Bayesian QAM demodulation and activity detection for multiuser communication systems

Gabor Hannak, Martin Mayer, Gerald Matz, Norbert Goertz
Institute of Telecommunications
TU Wien, Vienna, Austria
Email: {ghannak,mmayer,gmatz,ngoertz}@nt.tuwien.ac.at

***Abstract*—We consider overloaded (non-orthogonal) multiple access multiuser wireless communication systems with many transmitting devices and one central aggregation node, a typical scenario in e.g. machine-to-machine communications. The task of the central node is to detect the set of active devices and to separate and detect their data streams, whose number at any time instance is small compared to the total number of devices in the system. The payload bits are mapped to a quadrature amplitude modulation (QAM) symbol alphabet, transmitted by the active devices and received synchronously at the central node. The data detection can be cast as a compressed sensing (CS) problem due to the sparsity granted by the sporadic transmission of the typically low-complexity nodes. Separation of the real and imaginary parts of the measurement matrix, the unknown QAM symbols, and the received signal yields a group-sparsity problem. We utilize an efficient iterative Bayesian CS recovery scheme which, instead of separately solving for the real and imaginary parts, uses the Turbo principle to exchange and update parameters between the two solvers and thus comes to consensus regarding the sparsity structure. By tailoring this algorithm to QAM detection, joint activity detection, demodulation and data detection with high reliability is possible, even for very large-scale systems.**

I. Introduction

A. Compressed Sensing in Multiuser Detection

Wireless multiuser communication systems, and specifically machine-to-machine (M2M) communications, have received much attention in literature recently since the number of autonomously communicating devices is expected to grow tremendously [2]. Many practical scenarios can be represented by a star topology, in which a multitude of battery-driven and/or low-complexity devices communicate with one central aggregation node. In the uplink transmission, when the central node collects data from the multitude of spatially distributed devices, several multiple access schemes are available (time/code/frequency division multiple access) for simultaneous transmission such that the receiver is still able to separate the data streams. However, with the growing number of devices in one system, and due to the fact that typically only a small fraction of all devices transmit at the same time, orthogonal multiple access schemes introduce a significant overhead. Applications involving low-complexity and possibly battery-driven devices call for a low-overhead, simple transmission protocol and efficient signal processing methods with minimal cooperation between the users. We address this demand using the recent theory of compressed sensing (CS), and in particular Bayesian compressed sensing [11], [12].

CS allows us to solve underdetermined systems of linear equations with additional sparsity constraints on the unknown vector. For an overview we refer to [8] and references therein. It has been shown that CS methods can be applied successfully to achieve reliable joint activity *and* data detection for overloaded multiple access systems, i.e., when the spreading sequences are not orthogonal because their chip length is smaller than the number of users [6], [21]. Recently, a probabilistic approach for solving underdetermined systems of linear equations has been helpful in many areas: via the Bayesian formulation the signal prior distribution can be explicitly and optimally incorporated, thus not only exploiting sparsity, but also a more specific signal structure (e.g. the probability distribution of the nonzero components). Since in digital communications the signal vector entries are often constrained to a finite alphabet, utilizing this prior information is highly beneficial.

B. Existing Work

In [21], Zhu and Giannakis introduced a series of methods for data detection in sparse multiuser scenarios exploiting both sparsity and the finite alphabet constraint imposed on the data. These methods rely on solving a modified least absolute shrinkage and selection operator (LASSO) with lattice search and sphere decoding, limiting the problem dimension. In [20] the identification of the set of active users in a code division multiple access (CDMA) system is solved by a two stage algorithm. A subspace-based preprocessing determines the number of active users and is followed by the *multiple signal classification* algorithm [17] which identifies them. Schepker et al. investigate joint activity and data detection in the overloaded CDMA scenario in [16]. The authors formulate the CS interpretation of the detection and use orthogonal matching pursuit (OMP) to identify the active users and reconstruct the transmitted data. However, for larger problems, i.e., growing number of devices, group orthogonal matching pursuit (GOMP) becomes infeasible and has to be partitioned into subproblems, which introduces errors due to neglecting the multiple access interference. CDMA asynchronicity is considered in [15]. For every device in the system, the authors' method introduces a certain number of virtual nodes, which correspond to candidate delay lengths. Out of the virtual nodes, only one, the one corresponding to the actual delay, transmits, and this again can be solved by a slightly modified GOMP. The authors of [18] combine CS detection with a classical detection method to cope with varying sparsity in the transmitting devices. A Bayesian approach based on random set theory is presented and analyzed in [4]. Simultaneous activity detection and channel estimation using the approximate message passing (AMP) framework [9] is elaborated in [13].

C. Contribution

This work shows that performing device activity detection and data demodulation/detection simultaneously is strongly improved by exploiting the prior knowledge of the symbol alphabet (quadrature amplitude modulation (QAM)) and the node activity probability in an overloaded multiple access system. The formulation of the signal acquisition as a CS problem with known prior and group sparsity constraint allows the use of a slightly modified Bayesian-optimal approximate message passing (BAMP) [11], [12], a prominent recovery algorithm for underdetermined systems of linear equations. The the authors' best knowledge, applying BAMP and tailoring it to the QAM setting in the random access framework is novel. BAMP is computationally efficient as its steps do not involve matrix inversions, only multiplications and additions. This renders the computation extremely fast and applicable for very large problem dimensions (e.g. thousands of users). The modification of the original algorithm inherently solves the classification of nodes into active and inactive states, thus eliminating the need for the prior knowledge of the exact number of active nodes.

This paper is organized as follows. In Section II we formulate the multiuser detection problem as a system of linear equations, introduce the probabilistic formulation for known symbol alphabets and known node activity probability and present the BAMP algorithm. In Section III we present an extension of the BAMP algorithm that is suited for the complex channel and the QAM signal model. Section IV shows numerical simulations, and we conclude in Section V.

Notation. Uppercase (lowercase) boldface letters denote matrices (column vectors). For a matrix $\mathbf{A}$ (vector $\mathbf{a}$), $\mathbf{A}_s$ ($\mathbf{a}_s$) denotes its sth column (sth entry). Random variables and vectors are written in sans serif fonts such as x and $\mathsf{\mathbf{x}}$. The identity matrix of dimension M is denoted by $\mathbf{I}_M$. The superscripts $(\cdot)^{\mathcal{R}}$ and $(\cdot)^{\mathcal{I}}$ indicate the (component-wise) real and imaginary parts, respectively. The Dirac delta function is $\delta(\cdot)$, and $(\mathcal{C})\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\mathbf{C})$ denotes the (complex) normal probability density function (pdf) with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ evaluated at $\mathbf{x}$.

II. Problem Formulation

A. System Model

We consider a multiple access wireless communication system with star topology, i.e., N devices (denoted by $1,\ldots,N$) communicate with a central aggregation node C. We address the uplink scenario, where a subset of the N devices transmits simultaneously to C. We discretize time into frames: within each frame, each device is either active and transmits within the duration of a complete frame, or is inactive (and does not transmit). The nodes transmit simultaneously, where potential asynchronicity is accounted for by the channel impulse response. We consider only one frame in the following. Let us denote the set of active users by $\mathcal{A}$ and the set of inactive users by $\mathcal{I}$, where $\mathcal{A}\cup\mathcal{I}=\{1,\ldots,N\}$. An active user is transmitting $\log_2 M$ bits in a frame, that are Gray-mapped onto a M-QAM symbol:

$$\tilde{x}_n \in \tilde{\mathcal{M}}_M \quad \forall n \in \mathcal{A}\,,$$

where $\tilde{\mathcal{M}}_M=\{m_1,\ldots,m_M\}$ denotes the M-QAM symbol alphabet normalized to average symbol energy 1. The M-QAM symbols are composed of a real (in-phase) and an imaginary (quadrature phase) part, both taking values from the same real-valued alphabet:

$$m_q = a + jb\,, \quad a,b\in\mathcal{M}_M{=}\frac{1}{Z}\{\pm1,\pm3,\ldots,\pm(\sqrt{M}-1)\}$$

with Z being a normalization factor depending on the modulation order and $\mathcal{M}_M$ being the set of real values for the real/imaginary parts of the complex-valued symbol alphabet $\tilde{\mathcal{M}}_M$. In practice and in our simulations, $M\in\{4,16,64,256\}$. The inactive users are not transmitting, or equivalently they are modeled by transmitting a 0 pseudo-symbol, i.e.,

$$\tilde{x}_n = 0 \quad \forall n\in\mathcal{I}\,.$$

For now, we assume only one transmit symbol per frame to keep the theoretical description simple. The discussion for longer frames is topic of future work. If the transmit symbols of all users $\tilde{x}_n$ are stacked into $\tilde{\mathbf{x}}=(\tilde{x}_1,\ldots,\tilde{x}_N)^T$, the canonical input-output relationship of the transmission can be cast as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}\,, \tag{1}$$

where $\tilde{\mathbf{A}}\in\mathbb{C}^{(K\times N)}$, termed *measurement matrix*, subsumes the transmit and receive filters, the complex channel impulse responses, and the multiple access scheme signature sequences [21]; and $\tilde{\mathbf{w}}\in\mathbb{C}^K$ is circularly-symmetric complex additive white Gaussian noise (AWGN). In orthogonal systems, (1) is well determined. However, in very large communication systems one has to fall back to the underdetermined design, i.e., $K<N$, which can be handled by the CS framework. We can write (1) equivalently as

$$\underbrace{\begin{pmatrix}\tilde{\mathbf{y}}^{\mathcal{R}}\\ \tilde{\mathbf{y}}^{\mathcal{I}}\end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix}\tilde{\mathbf{A}}^{\mathcal{R}} & -\tilde{\mathbf{A}}^{\mathcal{I}}\\ \tilde{\mathbf{A}}^{\mathcal{I}} & \tilde{\mathbf{A}}^{\mathcal{R}}\end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix}\tilde{\mathbf{x}}^{\mathcal{R}}\\ \tilde{\mathbf{x}}^{\mathcal{I}}\end{pmatrix}}_{\mathbf{x}} + \underbrace{\begin{pmatrix}\tilde{\mathbf{w}}^{\mathcal{R}}\\ \tilde{\mathbf{w}}^{\mathcal{I}}\end{pmatrix}}_{\mathbf{w}}\,. \tag{2}$$

The new problem dimensions are now $\mathbf{A}\in\mathbb{R}^{K'\times N'}$, $\mathbf{y},\mathbf{w}\in\mathbb{R}^{K'}$ and $\mathbf{x}\in\mathbb{R}^{N'}$, with $N'=2N$ and $K'=2K$. Observe that $\mathbf{x}$ obeys group structure, since the real and imaginary parts of $\tilde{x}_n$ are either *both* zero or *both* nonzero. That is,

$$\tilde{x}_n^{\mathcal{R}}\neq 0 \Leftrightarrow \tilde{x}_n^{\mathcal{I}}\neq 0\,,$$

or

$$x_n\neq 0 \Leftrightarrow x_{n+N}\neq 0 \quad n\in\{1,\ldots,N\}\,.$$

For illustration consider the following example of a CDMA system: each user is assigned a binary chip sequence $\mathbf{s}^{(n)}\in\{-b,+b\}^{L_S}$ of length L_S, such that the transmitted signal is $\tilde{x}_n\mathbf{s}^{(n)}$. Further, the channel impulse response from user n to the receiver is written in the vector $\mathbf{h}^{(n)}$, thus the n-th column of $\tilde{\mathbf{A}}$ is the convolution of $\mathbf{s}^{(n)}$ and $\mathbf{h}^{(n)}$:

$$\tilde{\mathbf{A}}_n = \mathbf{h}^{(n)} * \mathbf{s}^{(n)}\,.$$

The entries of $\mathbf{A}$ are realizations of random variables, whose distributions are determined by the wireless channel characteristics and the chosen multiplexing scheme, and assumed to be i.i.d. Furthermore, we assume the knowledge of an approximate probability of node activity (and thus transmission) in a certain frame, which we denote by $\gamma^{(0)}$.

B. Theoretical Background: Bayesian Formulation

The entries of the unknown symbol vector $\tilde{\mathbf{x}}$ come from the augmented symbol alphabet $\tilde{\mathcal{M}}_M \cup \{0\}$. Since we model both the user activity and the nodes' symbol selection as i.i.d. random, $\tilde{\mathsf{x}}$ is a random variable with separable probability mass function (pmf)

$$p_{\tilde{\mathsf{x}}}(\tilde{\mathbf{x}}) = \prod_{n=1}^{N} p_{\tilde{\mathsf{x}}_n}(\tilde{x}_n)\,.$$

The variables are i.i.d. over the indices, but as defined in Section II-A by means of the symbol alphabet $\tilde{\mathcal{M}}_M$, the real and imaginary parts are not independent: either both are zero, or both are nonzero. Thus, the real-valued representation reads

$$\begin{aligned}p_{\tilde{\mathsf{x}}_{n'}}(\tilde{x}) =& p_{\tilde{\mathsf{x}}^{\mathcal{R}}_{n'},\tilde{\mathsf{x}}^{\mathcal{I}}_{n'}}(\tilde{x}^{\mathcal{R}},\tilde{x}^{\mathcal{I}}_{n'})\\ =& p_{\tilde{\mathsf{x}}^{\mathcal{R}}_{n'}|\tilde{\mathsf{x}}^{\mathcal{I}}_{n'}}(\tilde{x}^{\mathcal{R}}_{n'}|\tilde{x}^{\mathcal{I}}_{n'})p_{\tilde{\mathsf{x}}^{\mathcal{I}}}(\tilde{x}^{\mathcal{I}})\\ =& p_{\tilde{\mathsf{x}}^{\mathcal{I}}_{n'}|\tilde{\mathsf{x}}^{\mathcal{R}}_{n'}}(\tilde{x}^{\mathcal{I}}_{n'}|\tilde{x}^{\mathcal{R}}_{n'})p_{\tilde{\mathsf{x}}^{\mathcal{R}}}(\tilde{x}^{\mathcal{R}})\,.\end{aligned}$$

It is clear that the real and imaginary parts are not separable and thus are not independently distributed. With equally probable symbols from $\tilde{\mathcal{M}}_M$ the prior pmf can be written as

$$p_{\tilde{\mathsf{x}}_n}(\tilde{x}_n) = \gamma^{(0)}\delta(\tilde{x}_n) + \frac{1-\gamma^{(0)}}{M}\sum_{\tilde{m}_q\in\tilde{\mathcal{M}}_M}\delta(\tilde{x}_n - \tilde{m}_q)\,. \quad (3)$$

Transforming this for the real measurement model (2) yields the prior for both the real and imaginary parts

$$p_{\mathsf{x}_{n'}}(x_{n'}) = \gamma^{(0)}\delta(x_{n'}) + \frac{1-\gamma^{(0)}}{\sqrt{M}}\sum_{m_q\in\mathcal{M}_M}\delta(x_{n'} - m_q)\qquad n'=1,\dots,N'\,,$$

not independently over indices n'.
Based on the measurement model (2), the measured vector y will also be a random variable, transformed linearly by the measurement matrix and disturbed by random additive noise:

$$\mathsf{y} = \mathbf{A}\mathsf{x} + \mathsf{w}\,,$$

with $\mathsf{w} \sim \mathcal{N}(0, \sigma_w^2/2\mathbf{I}_{N'})$.
With the probabilistic measurement model at hand, we can formulate the minimum mean squared error (MMSE) estimator of x given the measurement $\mathbf{y}$:

$$\begin{aligned}\hat{\mathbf{x}} &= \underset{\mathbf{x}^*}{\operatorname{argmin}}\, \mathrm{E}_{\mathsf{x},\mathsf{w}}\left\{\|\mathsf{x}-\mathbf{x}^*\|_2^2 \,\middle|\, \mathsf{y}=\mathbf{y}\right\}\\ &= \mathrm{E}_{\mathsf{x}}\left\{\mathsf{x}\,|\,\mathsf{y}=\mathbf{y}\right\}\,. \end{aligned}\quad (4)$$

The expectation $\mathrm{E}_{\mathsf{x},\mathsf{w}}\{\cdot\}$ is taken over the random variables x and w. In general, the complexity of solving this exactly involves an N'-dimensional integral that is unfeasible in practice.

C. The BAMP Algorithm and Extensions

The BAMP algorithm, stated in [11] and shown in Algorithm 1, solves (4) approximately but efficiently for large N', given $\mathbf{y}$, $\mathbf{A}$ and the prior distribution $p_{\mathsf{x}}(\mathbf{x})$. One assumption on the measurement matrix $\mathbf{A}$ is that it satisfies the restricted isometry property (RIP) [7], which, for i.i.d. channel tap realizations and pseudo-randomly chosen transmission sequences is satisfied with high probability [1]. The key ingredients of Algorithm 1 are the following functions involving the signal prior:

$$\begin{aligned}F(u_{n'};\beta,\gamma(0)) &= \mathrm{E}_{\mathsf{x}_{n'}}\left\{\mathsf{x}_{n'}\,|\,\mathsf{u}_{n'}=u_{n'}\right\}\,,\\ F'(u_{n'};\beta,\gamma(0)) &= \frac{\mathrm{d}}{\mathrm{d}u_{n'}}F(u_{n'};\beta,\gamma(0))\,.\end{aligned}\quad (5)$$

The conditional pmf leading to (5) is

$$\begin{aligned}&p_{\mathsf{x}_{n'}|\mathsf{u}_{n'}}(x_{n'}|u_{n'};\beta,\gamma(0))\\ &\quad= \frac{1}{\sqrt{2\pi\beta}}\exp\left(-\frac{1}{2\beta}(x_{n'}-u_{n'})^2\right)\frac{p_{\mathsf{x}_{n'}}(x_{n'})}{f_{\mathsf{u}_{n'}}(u_{n'})}\,,\\ &\qquad n'=1,\dots,N'\,.\end{aligned}\quad (6)$$

The variance β of this distribution is computed in every iteration of BAMP (in the t-th iteration $\beta^{(t)}$) and is strictly larger than the variance $\sigma_w^2/2$ of the K' noise components in the measurement model. Moreover, this pdf applies for $n'=1,\dots,N'$, whereas $\mathbf{y}$ has only dimension K'. It has been proven (see e.g. [3], [10]) that asymptotically (as $N'\to\infty$ and $K'/N' = \text{const.}$) the pmf (6) is associated with a new *decoupled* measurement model

$$u_{n'} = x_{n'} + v_{n'}\,,\quad n'=1,\dots,N'\,,\quad \text{with } v_{n'}\sim\mathcal{N}(0,\beta)\,. \quad (7)$$

Here, the *effective noise* $v_{n'}$ combines the measurement noise $w_{n'}$ and the *undersampling noise*, the latter resulting from the fact that $K' < N'$. Its Gaussian distribution results from $N\to\infty$. In practice, due to the central limit theorem, $N'\gg 1$ suffices for adequate estimation. The algorithm runs until the maximum number of iterations $t_{\max}$ is reached or the residual $\mathbf{z}^{(t)}$ does not change significantly (as defined by ϵ_{TOL}).
The BAMP algorithm delivers an MMSE estimate $\hat{\mathbf{x}}_{\text{BAMP}}$ (respectively $\hat{\tilde{\mathbf{x}}}_{\text{BAMP}}$) of $\mathbf{x}$, thus, in general its entries will not be equal to values in the augmented alphabet $\mathcal{M}_M\cup\{0\}$. Moreover and most importantly, it might happen that the real part estimate is far from zero whereas the imaginary part estimate is close to zero, i.e.,

$$\hat{\tilde{\mathbf{x}}}^{\mathcal{R}}_{\text{BAMP}} \gg 0\,, \hat{\tilde{\mathbf{x}}}^{\mathcal{I}}_{\text{BAMP}} = 0\,,$$

which is a significant disagreement that the later stages of detection have to cope with. Thus, we wish to utilize an algorithm for solving (2) that respects the group structure of the real-valued vector $\mathbf{x}$. In order to formally treat this dependency, we define a latent random variable, the *activity variable* a_n, $n=1,\dots,N$, with the Bernoulli distribution

$$p_{\mathsf{a}_n}(a_n) = \gamma^{(0)}\delta(a_n) + (1-\gamma^{(0)})\delta(1-a_n)\,, \quad (8)$$

Algorithm 1 BAMP

Input: $t=0\,,\hat{\mathbf{x}}^{(t)} = \mathbf{0}_{N'\times 1}\,,\mathbf{z}^{(t)}=\mathbf{y}\,,\beta^{(t)} = \frac{1}{K}\|\mathbf{z}^{(t)}\|_2^2$
do:
1: $t \leftarrow t+1$
2: $\mathbf{u}^{(t-1)} = \hat{\mathbf{x}}^{(t-1)} + \mathbf{A}^T\mathbf{z}^{(t-1)}$
3: $\beta^{(t-1)} = \frac{1}{K'}\|\mathbf{z}^{(t-1)}\|_2^2$
4: $\hat{\mathbf{x}}^{(t)} = F\left(\mathbf{u}^{(t-1)};\beta^{(t-1)},\gamma^{(0)}\right)$,
5: $\mathbf{z}^{(t)} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{(t)} + \frac{1}{K'}\mathbf{z}^{(t-1)}\sum_{n'=1}^{N'}F'\left(u_{n'}^{(t-1)};\beta^{(t-1)},\gamma^{(0)}\right)$
while $\|\mathbf{z}^{(t)}-\mathbf{z}^{(t-1)}\|_2^2/\|\mathbf{z}^{(t-1)}\|_2^2 > \epsilon_{TOL}$ **and** $t < t_{\max}$
Output: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(t)}$, $\hat{\mathbf{u}} = \hat{\mathbf{u}}^{(t-1)}$, $\beta = \beta^{(t-1)}$

Algorithm 2 BOSSAMP for group sparse QAM signals

Input: $t = 0$, $\hat{\mathbf{x}}^{(t)} = \mathbf{0}_{N'\times 1}$, $\mathbf{z}^{(t)} = \mathbf{y}$,
$\boldsymbol{\gamma}^{(t)} = \gamma^{(0)}\mathbf{1}_{N'\times 1}$,

do:

$t \leftarrow t+1$

$\mathbf{u}^{(t-1)} = \hat{\mathbf{x}}^{(t-1)} + \mathbf{A}^T\mathbf{z}^{(t-1)}$

$\beta^{(t-1)} = \frac{1}{K'}\|\mathbf{z}^{(t-1)}\|_2^2$

$\hat{\mathbf{x}}^{(t)} = F\left(\mathbf{u}^{(t-1)}; \beta^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}\right)$

$\mathbf{z}^{(t)} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{(t)}$
$+\frac{1}{K'}\mathbf{z}^{(t-1)}\sum_{n'=1}^{N'} F'\left(u_{n'}^{(t-1)}; \beta^{(t-1)}, \gamma_{n'}^{(t-1)}\right)$

$\mathbf{l}^{(t)} = U(\mathbf{u}^{(t-1)}, \beta^{(t-1)}, \boldsymbol{\gamma}^{(0)})$

$\boldsymbol{\gamma}^{(t)} = V(\mathbf{l}^{(t)})$

while $\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\|_2^2/\|\mathbf{z}^{(t-1)}\|_2^2 > \epsilon_{TOL}$ **and** $t < t_{\max}$

Output: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(t)}$, $\mathbf{u} = \mathbf{u}^{(t-1)}$, $\beta = \beta^{(t-1)}$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(t)}$

independently across the indices n. The realization of the activity variable, a_n, indicates whether $\tilde{x}_n$ is zero or a (nonzero) QAM symbol. Then, we can formulate the prior pmf (3) conditioned on the activity variable as

$$p_{\tilde{\mathsf{x}}_n|\mathsf{a}_n}(\tilde{x}_n|\mathsf{a}_n = 0) = \delta(\tilde{x}_n) \tag{9}$$

$$p_{\tilde{\mathsf{x}}_n|\mathsf{a}_n}(\tilde{x}_n|\mathsf{a}_n = 1) = \frac{1-\gamma^{(0)}}{M}\sum_{m_q\in\tilde{\mathcal{M}}_M}\delta(\tilde{x}_n - m_q)\,. \tag{10}$$

This way, the two parts are connected via a mutual underlying vector $\mathbf{a} = (a_1, \ldots, a_N)^T$. Further, if the estimation of $\mathbf{a}$ and $\mathbf{x}$ are treated separately, the potential problem of acquiring nonidentical supports is resolved.

III. Proposed Method

A. The Group-sparse Extension

In [14], the Bayesian optimal structured signal approximate message passing (BOSSAMP) algorithm, a simple yet effective extension of BAMP was introduced, which deals with group sparsity, i.e., if the unknown vector is known to have index groups whose entries are known to be all zero or all nonzero. Our version of BOSSAMP is shown in Algorithm 2. Let us use the following notation:

$$n'' = \begin{cases} n' + N \text{ if } 1 \le n' \le N \\ n' - N \text{ if } N+1 \le n' \le 2N\,. \end{cases} \tag{11}$$

That is, indices n' and n'' always represent the real and imaginary parts (irrespective of their order) of the same component of the complex-valued vector $\tilde{\mathbf{x}}$.
The key feature of this algorithm is to update the *zero probability* $\gamma_{n'}$ of the real (imaginary) parts $\hat{x}_{n'}$ individually during iterations based on the estimates of the imaginary (real) parts $\hat{x}_{n''}$, and to use it in the next iteration. Let $u_{n'}^{(t-1)}$ be the auxiliary variable in BOSSAMP, with effective noise variance $\beta^{(t-1)}$. Based on the decoupled measurement model (7),

$$u_{n'}^{(t-1)} = x_{n'} + v_{n'}^{(t-1)}$$

with $v_{n'}^{(t-1)}$ being a realization of $\mathcal{N}(0, \beta^{(t-1)})$. Calculating the conditional likelihoods for $x_{n'}$ being zero and for $x_{n'}$ being each value from $\mathcal{M}_M$ given $u_{n'}^{(t-1)}$ results in:

$$\begin{aligned}
\mathrm{P}(\mathsf{x}_{n'} = 0|\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)}) &= \\
&= \frac{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)}|\mathsf{x}_{n'} = 0)}{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)})}\mathrm{P}(\mathsf{x}_{n'} = 0) \\
&= \gamma^{(0)}\frac{\mathcal{N}(u_{n'}^{(t-1)}; 0, \beta^{(t-1)})}{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)})}\,,
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(\mathsf{x}_{n'} = m_q|\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)}) &= \\
&= \frac{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)}|\mathsf{x}_{n'} = m_q)}{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)})}\mathrm{P}(\mathsf{x}_{n'} = m_q) \\
&= \frac{1-\gamma^{(0)}}{\sqrt{M}}\frac{\mathcal{N}(u_{n'}^{(t-1)}; m_q, \beta^{(t-1)})}{\mathrm{P}(\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)})}\,.
\end{aligned}$$

Observe the common denominator. To obtain the true conditional log-likelihood ratio (LLR), we can write:

$$\begin{aligned}
l_{n'}^{(t)} &= \log\frac{\gamma^{(0)}\mathcal{N}(u_{n'}^{(t-1)}; 0, \beta^{(t-1)})}{\sum_{m_q\in\mathcal{M}_M}\frac{1-\gamma^{(0)}}{\sqrt{M}}\mathcal{N}(u_{n'}^{(t-1)}; m_q, \beta^{(t-1)})} \\
&= \log\frac{\gamma^{(0)}}{1-\gamma^{(0)}} + \frac{1}{2}\log M + \\
&\quad \log\frac{\mathcal{N}(u_{n'}^{(t-1)}; 0, \beta^{(t-1)})}{\sum_{m_q\in\mathcal{M}_M}\mathcal{N}(u_{n'}^{(t-1)}; m_q, \beta^{(t-1)})}\,.
\end{aligned} \tag{12}$$

Note that in order to calculate (12), in every iteration of BOSSAMP a normal distribution has to be evaluated at $2N'(\sqrt{M}+1)$ values. A meaningful simplification is obtained if only the most probable candidate QAM symbol is considered and compared to the zero pseudo-symbol. The same simplification is obtained by using the well-known log-sum-exp identity:

$$\log\sum_i \exp a_i \approx \max a_i\,. \tag{13}$$

Let us denote the most probable (nonzero) candidate symbol (for component n') by $\hat{m}_{q,n'}$, i.e.,

$$\hat{m}_{q,n} = \underset{m_q\in\mathcal{M}_M}{\operatorname{argmax}}\,\mathrm{P}(\mathsf{x}_{n'} = m_q|\mathsf{u}_{n'}^{(t-1)} = u_{n'}^{(t-1)})\,. \tag{14}$$

Since the normal distributions with means $m_q \in \mathcal{M}_M$ have identical variance, this is equivalent to the symbol closest to $u_{n'}^{(t-1)}$, i.e.,

$$\hat{m}_{q,n} = \underset{m_q\in\mathcal{M}_M}{\operatorname{argmin}}\,|u_{n'}^{(t-1)} - m_q|\,,$$

whose evaluation is computationally significantly less complex than that of (14). Using the identity (13), (12) simplifies to

$$\begin{aligned}
l_{n'}^{(t)} &= U(u_{n'}^{(t-1)}, \beta^{(t-1)}, \gamma^{(0)}) \\
&\approx \log\frac{\gamma^{(0)}}{1-\gamma^{(0)}} + \frac{1}{2}\log M + \log\frac{\mathcal{N}(u_{n'}^{(t-1)}; 0, \beta^{(t-1)})}{\mathcal{N}(u_{n'}^{(t-1)}; \hat{m}_q, \beta^{(t-1)})}\,.
\end{aligned}$$

The third summand can be interpreted as the extrinsic information that the real (imaginary) part estimate provides on top of the general prior knowledge for the imaginary (real) part estimation. The interpretation of this LLR is as follows: if $\tilde{l}_{n'}^{(t)}$ is positive and far from zero, the estimate $u_{n'}^{(t-1)}$ suggests that $x_{n'}$ is probably zero; whereas if $\tilde{l}_{n'}^{(t)}$ is negative and far from zero, it suggests the opposite, i.e., that $x_{n'}$ and $x_{n''}$ are nonzero and the corresponding node was transmitting a QAM symbol. Note that the LLR does not favor or discriminate any of the QAM symbols, it only helps to decide between the zero and nonzero values. The prior update of the complementary component $x_{n''}$ for the next iteration is as follows:

$$\gamma_{n''}^{(t)} = V(l_{n'}^{(t)}) = \frac{1}{1+\exp(-l_{n'}^{(t)})},$$

which is a well known relationship between the LLR and the probability of a binary random variable.

Our version of the BOSSAMP algorithm performs the conventional BAMP iterations with the modification that after every iteration, based on the relationship between the amplitude of the real/imaginary part estimate, the likelihood of that component being zero is calculated. These likelihoods are then exchanged between the estimators of the real and imaginary parts, and are then transformed into prior probabilities for the next iteration. This way, in case of convergence, the estimates of the real and imaginary parts reach a consensus regarding whether the corresponding node was active or not. Moreover, the prior distribution describing the QAM constellation will favor the nonzero estimates to be closer to true QAM symbols.

B. *Post-processing: Activity Detection and Quantization*

As a final step, one could perform nearest neighbor quantization of the ultimate BOSSAMP estimate $\hat{\mathbf{x}}$ to the augmented real-valued alphabet $\mathcal{M}_M$. With this method, however, it might still happen that real and imaginary estimates of the same node disagree on its activity, i.e., one is nonzero while the other is zero. Another possibility is to first generate the complex estimate by combining the respective real and imaginary estimates as $\hat{\tilde{x}}_n = \hat{x}_{n'} + j\hat{x}_{n''}$; and then to perform nearest-neighbor quantization to the augmented QAM alphabet $\tilde{\mathcal{M}}_M \cup \{0\}$. In the case of the first method, however, it might still happen that real and imaginary estimates of the same node disagree on its activity, i.e., one is nonzero while the other is zero. Hereby we introduce a third method which uses soft information acquired from BOSSAMP, not only amplitude values.

We call the decoupled measurement model (7) once more into action. Let us denote the ultimate BOSSAMP variables $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(t)}$, $\mathbf{u} = \mathbf{u}^{(t-1)}$, $\beta = \beta^{(t-1)}$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(t)}$. Furthermore, consider the estimation of $\tilde{x}_n$, whose corresponding real and imaginary part indices in $\mathbf{x}$ are n' and n'', respectively:

$$\tilde{x}_n = x_{n'} + jx_{n''}\,.$$

Recall that both $u_{n'}$ and $u_{n''}$ are realizations of a normal distribution with variances β and unknown means, which come from the finite set $\mathcal{M}_M \cup \{0\}$. If we treat $(u_{n'}, u_{n''})^T$ as a vector variable, it is the realization of a normal distribution with covariance matrix $\beta\mathbf{I}_2$ and unknown mean, which is one of the values in the set $\mathcal{M}_M \times \mathcal{M}_M \cup \{\mathbf{0}_{2\times1}\}$. A theoretically sound way of clustering values that come from different (and possibly unknown) distributions is the expectation-maximization (EM) algorithm [5]. The final decision boils down to comparing the *responsibilities* $\rho(\cdot)$ for $(u_{n'}, u_{n''})^T$ coming from the zero-mean and the non-zero-mean distributions:

$$\rho(a_n = 0) \propto \gamma_{n'}\gamma_{n''}\mathcal{N}(u_{n'};0,\beta)\mathcal{N}(u_{n''};0,\beta)\,,$$
$$\rho(a_n = 1) \propto \sum_{m'_q\in\mathcal{M}_M}\sum_{m''_q\in\mathcal{M}_M}\frac{(1-\gamma_{n'})(1-\gamma_{n''})}{M^2}\mathcal{N}(u_{n'};m'_q,\beta)\mathcal{N}(u_{n''};m''_q,\beta)\,,$$

In practice, however, for $\rho(a_n = 1)$ it suffices to consider the summands corresponding to the most probable symbol candidate(s) $\hat{m}_{q,n'}, \hat{m}_{q,n''}$:

$$\rho(a_n = 1) \propto (1-\gamma_{n'})(1-\gamma_{n''})\mathcal{N}(u_{n'};\hat{m}_{q,n'},\beta)\mathcal{N}(u_{n''};\hat{m}_{q,n''},\beta)\,.$$

The responsibility can be interpreted as the "relative contribution" of a particular distribution to the observation u_n. To arrive at the final unquantized estimate $\tilde{x}_n$, we compare the responsibilities as follows:

$$\tilde{x}_n \leftarrow \begin{cases} 0 & \text{if } \rho(a_n=0) > \rho(a_n=1) \\ x_{n'} + jx_{n''} & \text{if } \rho(a_n=0) \le \rho(a_n=1)\,. \end{cases}$$

Even though the EM-algorithm is in principle optimal [5], numerical results have shown that the advantage of the EM-based method is insignificant for higher modulation orders ($M = 16, 64, 256$), and marginal for $M = 4$. The intuitive reason is that this method only helps the detection in case of the 4 symbols that are closest to 0, because in in case of other symbols further from 0 there is no need for more precise decision between the symbol and 0. For e.g. 64-QAM, this is approximately one-sixteenth of all symbols, so even if EM outperforms the nearest-neighbor quantization, its overall effect is minimal. To sum up, we state that the most efficient method is the nearest-neighbor quantization in the complex symbol domain.

IV. Numerical Results

To show the performance of our method, we simulated an overloaded CDMA system with $N = 1000$ nodes, each with a pseudo-random binary chip sequence of length K with amplitudes $\{-1, +1\}$. For each node, the channel impulse response is a sequence of complex normally distributed variables of

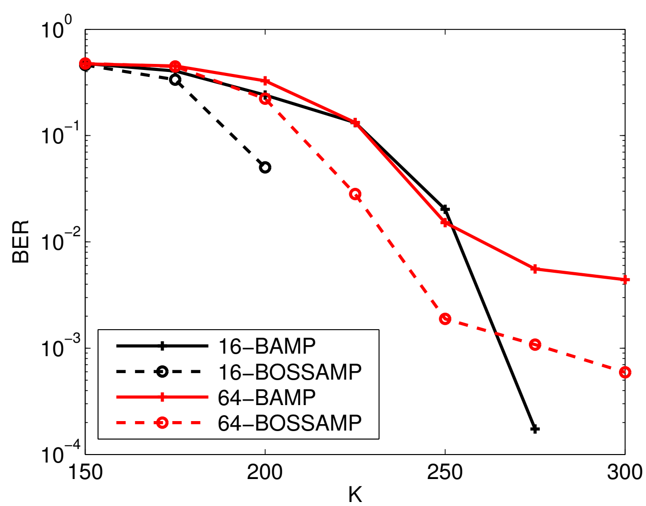

Fig. 1. Average BER of only the active nodes over the chip sequence length K for 16- and 64-QAM when decoded with BAMP and BOSSAMP.

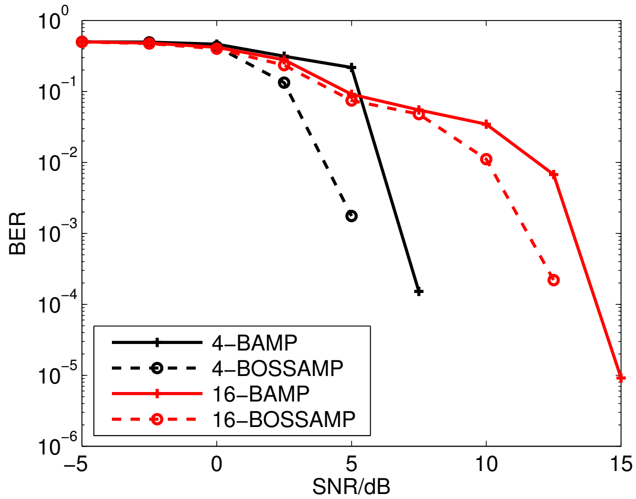

Fig. 2. Average BER of only the active nodes over SNR for 16- and 64-QAM when decoded with BAMP and BOSSAMP.

length $K'/8$. The measurement matrix $\mathbf{A}$, defined by the user sequences and the channel impulse responses, is constructed to have normalized columns, which is in practice feasible. The activity probability is $\gamma^{(0)} = 0.05$. The SNR is defined as

$$\mathrm{SNR} = \mathrm{E}\left\{ \frac{\|\tilde{\mathbf{A}}\tilde{\mathbf{x}}\|_2^2}{\|\tilde{\mathbf{w}}\|_2^2} \right\}. \tag{15}$$

In each simulation, 1000 independent realizations of the set of active nodes, their symbols, chip sequences, channels and the AWGN were simulated and the obtained values averaged.

In the first simulation, we compare BAMP with BOSSAMP, both followed by nearest-neighbor decision in a 16-QAM and a 64-QAM system. The SNR is fixed at 20dB and we varied the chip sequence length K between 150 and 300. In Figure 1 we plotted the averaged bit error ratio (BER) of only the active nodes over K.

In the second simulation, we compare BAMP with BOSSAMP, both followed by nearest-neighbor decision in a 4-QAM and a 16-QAM system. We fixed the chip sequence length to $K = 300$ and varied the SNR between -5dB and 15dB. The averaged BER of the active nodes is shown in Figure 2.

Both simulations show that our proposed method outperforms the existing one in terms of average BER. We note that due to the simplicity and efficacy of both methods they are well suited for time-critical applications. For the above parameters they tend to converge within at most $10 - 20$ iterations.

We refer the interested user to a previous work [13], where the *correlation receiver* [19] was compared to the non-Bayesian AMP for real-valued channels and symbol sets.

V. Conclusion and Outlook

We presented the complex-valued communication model for multiuser systems and the uplink scenario. We have shown that (opposed to the classic CS methods) the probabilistic formulation is advantageous when the symbol alphabet in the digital communication is known in advance. We proposed an efficient algorithm that exploits the dependency between the real and imaginary parts of the unknown data vector and that also forgoes complicated post-processing. Numerical results have shown that the real-imaginary joint solvers outperforms the standard BAMP algorithm, the state-of-the-art algorithm for underdetermined real-valued systems of linear equations.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[2] A. Bartoli, M. Dohler, J. Hernandez-Serrano, A. Kountouris, and D. Barthel, "Low-power low-rate goes long-range: The case for secure and cooperative machine-to-machine communications," in *NETWORKING 2011 Workshops*. Springer Berlin Heidelberg, 2011, vol. 6827, pp. 219–230.

[3] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.

[4] E. Biglieri and M. Lops, "Multiuser detection in a dynamic environment; part I: User identification and data detection," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3158–3170, Sept. 2007.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer New York, 2006.

[6] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 389–400, 2013.

[7] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[8] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[9] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[10] D. L. Donoho and A. Maleki, A. Montanari. (2011) How to design message passing algorithms for compressed sensing. [Online]. Available: http://www.ece.rice.edu/~mam15/bpist.pdf

[11] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory*, Jan. 2010, pp. 1–5.

[12] ——, "Message passing algorithms for compressed sensing: II. Analysis and validation," in *2010 IEEE Information Theory Workshop on Information Theory*, Jan. 2010, pp. 1–5.

[13] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015, pp. 2086–2091.

[14] M. Mayer and N. Goertz, "Bayesian Optimal Approximate Message Passing to Recover Structured Sparse Signals," *ArXiv e-prints*, Aug. 2015. [Online]. Available: http://adsabs.harvard.edu/abs/2015arXiv150801104M

[15] H. Schepker, C. Bockelmann, and A. Dekorsy, "Coping with CDMA asynchronicity in compressive sensing multi-user detection," in *Proc. 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[16] H. Schepker and A. Dekorsy, "Sparse multi-user detection for CDMA transmission using greedy algorithms," in *2011 8th International Symposium on Wireless Communication Systems (ISWCS)*, Nov. 2011, pp. 291–295.

[17] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[18] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Communications Letters*, vol. 16, no. 7, pp. 972–974, Jul. 2012.

[19] S. Verdú, *Multiuser Detection*, Cambridge (UK), 1998.

[20] W. Wu and K. Chen, "Identification of active users in synchronous CDMA multiuser detection," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 9, pp. 1723–1735, Dec. 1998.

[21] H. Zhu and G. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454–465, Feb. 2011.