Improving Approximate Message Passing Recovery of Sparse Binary Vectors by Post Processing

Martin Mayer and Norbert Goertz

Institute of Telecommunications, Vienna University of Technology
Gusshausstrasse 25/E389, A-1040 Vienna, Austria
Email: mmayer@nt.tuwien.ac.at and norbert.goertz@nt.tuwien.ac.at.

***Abstract*—Compressed sensing allows to recover a sparse high-dimensional signal vector from fewer measurements than its dimension would suggest. Approximate Message Passing (AMP) was introduced by Donoho et al. to perform iterative recovery with low complexity. We investigate the behavior of AMP for a low number of measurements and low signal-to-noise ratio. In these cases, recovered signals exhibit many false alarms and strong amplitude deviations from the original. We propose post processing schemes that try to retain the correct detections while voiding false alarms. The best performance is obtained by Maximum Likelihood (ML) detection that is performed on a well selected subset of the recovered signal vector's support. In order to maintain feasibility, ML detection is implemented by a *sphere decoder* and the sparse signal vector is assumed to be *binary*, i.e., its few nonzero elements have value one. We show by simulation that post processing enables accurate estimation of the true signal vector and its support, despite strong noise or an insufficient number of measurements.**

I. Introduction

The introduction of compressed sensing [1]–[4] led to a shift of paradigm in sampling theory. It was shown that sparse or compressible signals can be acquired by significantly fewer measurements (samples) than suggested by the long-established Shannon-Nyquist sampling theorem. This is enabled by a proper choice of the sampling basis functions, also called dictionary, and a recovery formulation that accounts for the solution accuracy as well as the sparsity; details follow in Section II.

The goal of the recovery scheme is to reconstruct the signal from a seemingly incomplete set of linear measurements. In this work, we focus on Approximate Message Passing (AMP) introduced in [5]–[7]. It allows for efficient and fast iterative recovery with low computational complexity. An implementation overview is provided in [8].

Loosely speaking, AMP searches for an optimum by iteratively adapting local optima, and the resulting signal recovery is an approximation of the true signal. It exhibits good performance and finds an accurate signal estimate with relatively few iterations, provided that the number of compressed sensing measurements or the Signal-to-Noise Ratio (SNR) is sufficiently large. In [9], we presented simulations that suggest that with increasing noise power, the number of measurements has to increase as well in order to keep the reconstruction error small. If the number of measurements is too small for a given noise power, AMP yields an imperfect signal reconstruction with false alarms and large amplitude deviations.

In this work, we present two methods that improve such imperfect signal reconstructions by applying post processing. A ”slicer” and a more complex Maximum Likelihood (ML) detector are introduced in order to estimate the amplitude information and in particular the true signal support in terms of vector indices. The schemes perform well if the signal vector entries belong to a finite set with very small cardinality, and the number of nonzero entries in the original signal vector is rather small (high sparsity). We therefore stick to signal vectors with many zeros and few ones (sparse binary vectors). For both approaches, we consider the cases of known and unknown sparsity.

The acquisition of sparse binary vectors has been researched in literature such as [10], [11]. Note that these approaches are invasive to the measurement process, while our approach does not change the measurements but the subsequent processing thereof.

To perform ML detection, we utilize a soft output sphere decoder [12]–[15]. While this concept was originally introduced to ease the computational complexity of transmit symbol detection in wireless transmissions over multiple-input multiple-output channels, we show that it can be adopted to our case. Furthermore, the soft output in terms of Log Likelihood Ratios (LLRs) provides valuable information on the solution's reliability.

The remainder of this paper is organized as follows. Section II recapitulates the measurement model and signal recovery with AMP. Section III investigates the support regions of signal vectors recovered by AMP. Section IV discusses different methods to improve the reconstruction of imperfectly recovered signal vectors and finally introduces the ML detector formulation in terms of a soft output sphere decoder. Section V compares the performance of the proposed detection schemes.

Notation: Boldface letters such as $\mathbf{A}$ and $\mathbf{a}$ denote matrices and vectors, respectively. The $N \times N$ identity matrix is denoted $\mathbf{I}_N$. Calligraphic letters $\mathcal{S}$ denote sets, their usage as subscript $\mathbf{A}_\mathcal{S}$ or $\mathbf{a}_\mathcal{S}$ implies that only the matrix columns or vector entries defined by the elements in $\mathcal{S}$ are selected. The cardinality of a set is denoted by $|\mathcal{S}|$. The superscript $^\mathrm{T}$ denotes the transposition of a vector or matrix.

II. Compressed Sensing and AMP

Let us introduce an N-dimensional signal vector $\mathbf{x}$ that is assumed to be K sparse, i.e., only $K \ll N$ entries are nonzero. We intend to recover $\mathbf{x}$ from $M < N$ non-adaptive, noisy linear measurements

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

where $\mathbf{y} \in \mathbb{R}^M$ is the measurement vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the sensing matrix and $\mathbf{w} \in \mathbb{R}^M$ is additive white Gaussian measurement noise with zero mean entries and covariance matrix $\sigma_w^2 \mathbf{I}_M$.

Compressed sensing theory [1]–[4] indicates that $\mathbf{x}$ can be recovered despite (1) being an underdetermined system of linear equations with noise present. To ensure stable recovery from noisy measurements, the sensing matrix $\mathbf{A}$ has to satisfy the Restricted Isometry Property (RIP) [3] $(1-\delta_K)\left\|\mathbf{v}\right\|_2^2 \leq \left\|\mathbf{A}\mathbf{v}\right\|_2^2 \leq (1+\delta_K)\left\|\mathbf{v}\right\|_2^2$ for all K-sparse vectors $\mathbf{v}$, which implies that $\mathbf{A}$ preserves their Euclidean length up to a small constant defined by δ_K. Randomly generated matrices with i.i.d. Gaussian entries were proven in [16] to almost surely satisfy the RIP, while the number of required measurements for successful recovery is lower bounded by $M = cK \log \frac{N}{K}$. The simulations in this work utilize such randomly generated sensing matrices. Note that the measurement multiplier $c \in \mathbb{R}$ has to increase with increasing noise power to maintain a certain recovery performance [9].

The recovery of $\mathbf{x}$ from (1) can be formulated according to the so called Least Absolute Shrinkage and Selection Operator (LASSO) [17], also known as basis pursuit denoising [18], which yields a non-linear convex optimization problem that reads

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2^2 + \lambda \left\|\mathbf{x}\right\|_1 \right\}. \tag{2}$$

The underlying intuition is to find the most accurate solution with the smallest support (motivated by the assumed sparsity of $\mathbf{x}$), where λ allows a trade-off between accuracy with respect to the observation error $\left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2^2$ and the sparsity[1] $\left\|\mathbf{x}\right\|_1$ of the solution.

Many schemes [8] have been proposed to solve the LASSO problem. The Approximate Message Passing (AMP) algorithm introduced in [5]–[7] constitutes a low complexity iterative approach with fast convergence, well suited for practical implementations. Our implementation is provided in Algorithm 1; $\eta_\tau(.)$ depicts the soft thresholding [19] operation with threshold τ and the iterations are stopped if residual $\mathbf{r}^t$ does not change notably anymore (controlled by a small constant ϵ) or $t_{\max}$ is reached which limits the number of iterations.

III. Investigation of Signal Support Sets

In this section, we introduce signal support regions in terms of vector index sets, or briefly, *signal support sets*. Their investigation will motivate the necessity of post processing.

[1]Sparsity is usually expressed by the l_0-"norm" according to $\left\|\mathbf{x}\right\|_0 \leq K$. The l_1-norm relaxation was proven in [1] to often yield the same result in high dimensions while introducing a favorable convex optimization problem.

Algorithm 1 Approximate Message Passing (AMP) [5]

initialize $\mathbf{r}^t = \mathbf{y}$ and $\mathbf{x}^t = \mathbf{0}$ for $t = 0$
do
  $t = t + 1$ ▷ advance iterations
  $\tau = \frac{\lambda}{\sqrt{M}} \left\|\mathbf{r}^{t-1}\right\|_2$ ▷ compute threshold
  $\mathbf{x}^t = \eta_\tau(\mathbf{x}^{t-1} + \mathbf{A}^{\mathrm{T}}\mathbf{r}^{t-1})$ ▷ soft thresholding
  $b = \frac{1}{M}\left\|\mathbf{x}^t\right\|_0$ ▷ compute sparsity
  $\mathbf{r}^t = \mathbf{y} - \mathbf{A}\mathbf{x}^t + b\mathbf{r}^{t-1}$ ▷ compute residual
while $\left\|\mathbf{r}^t - \mathbf{r}^{t-1}\right\|_2 > \epsilon$ and $t < t_{\max}$
return $\widehat{\mathbf{x}} = \mathbf{x}^t$ ▷ recovered sparse vector

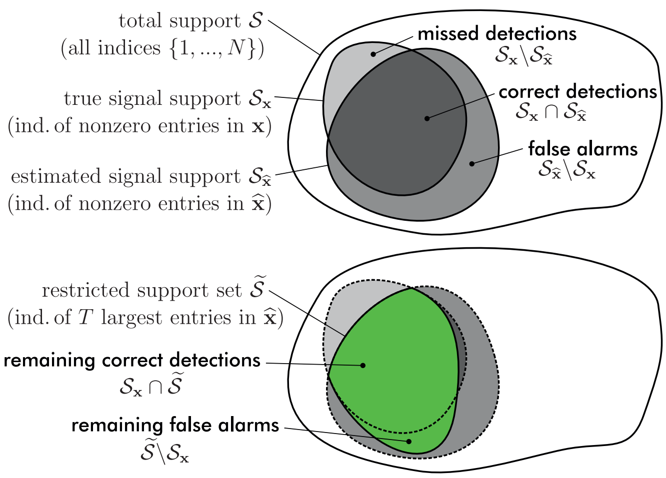

Fig. 1. Illustration of signal support sets. All sets comprise vector indices.

The following signal support sets – illustrated in Figure 1 – are defined:

- Total support $\mathcal{S}$: all indices $i = \{1, ..., N\}$ of a signal vector $\mathbf{x}$. The cardinality is $|\mathcal{S}| = N$.
- True signal support $\mathcal{S}_{\mathbf{x}}$: indices of the nonzero entries in $\mathbf{x}$. The cardinality is $|\mathcal{S}_{\mathbf{x}}| = K$.
- Estimated signal support $\mathcal{S}_{\widehat{\mathbf{x}}}$: indices of the nonzero entries in the AMP recovered signal vector $\widehat{\mathbf{x}}$. The cardinality is typically $|\mathcal{S}_{\widehat{\mathbf{x}}}| \geq K$ and influenced by λ in (2). An increasing λ reduces $|\mathcal{S}_{\widehat{\mathbf{x}}}|$.
- Restricted support set $\widetilde{\mathcal{S}}$: indices of the T largest entries in $\widehat{\mathbf{x}}$, i.e., a subset of $\mathcal{S}_{\widehat{\mathbf{x}}}$. Note that T is a design parameter. In case of $T \geq |\mathcal{S}_{\widehat{\mathbf{x}}}|$, we define $\widetilde{\mathcal{S}} = \mathcal{S}_{\widehat{\mathbf{x}}}$.

Introducing an indicator function

$$\mathbb{I}(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}, \tag{3}$$

we define the number of **correct detections** in $\widehat{\mathbf{x}}$ as set cardinality

$$N_{\mathrm{CD}}(\widehat{\mathbf{x}}) = |\mathcal{S}_{\mathbf{x}} \cap \mathcal{S}_{\widehat{\mathbf{x}}}| = \sum_{i \in \mathcal{S}_{\mathbf{x}}} \mathbb{I}(\widehat{x}_i), \tag{4}$$

the number of **missed detections** as

$$N_{\mathrm{MD}}(\widehat{\mathbf{x}}) = |\mathcal{S}_{\mathbf{x}} \backslash \mathcal{S}_{\widehat{\mathbf{x}}}| = K - N_{\mathrm{CD}}(\widehat{\mathbf{x}}), \tag{5}$$

and the number of **false alarms** as

$$N_{\mathrm{FA}}(\widehat{\mathbf{x}}) = |\mathcal{S}_{\widehat{\mathbf{x}}} \backslash \mathcal{S}_{\mathbf{x}}| = \sum_{i \in \mathcal{S} \backslash \mathcal{S}_{\mathbf{x}}} \mathbb{I}(\widehat{x}_i). \tag{6}$$

Similarly, we define the number of **remaining correct detections** after support restriction, i.e., the number of correct detections in the restricted support set, as

$$N_{\mathrm{CD},\widetilde{\mathcal{S}}}(\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}) = |\mathcal{S}_{\mathbf{x}} \cap \widetilde{\mathcal{S}}| = \sum_{i \in \mathcal{S}_{\mathbf{x}}} \mathbb{I}\left((\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}})_i\right), \tag{7}$$

and the number of **remaining false alarms** after support restriction as

$$N_{\mathrm{FA},\widetilde{\mathcal{S}}}(\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}) = |\widetilde{\mathcal{S}} \backslash \mathcal{S}_{\mathbf{x}}| = \sum_{i \in \mathcal{S} \backslash \mathcal{S}_{\mathbf{x}}} \mathbb{I}\left((\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}})_i\right). \tag{8}$$

Let us investigate the behavior of these numbers w.r.t. the number of measurements M and the Signal-to-Noise Ratio

$$\mathrm{SNR} = \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{M\sigma_w^2} = \frac{P}{\sigma_w^2}, \tag{9}$$

where $P = \|\mathbf{A}\mathbf{x}\|_2^2 / M$ denotes the average signal power per measurement.

An example of a sparse binary signal vector $\mathbf{x} \in \{0,1\}^N$ with $K = 10$ nonzero entries and $N = 1000$ is illustrated in Figure 2. Due to low $\mathrm{SNR} = 3\,\mathrm{dB}$ and only few measurements $M = 140$, its AMP recovery $\widehat{\mathbf{x}}$ is imperfect with strong amplitude deviations in the true signal support region. It has many false alarms because low sparsity was enforced by small $\lambda = 1.6$ in Algorithm 1. The crux in the λ selection lies in the fact that K is generally unknown a priori — if K is small, high sparsity can be enforced by large λ. If one does not know the sparsity in advance, large λ may exclude solutions. The numbers defined above amount to $N_{\mathrm{CD}} = 10$, $N_{\mathrm{MD}} = 0$ and $N_{\mathrm{FA}} = 86$. In this example, all correct detections are included as nonzero entries in $\widehat{\mathbf{x}}$ and thus as indices in $\mathcal{S}_{\widehat{\mathbf{x}}}$, but there are many false alarms and it is not clear which elements actually are correct detections.

In Figure 3, the restricted support set $\widetilde{\mathcal{S}}$ of the signal vector from Figure 2 is plotted for two values of M. For large number of measurements $M = 600$ (see Figure 3 (a)), the K largest entries in $\widehat{\mathbf{x}}$ constitute the correct detections, and there is a visible gap (in terms of amplitude) between correct detections and false alarms. For small number of measurements $M = 140$ (see Figure 3 (b)), correct detections are no longer among the K largest entries in $\widehat{\mathbf{x}}$ but still in the restricted support set of cardinality $T = |\widetilde{\mathcal{S}}| = 3K = 30$. We obtain $N_{\mathrm{CD},\widetilde{\mathcal{S}}} = 10$ and $N_{\mathrm{FA},\widetilde{\mathcal{S}}} = 20$. The number of false alarms was reduced while retaining the correct detections.

We still have to find a way to detect the correct true support, i.e., the K entries that were nonzero in $\mathbf{x}$. We first investigate the general behavior of the sets. All subsequent illustrations are based on simulation results considering a signal vector $\mathbf{x}$ of dimension $N = 1000$ with $K = 10$ nonzero entries. The results are averaged over 10000 independent realizations.

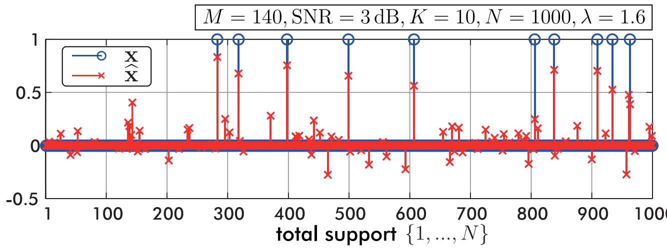

Fig. 2. Example of a sparse binary signal vector $\mathbf{x}$ and AMP recovery $\widehat{\mathbf{x}}$.

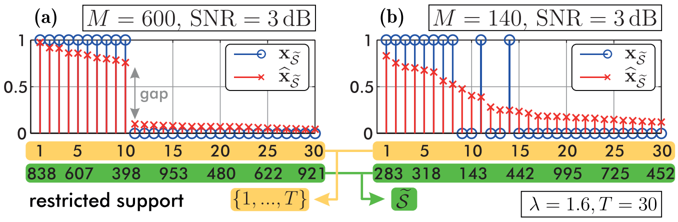

Fig. 3. Depicted are the $T = 30$ largest values of AMP recovery $\widehat{\mathbf{x}}$ arranged in descending order and the corresponding (same indices) original signal vector $\mathbf{x}$. The indices corresponding to the T largest values in $\widehat{\mathbf{x}}$ comprise the restricted support set $\widetilde{\mathcal{S}}$. (a) illustrates the case with a large number of measurements M, (b) with small M.

A. Behavior of Signal Support after AMP Recovery

Let us investigate the percentages of correct detections, missed detections and false alarms after AMP recovery by defining $N_{\mathrm{CD}} + N_{\mathrm{MD}} + N_{\mathrm{FA}} = 100\%$.

Figure 4 (a) illustrates the support set percentages over variable number of measurements M at fixed $\mathrm{SNR} = 3\,\mathrm{dB}$. A low choice of λ generates many false alarms. Increasing λ enforces higher sparsity in the LASSO solution and thereby reduces the number of false alarms. If the number of nonzero entries K is unknown, λ should be chosen carefully (not too large) so that no correct detections are excluded. A low SNR coupled with a low number of measurements M results in missed detections, i.e., true solutions (correct detections) are lost by AMP.

Figure 4 (b) illustrates the support set percentages over variable SNR range at fixed number of measurements $M = 140$. An overall similar behaviour to Figure 4 (a) is observed.

Note that we deal with extreme cases — by increasing the number of measurements and SNR, the false alarms will shrink and are orders of magnitude lower than the correct detections.

B. Behavior of Restricted Support Set

ML detection that will be introduced in Section IV is only feasible on a heavily reduced index set, namely the restricted support set $\widetilde{\mathcal{S}}$. We therefore investigate how many correct detections remain after we restrict the support to the T largest values in $\widehat{\mathbf{x}}$. Note that detections that are missed by AMP will be missed by ML detection as well as the ML search range was limited by AMP.

Figure 5 (a) depicts the number of correct detections and false alarms over variable number of measurements M at fixed $\mathrm{SNR} = 3\,\mathrm{dB}$. The restricted support set was chosen to have cardinality $T = |\widetilde{\mathcal{S}}| = 2K = 20$. We observe that most of the

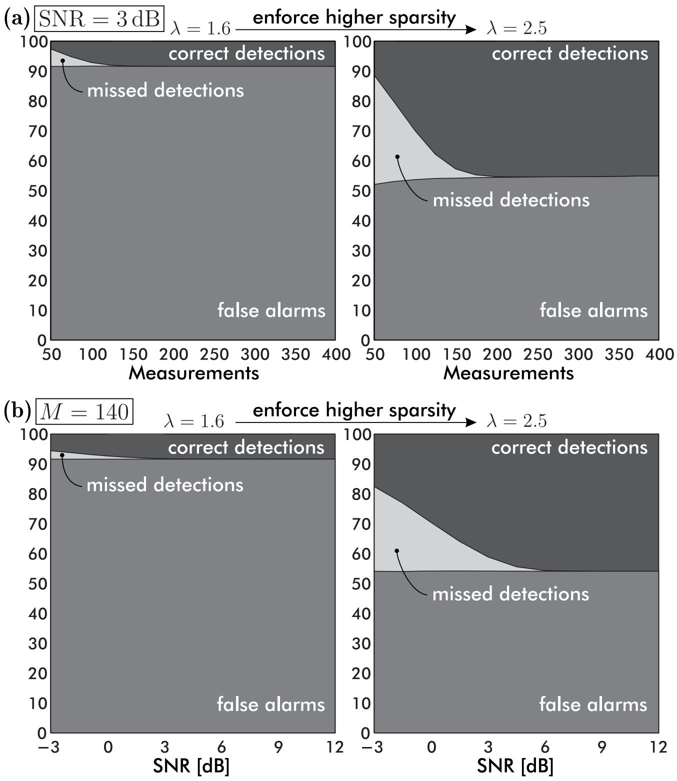

Fig. 4. Support set percentages so that $N_{\mathrm{CD}} + N_{\mathrm{MD}} + N_{\mathrm{FA}} = 100\%$.

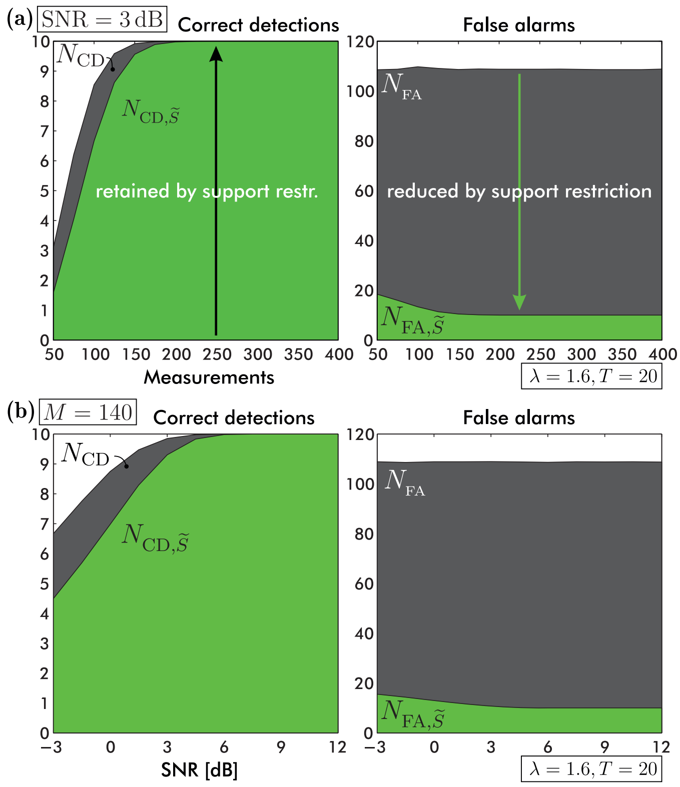

Fig. 5. Number of correct detections and false alarms before and after support restriction.

correct detections are retained by support restriction, i.e., by considering only the T largest values of AMP recovery $\widehat{\mathbf{x}}$. The number of false alarms is strongly reduced.

Figure 5 (b) shows the number of correct detections and false alarms over variable SNR at fixed $M = 140$. A similar behavior to Figure 5 (a) is observed.

We hence conclude that the restricted support set $\widetilde{\mathcal{S}}$ with $T = |\widetilde{\mathcal{S}}| = 2K$ is very likely to retain the correct detections from AMP while many false alarms are excluded if M or SNR are sufficiently large.

IV. Post Processing by Detectors

Let us now discuss post processing by introducing detectors that try to find the true signal support, i.e., the correct detections, while voiding the false alarms. Remember that $\widehat{\mathbf{x}} = [\widehat{x}_1, ..., \widehat{x}_N]^{\mathrm{T}}$ is the AMP recovery without post processing. The following detectors know that the true signal vector is binary and can only take values zero and one.

A. Simple Approach: Slicer

Assuming that the true support is likely to be contained in the set of largest values in $\widehat{\mathbf{x}}$ (see Figure 3), the standard slicer performs the following mapping:

$$\widehat{x}_i^{\mathrm{SLICER}} = \begin{cases} 1 & \widehat{x}_i > 0.5 \\ 0 & \widehat{x}_i \leq 0.5 \end{cases}, \forall i \in \mathcal{S} = \{1, ..., N\}. \quad (10)$$

If the number of nonzero elements K is known, the slicer maps the K largest values in $\widehat{\mathbf{x}}$ to one:

$$\widehat{x}_i^{\mathrm{SLICER}|K} = \begin{cases} 1 & i \in \widetilde{\mathcal{S}} \\ 0 & \text{else} \end{cases}, \forall i \in \mathcal{S}, |\widetilde{\mathcal{S}}| = T = K. \quad (11)$$

B. Advanced Approach: Maximum Likelihood Detector

We now formulate an ML detector scheme that is applied on the restricted support set $\widetilde{\mathcal{S}}$ with $|\widetilde{\mathcal{S}}| = T$. The standard ML detector formulation reads

$$\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}(\mathbf{y}) = \arg \min_{\mathbf{x}_{\widetilde{\mathcal{S}}} \in \{0,1\}^T} \|\mathbf{y} - \mathbf{A}_{\widetilde{\mathcal{S}}} \mathbf{x}_{\widetilde{\mathcal{S}}}\|_2^2, \quad (12)$$

it assumes white Gaussian noise $\mathbf{w}$ in (1) and a uniform prior distribution of $\mathbf{x}$. It implies that we have to compute 2^T Euclidean distances in order to find the "most likely" signal $\mathbf{x}$ to produce measurement $\mathbf{y}$. In Section III, we concluded that the correct detections are very likely to be contained in the restricted support set if its cardinality is $T = 2K$.

To ease the computational burden, (12) can be reformulated in order to be solvable by a Schnorr-Euchner sphere decoder [12]–[15]. First, the restricted support sensing matrix is decomposed by a QR-decomposition $\mathbf{A}_{\widetilde{\mathcal{S}}} = \mathbf{QR}$ with a unitary matrix $\mathbf{Q} \in \mathbb{R}^{M \times T}$ and an upper triangular matrix $\mathbf{R} \in \mathbb{R}^{T \times T}$. By left-multiplying (1) with $\mathbf{Q}^{\mathrm{T}}$ after applying the support restriction, we obtain the alternative measurement formulation

$$\begin{aligned} \mathbf{Q}^{\mathrm{T}}\mathbf{y} = \mathbf{Q}^{\mathrm{T}}\mathbf{A}_{\widetilde{\mathcal{S}}}\mathbf{x}_{\widetilde{\mathcal{S}}} + \mathbf{Q}^{\mathrm{T}}\mathbf{w} = \\ \widetilde{\mathbf{y}} = \mathbf{Rz} + \widetilde{\mathbf{w}}, \end{aligned} \quad (13)$$

where we used $\mathbf{z} = \mathbf{x}_{\widetilde{\mathcal{S}}}$ for convenience of notation in the following. The refined ML detector reads

$$\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}(\widetilde{\mathbf{y}}) = \arg \min_{\mathbf{z} \in \{0,1\}^T} \|\widetilde{\mathbf{y}} - \mathbf{Rz}\|_2^2. \quad (14)$$

This transforms the Euclidean distance computation into a tree search with T layers consisting of the rows in $\mathbf{R}$; each tree node entails two branches due to the binary entries in $\mathbf{z}$. The bottom row $i = T$ of matrix $\mathbf{R}$ constitutes the first branching below the root node. Traversing the tree from root to leaf, each branch adds a distance increment $|e_i|^2 = \left|\widetilde{y}_i - \sum_{j=1}^{T} R_{ij} z_j\right|^2$, and a node at layer t is associated with a partial Euclidean distance $d_t = \sum_{i=t}^{T} |e_i|^2$. At leaf node layer $t = 1$, the partial Euclidean distance d_1 is equivalent to the Euclidean distance $\|\widetilde{\mathbf{y}} - \mathbf{R}\mathbf{z}\|_2^2$.

Each leaf node corresponds to one of the 2^T possible vectors $\mathbf{z}$, the goal is to find the leaf node with the smallest Euclidean distance. The search is restricted by a search radius that prunes subtrees if the partial Euclidean distance d_t exceeds the radius. This significantly reduces the number of necessary computations. A good initial radius could be provided by using the slicer solution in (10) or (11) — the closer the initial "guess" is to the correct solution, the smaller becomes the search radius and the amount of remaining nodes to compute. If fixed detection throughput is aspired, ML detection can be approximated utilizing a K-best sphere decoder [14], [15].

After detection was performed on the restricted support set, the nonzero entries have to be mapped back to a vector of dimension N (thereby lifting the support restriction):

$$\widehat{x}_i^{\mathrm{ML}} = \begin{cases} 1 & i \in \widetilde{\mathcal{S}} \wedge \left(\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}\right)_i = 1 \\ 0 & \text{else} \end{cases}, \forall i \in \mathcal{S}. \tag{15}$$

The usage of a *soft output* sphere decoder allows to compute Log Likelihood Ratios (LLRs) [12], [15] for all T elements in the restricted support set $\widetilde{\mathcal{S}}$. Let us first define the maximum likelihood distance

$$\lambda^{\mathrm{ML}} = \min_{\mathbf{z} \in \{0,1\}^T} \|\widetilde{\mathbf{y}} - \mathbf{R}\mathbf{z}\|_2^2 = \left\|\widetilde{\mathbf{y}} - \mathbf{R}\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}(\widetilde{\mathbf{y}})\right\|_2^2, \tag{16}$$

which can be computed by a sphere decoder. For each entry $(\mathbf{x}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}})_i$ in $\mathbf{x}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}$, the counter hypothesis distance is defined as

$$\lambda_i^{\mathrm{CH}} = \min_{\mathbf{z} \in \{0,1\}^T | z_i = \overline{(\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}})_i}} \|\widetilde{\mathbf{y}} - \mathbf{R}\mathbf{z}\|_2^2, \tag{17}$$

i.e., the minimum distance given that the i^{th} binary entry in $\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}}$ is flipped. Again, these distances can be computed by a sphere decoder. The LLRs with max-log approximation [20] are now defined as

$$\mathrm{LLR}\left((\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}})_i\right) = \begin{cases} \lambda^{\mathrm{ML}} - \lambda_i^{\mathrm{CH}} & (\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}})_i = 0 \\ \lambda_i^{\mathrm{CH}} - \lambda^{\mathrm{ML}} & (\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}^{\mathrm{ML}})_i = 1 \end{cases}. \tag{18}$$

The ML detector with knowledge of K exploits this by setting only the entries corresponding to the K largest LLRs to 1, the rest is set to 0. Let $\widetilde{\mathcal{S}}_L$ denote the support set of the K largest LLRs, the detector is then defined as

$$\widehat{x}_i^{\mathrm{ML}|K} = \begin{cases} 1 & i \in \widetilde{\mathcal{S}}_L \\ 0 & \text{else} \end{cases}, \forall i \in \mathcal{S} = \{1, ..., N\}. \tag{19}$$

Figure 6 illustrates the performance of ML detection on the example from Figure 3. Considering a large number of measurements $M = 600$ (see Figure 6 (a)), ML detection works flawlessly. Note the salient "jump" in the LLRs. In case of low number of measurements $M = 140$ (see Figure 6 (b)), $\widehat{\mathbf{x}}^{\mathrm{ML}}$ features all correct detection and one false alarm. $\widehat{\mathbf{x}}^{\mathrm{ML}|K}$ that considers only the K largest LLRs would find the correct solution without false alarms.

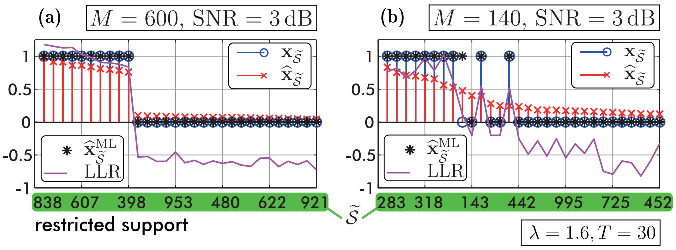

Fig. 6. Signal vector $\mathbf{x}_{\widetilde{\mathcal{S}}}$ and its recovery $\widehat{\mathbf{x}}_{\widetilde{\mathcal{S}}}$ with applied ML detection. (a) illustrates the case with a large number of measurements M, (b) with small M.

V. Detector Comparison and Discussion

Figure 7 shows a detector performance comparison in case of variable number of measurements M at fixed SNR $= 3\,\mathrm{dB}$. The shaded areas indicate the case of unknown sparsity K, the (blue) solid lines indicate the shifted borders in case of known K. Similarly, Figure 8 shows a detector performance comparison in case of variable SNR at fixed number of measurements $M = 140$. We conclude:

- A large λ in LASSO floods the standard slicer (unknown K) with missed detections. The slicer misses detections because, despite a reduced number of false alarms, the entries of the true signal support are estimated poorly and their amplitude is often lower than 0.5, see (10).
- The knowledge of K strongly improves the slicer, particularly in case of large λ.
- The ML detector works superior even without knowledge of K. Knowing K marginally improves the results.

Overall, post processing strongly improves imperfectly recovered signal vectors from AMP. For ML detection, sparsity K does not need to be known for excellent performance.

VI. Conclusion

We introduced and investigated detection schemes to improve imperfect AMP recoveries. Proposed detectors offer a trade-off between the number of measurements and the subsequent detection effort — reducing the number of measurements generates worse AMP recoveries which in turn requires more sophisticated detector schemes. While the simple slicer performs well if sparsity K is known, an ML detector is recommended in case of unknown K or simply to obtain the best performance.

Acknowledgment

This work has been funded by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, and its industrial partner Infineon Technologies. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged.

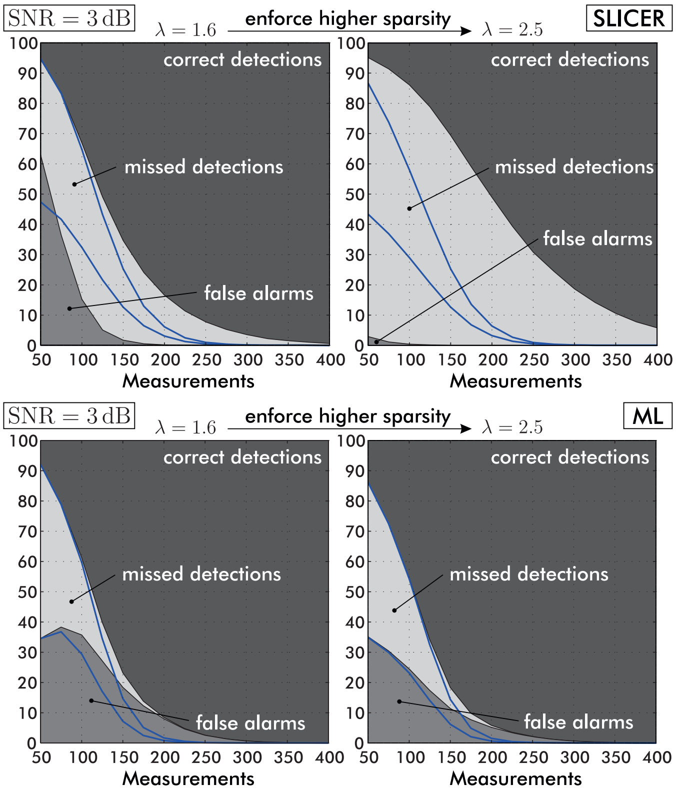

Fig. 7. Support set percentages vs. number of measurements M. ML detection is applied on a restricted support set with $T = |\widetilde{\mathcal{S}}| = 20$. The shaded areas illustrate the percentages in case of unknown K, the (blue) solid lines indicate the shifted borders due to known K.

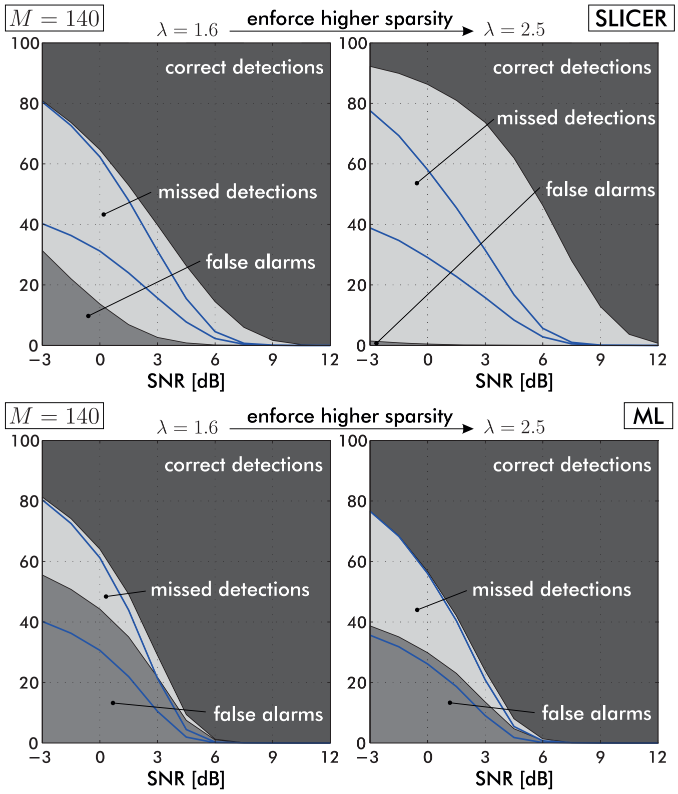

Fig. 8. Support set percentages vs. SNR. ML detection is applied on a restricted support set with $T = |\widetilde{\mathcal{S}}| = 20$. The shaded areas illustrate the percentages in case of unknown K, the (blue) solid lines indicate the shifted borders due to known K.

REFERENCES

[1] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3, pp. 1433–1452, Madrid, Spain, 2006.

[3] E. J. Candès, 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.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] 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. 18914–18919, 2009.

[6] M. A. Maleki, "Approximate message passing algorithms for compressed sensing," *PhD thesis at Stanford University*, 2010.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message Passing Algorithms for Compressed Sensing: I. Motivation and Construction," in *2010 IEEE Information Theory Workshop (ITW)*, pp. 1–5.

[8] P. Maechler, C. Studer, D. E. Bellasi, A. Maleki, A. Burg, N. Felber, H. Kaeslin, and R. G. Baraniuk, "VLSI Design of Approximate Message Passing for Signal Restoration and Compressive Sensing," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 579–590, 2012.

[9] M. Mayer, N. Goertz, and J. Kaitovic, "RFID Tag Acquisition via Compressed Sensing," in *Proceedings of IEEE RFID Technology and Applications (RFID-TA)*, 2014.

[10] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *online available at http://arxiv.org/abs/1104.3160*, 2011.

[11] M. Yan, Y. Yang, and S. Osher, "Robust 1-bit compressive sensing using adaptive outlier pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3868–3875, 2012.

[12] C. Studer, M. Wenk, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Performance and implementation aspects," in *40th Asilomar Conference on Signals, Systems and Computers (ACSSC'06)*, pp. 2071–2076, IEEE, 2006.

[13] C. Studer, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Algorithms and VLSI implementation," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 2, pp. 290–300, 2008.

[14] Z. Guo and P. Nilsson, "Algorithm and Implementation of the K-Best Sphere Decoding for MIMO Detection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 491–503, 2006.

[15] M. Mayer, M. Simko, and M. Rupp, "Soft-output sphere decoding: Single tree search vs. improved k-best," in *Proceedings of 18th International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, 2011.

[16] 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.

[17] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[18] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[19] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[20] B. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.