), message-passing update rule.

Further speed-up of GaBP can be achieved by adopting known acceleration techniques, like Aitken's method and Steffensen's iterations [26], yet to be employed with BP schemes. Consider a sequence $\{x_n\}$ (*e.g.*, obtained by using GaBP iterations) linearly converging to the limit $\hat{x}$, and $x_n \neq \hat{x}$ for $n \geq 0$. According to Aitken's method, if there exists a real number a such that $|a| < 1$ and $\lim_{n\to\infty}(x_n - \hat{x})/(x_{n-1} - \hat{x}) = a$, then the sequence $\{y_n\}$ defined by

$$y_n = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}$$

converges to $\hat{x}$ faster than $\{x_n\}$ in the sense that $\lim_{n\to\infty} |(\hat{x} - y_n)/(\hat{x} - x_n)| = 0$. Aitken's method can be viewed as a generalization of over-relaxation, since one uses values from three, rather than two, consecutive iteration rounds. This method can be easily implemented in GaBP as every node computes values based only on its own history.

Steffensen's iterations encapsulate Aitken's method. Starting with x_n, two iterations are run to get x_{n+1} and x_{n+2}. Next, Aitken's method is used to compute y_n, this value is replaced with the original x_n, and GaBP is executed again to get a new value of x_{n+1}. This process is repeated iteratively until convergence. Table III demonstrates the speed-up of GaBP using these acceleration methods, in comparison with the modified Jacobi algorithm.[8]

[7]This moving average improvement of Jacobi and GS algorithms is equivalent to what is known in the BP literature as 'damping' [29].

[8]Application of Aitken and Steffensen's methods for speeding-up the convergence of standard (non-BP) iterative solution algorithms in the context of MUD was introduced by Leibig *et al.* [30].

Fig. 1-(a) displays the Euclidean distance between the tentative (intermediate) results and the fixed-point solution as a function of the iteration rounds, for the algorithms we examined. As expected, all linear algorithms exhibit a logarithmic convergence behaviour. Note that GaBP converges faster on average, although there are some fluctuations in the GaBP curves, in contrast to the monotonicity of the other curves.

An interesting question concerns the origin of this convergence speed-up associated with GaBP. Better understanding may be gained by visualizing the iterations of the different methods for the matrix $\mathbf{R}_1$ case. The convergence contours are plotted in the space of $\{x_1, x_2, x_3\}$ in Fig. 1-(b). As expected, the Jacobi algorithm converges in zigzags towards the fixed point (this behavior is well-explained in Bertsekas and Tsitsiklis [31]). The fastest algorithm is serial GaBP. It is interesting to note that GaBP convergence is in a spiral shape, hinting that despite the overall convergence improvement, performance improvement is not guaranteed in successive iteration rounds. The spiral nature of GaBP convergence is better viewed in Fig. 1-(c). In this case the system was simulated with a certain $\mathbf{R}$ matrix for which Jacobi algorithm and other standard methods did not even converge. Using Aitken's method a further speed-up in GaBP convergence is obtained.

Despite the fact that we are using examples of small systems, we believe that these examples capture the typical behavior of the various algorithms. Note, in passing, that GaBP was experimentally shown to converge in a logarithmic number of iterations in the cases of very large matrices both dense (with up to hundreds of thousands of dimensions [32]) and sparse (with up to millions of dimensions [33]).

ACKNOWLEDGMENT

The work of OS, PS and JW is supported in part by NSF Grant No. CCR-0514859. DB has been partially supported by EVERGROW, IP 1935 of the EU Sixth Framework. Part of the work of DD was done while visiting Cornell university and was funded in part by ISF, ISOC, NSF, CCR, and AFOSR. The authors wish to thank A. Montanari for sharing the source code for the algorithms in [34] and J. K. Johnson for valuable comments on the manuscript.

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.

[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*, submitted for publication.

[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] 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.

[20] 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.

[21] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *Journal of Machine Learning Research*, vol. 7, Oct. 2006.

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

[23] 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.

[24] 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.

[25] 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.

[26] P. Henrici, *Elements of Numerical Analysis*. John Wiley and Sons, 1964.

[27] D. Bickson, D. Dolev, and Y. Weiss, "Modified belief propagation for energy saving in wireless and sensor networks," in *Leibniz Center TR-2005-85, School of Computer Science and Engineering, The Hebrew University*, 2005. [Online]. Available: http://leibniz.cs.huji.ac.il/tr/842.pdf

[28] O. Axelsson, *Iterative Solution Methods*. Cambridge, UK: Cambridge University Press, 1994.

[29] K. M. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. of UAI*, 1999.

[30] C. Leibig, A. Dekorsy, and J. Fliege, "Power control using Steffensen iterations for CDMA systems with beamforming or multiuser detection,"

in *Proc. IEEE International Conference on Communications (ICC)*, Seoul, Korea, 2005.

[31] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Calculation. Numerical Methods.* Prentice Hall, 1989.

[32] D. Bickson, D. Dolev, and E. Yom-Tov, "Solving large scale SVMs using Gaussian belief propagation," in *Leibniz Center TR-2007-117, School of Computer Science and Engineering, The Hebrew University*, 2007. [Online]. Available: http://leibniz.cs.huji.ac.il/tr/1050.pdf

[33] D. Bickson, D. Malkhi, and L. Zhou, "Peer-to-Peer rating," in *7th IEEE P2P computing*, Galway, Ireland, 2007.

[34] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multi–user detection," 2005. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0510044