Exploiting Sparse User Activity in Multiuser Detection

Hao Zhu, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—The number of *active* users in code-division multiple access (CDMA) systems is often much lower than the spreading gain. The present paper exploits fruitfully this *a priori* information to improve performance of multiuser detectors. A low-activity factor manifests itself in a sparse symbol vector with entries drawn from a finite alphabet that is augmented by the zero symbol to capture user inactivity. The non-equiprobable symbols of the augmented alphabet motivate a sparsity-exploiting maximum *a posteriori* probability (S-MAP) criterion, which is shown to yield a cost comprising the ℓ_2 least-squares error penalized by the p-th norm of the wanted symbol vector ($p = 0, 1, 2$). Related optimization problems appear in variable selection (shrinkage) schemes developed for linear regression, as well as in the emerging field of compressive sampling (CS). The contribution of this work to such *sparse* CDMA systems is a gamut of sparsity-exploiting multiuser detectors trading off performance for complexity requirements. From the vantage point of CS and the least-absolute shrinkage selection operator (Lasso) spectrum of applications, the contribution amounts to sparsity-exploiting algorithms when the entries of the wanted signal vector adhere to finite-alphabet constraints.**

***Index Terms*—Sparsity, multiuser detection, compressive sampling, Lasso, sphere decoding.**

I. Introduction

MULTIUSER detection (MUD) algorithms play a major role for mitigating multi-access interference present in code-division multiple access (CDMA) systems; see e.g., [17] and references therein. These well-appreciated MUD algorithms simultaneously detect the transmitted symbols of all active user terminals. However, they require knowledge of which terminals are active, and exploit no possible user (in)activity.

In this paper, MUD algorithms are developed when the active terminals are unknown and the activity factor (probability of each user being active) is low – a typical scenario in tactical or commercial CDMA systems deployed. The inactivity per user can be naturally incorporated by augmenting the underlying alphabet with an extra zero constellation point. Low activity thus implies a sparse symbol vector to be detected. With non-equiprobable symbols in the augmented alphabet, the optimal sparsity-embracing MUD naturally suggests a maximum *a posteriori* (MAP) criterion for detection. Sparsity has been used for estimating parameters of communication systems, see e.g., [1], [3], [6], [10], but not for multiuser detection[1].

Reconstruction of sparse signals has become very popular recently, especially after the emergence of the compressive sampling (CS) theory; see e.g., [2], [4], [7], [16] and references therein. Exploiting sparsity has also received growing interest in regression problems because it offers parsimonious statistical models that prevent data overfitting. In particular, the least-absolute shrinkage and selection operator (Lasso) has been applied for variable selection (VS) to diverse applications involving sparse signals [15]. Lasso regression amounts to regularizing the ordinary least-squares (LS) with the ℓ_1 norm of the wanted vector, thus retaining the attractive features of both VS and LS regression. However, neither CS nor Lasso have dealt with sparse signals under finite-alphabet constraints; and this is the subject of the present work.

Different from sparse signal reconstruction or regression, the multiuser detectors developed in this paper must account for sparsity but also for the finite alphabet of the desired symbol vector. Taking into consideration the sparsity in MUD under integer constraints however, incurs complexity that is exponential in the vector length. To cope with this challenge, we will exploit sparsity to either relax or judiciously search over subsets of the alphabet. The resultant MUD algorithms trade off optimality in detection error performance with computational complexity. In a nutshell, this paper's contribution is the development of efficient algorithms for MUD under sparsity and finite-alphabet constraints.

Also in the CS literature a number of works is available to deal with the performance of recovering sparse sequences of size exceeding the number of observations; see e.g., [4], [5] and references therein. These works establish the minimum number of observations needed to estimate sparse vectors with overwhelming probability. In the design of practical CDMA systems however, one is also interested in saving bandwidth and power resources. This becomes possible by reducing the size of the required spreading gain, which in turn reduces latency and energy consumption. Taking advantage of the low activity factor, enables designs with spreading gains

Paper approved by X. Wang, the Editor for Multiuser Detection and Equalization of the IEEE Communications Society. Manuscript received September 21, 2009; revised April 2, 2010, July 3, 2010, and August 30, 2010.

This work was supported by the NSF grants CCF-0830480, 1016605, and ECCS-0824007, 1002180; and also through the collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. Part of this work was presented at the IEEE Intl. Symp. on Info. Theory, Seoul, South Korea, June 2009.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA (e-mail: {zhuh, georgios}@umn.edu).

Digital Object Identifier 10.1109/TCOMM.2011.121410.090570

[1]Recently, along with the conference pre-cursor of this paper in [20], work related to sparse sphere decoding was reported independently in [14].

smaller than the number of candidate users. Efficient multiuser detectors are also developed here for this kind of under-determined systems encountered in CDMA communications.

The rest of the paper is organized as follows. Section II presents the model and formulates the problem; while Section III introduces the optimal sparsity-exploiting MAP detector. (Sub-) optimal algorithms are developed in Sections IV and V. Section VI lists several extensions. All the detection algorithms are tested and compared numerically in Section VII, before concluding in Section VIII.

II. Modeling and Problem Statement

Consider the uplink of a CDMA system with K user terminals and spreading gain N. Assume first that $N \geq K$. The under-determined case ($N < K$) will also be addressed later on in Section VI-B. Suppose the system has a relatively low activity factor, which analytically means that each terminal is active with probability (w.p.) $p_a < 1/2$ per symbol, and the events "*active*" are independent across symbols and across users. The case of correlated (in)activity of users across symbols will be dealt with in Section VI-A. Let $b_k \in \mathcal{A}$ denote the symbol, drawn from a finite alphabet by the k-th user, when active; otherwise, $b_k = 0$. Incorporating possible (in)activity per user is equivalent to having b_k take values from an *augmented* alphabet $\mathcal{A}_a := \mathcal{A} \bigcup \{0\}$.

The access point (AP) receives the superimposed modulated (quasi-) synchronous signature waveforms through (possibly frequency-selective) fading channels in the presence of additive white Gaussian noise (AWGN); and projects on the orthonormal space spanned by the aggregate waveforms to obtain the received chip samples collected in the $N \times 1$ vector $\mathbf{y}$. With the $K \times 1$ vector $\mathbf{b}$ containing the symbols of all (active and inactive) users, the canonical input-output relationship is, see e.g., [17, Sec. 2.9]

$$\mathbf{y} = \mathbf{H}\mathbf{b} + \mathbf{w} \tag{1}$$

where $\mathbf{H}$ is an $N \times K$ matrix capturing transmit-receive filters, spreading sequences, channel impulse responses, and timing offsets; and the $N \times 1$ vector $\mathbf{w}$ is the AWGN. Without loss of generality (w.l.o.g.), $\mathbf{w}$ can be scaled to have unit variance. Note that (1) holds for quasi-synchronous systems too, where relative user asynchronism is bounded to a few chips per symbol, provided that: either i) user transmissions include guard bands to eliminate inter-symbol interference (ISI); or ii) the received vector $\mathbf{y}$ collects only the chips of each user that belong to the part of the common symbol interval under consideration.

The low activity factor implies that $\mathbf{b}$ is a sparse vector. However, the AP is neither aware of the positions nor the number of zero entries in $\mathbf{b}$. In order to perform multiuser detection (MUD) needed to determine the optimal $\hat{\mathbf{b}}$, the AP must account for the augmented alphabet $\mathcal{A}_a$, i.e., consider all the possible candidate vectors $\mathbf{b} \in \mathcal{A}_a^K$. This way, the MUD also determines the k-th user's activity captured by the extra constellation point $b_k = 0$. Supposing that the AP has the channel matrix $\mathbf{H}$ available (e.g., via training), the goal of this paper is to detect the optimal $\hat{\mathbf{b}}$ given the received vector $\mathbf{y}$ by exploiting the sparsity of active users.

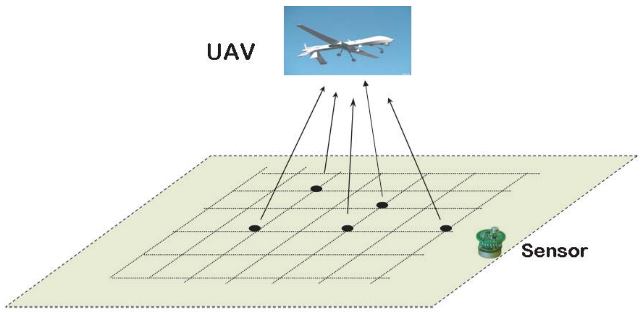

Fig. 1. Wireless sensors access a UAV.

To motivate this sparsity-exploiting MUD setup in the CDMA context, consider a set of terminals wishing to link with a common AP. Suppose that the AP acquires the full matrix $\mathbf{H}$ (with all terminals active) during a training phase. Those channels may include either non-dispersive or multipath fading, and are assumed invariant during the coherence time, which is typically larger compared to the symbol period. Each terminal accesses the channel randomly, and the AP receives the superposition of signals from the active terminals only. The AP is interested in determining both the active terminals and the symbols transmitted.

Another scenario where $\mathbf{H}$ is known and sparsity-exploiting MUD is well motivated, entails an unmanned aerial vehicle (UAV) collecting information from a ground wireless sensor network (WSN) placed over a grid, as depicted in Fig. 1. As the UAV flies over the grid of sensors, it collects the signals from a random subset of them. If the channel fading is predominantly affected by path loss, the UAV can acquire $\mathbf{H}$ based on the relative AP-to-sensor distances. Again, the UAV faces the problem of determining both identities of active sensors, and the symbols each active sensor transmits.

With this problem setup in mind, we will develop different MUD strategies, which account for the low activity factor. First, we will look at the maximum *a posteriori* probability (MAP) optimum MUD that exploits the sparsity present.

III. Sparsity-Exploiting MAP Detector

The goal is to detect $\mathbf{b}$ in (1), given a prescribed activity factor, the received vector $\mathbf{y}$, and the matrix $\mathbf{H}$. Recall though that the low activity factor leads to a sparse $\mathbf{b}$, i.e., each entry b_k is more likely to take the value 0 from the alphabet. Because entries $\{b_k\}_{k=1}^K$ are non-equiprobable, the optimal detector in the sense of minimizing the detection error rate is the MAP one.

Aiming at a sparsity-aware MAP criterion, consider first the prior probability for $\mathbf{b}$. For simplicity in exposition, suppose for now that each terminal transmits binary symbols when active, i.e., $\mathcal{A} = \{\pm 1\}$. (It will become clear later on that all sparsity-cognizant MUD schemes are applicable to finite alphabets of general constellations, not necessarily binary.) If b_k takes values from $\{-1, 0, 1\}$, with corresponding probabilities $\{p_a/2, 1 - p_a, p_a/2\}$, and since each entry b_k is independent from $b_{k'}$ for $k \neq k'$, the prior probability for $\mathbf{b}$

can be expressed as

$$\Pr(\mathbf{b}) = \prod_{k=1}^{K} \Pr(b_k) = (1-p_a)^{K-\|\mathbf{b}\|_0}(p_a/2)^{\|\mathbf{b}\|_0} \quad (2)$$

where $\|\mathbf{b}\|_0$ denotes the ℓ_0 (pseudo) norm that is by definition equal to the number of non-zero entries in the vector $\mathbf{b}$. Upon taking logarithms, (2) yields

$$\ln \Pr(\mathbf{b}) = -\lambda\|\mathbf{b}\|_0 + K\ln(1-p_a) \quad (3)$$

where

$$\lambda := \ln \frac{1-p_a}{p_a/2}. \quad (4)$$

Since low activity factor means $p_a < 1/2$, it follows readily from (4) that $\lambda > 0$. With the prior distribution of $\mathbf{b}$ in (3), the sparsity-aware MAP (S-MAP) detector is

$$\begin{aligned}\hat{\mathbf{b}}^{\text{MAP}} &= \arg\max_{\mathbf{b}\in\mathcal{A}_a^K} \Pr(\mathbf{b}|\mathbf{y}) = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} -\ln p(\mathbf{y}|\mathbf{b}) - \ln\Pr(\mathbf{b}) \\ &= \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_0 \quad (5)\end{aligned}$$

where the last equality follows from (3) and the Gaussianity of $\mathbf{w}$. Hence, the S-MAP detection task amounts to finding the vector in the constraint set $\mathcal{A}_a^K$, which minimizes the cost in (5).

Two remarks are now in order.

Remark 1: *(General constellations).* Beyond binary alphabets, it is easy to see why the S-MAP detector in (5) applies to more general constellations, including pulse amplitude modulation (PAM), phase-shift keying (PSK), and quadrature amplitude modulation (QAM). Specifically, for general M-ary constellations with $M \geq 2$ it suffices to adjust accordingly the prior probability as a function of $\|\mathbf{b}\|_0$ in (2). This will render the S-MAP MUD in (5) applicable to general M-ary constellations, provided that λ in (4) is replaced by $\lambda := \ln\frac{1-p_a}{p_a/M}$.

Remark 2: *(Scale λ as a function of p_a).* The definition in (4) reveals the explicit relationship of λ with the activity factor p_a. Different from CS and VS approaches, where λ is a tuning parameter often chosen with cross-validation techniques as a function of the data size N and K, here it is directly coupled with the user activity factor p_a. Such a coupling carries over even when users have distinct activity factors. Specifically, if the user k is active w.p. $p_{a,k}$, then the term $\lambda\|\mathbf{b}\|_0 := \lambda\sum_{k=1}^{K}|b_k|$ in (5) should change to $\sum_{k=1}^{K}\lambda_k|b_k|$, where λ_k is defined as in (4) with $p_{a,k}$ substituting p_a. This user-specific regularization will be used in Section VI-A to adaptively estimate user activity factors on-the-fly, and thus enable sparsity-aware MUD which accounts for correlated user (in)activity across the time slots.

With the variable b_k only taking values from $\{\pm1, 0\}$, it holds for $p \geq 1$ that

$$\|\mathbf{b}\|_0 = \sum_{k=1}^{K}|b_k|^p = \|\mathbf{b}\|_p^p, \quad \forall\, \mathbf{b}\in\mathcal{A}_a^K. \quad (6)$$

Hence, the S-MAP detector (5) for binary transmissions is equivalent to

$$\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_p^p\ , \ \ \forall\, p \geq 1. \quad (7)$$

Notice that the equivalence between (5) and (7) is based on the norm equivalence in (6), which holds only for constant modulus constellations. Although the cost in (7) will turn out to be of interest on its own, it is not an S-MAP detector for non-constant modulus constellations.

Interestingly, since low-activity factor implies a positive λ, the problem (7) entails a convex cost function, compared to the non-convex one in (5). In lieu of the finite-alphabet constraint, the criterion of (7) consists of the least-squares (LS) cost regularized by the ℓ_p norm, which in the context of linear regression has been adopted to mitigate data overfitting. In contrast, the LS estimator only considers the goodness-of-fit, and thus tends to overfit the data. Shrinking the LS estimator, by penalizing its size through the ℓ_p norm, typically outperforms LS in practice. For example, the Lasso adopts the ℓ_1 norm through which it effects sparsity [15]. In the MUD context for CDMA systems with low-activity factor, the vector $\mathbf{b}$ has a sparse structure, which motivates well this regularizing strategy. What is distinct and interesting here is that this penalty-augmented LS approach under finite-alphabet constraints emerges naturally as the logarithm of the prior in the S-MAP detector.

However, the finite-alphabet constraint renders the solution of (7) combinatorially complex. For general $\mathbf{H}$ and $\mathbf{y}$, the solution of (7) requires exhaustive search over all the 3^K feasible points, with the complexity growing exponentially in the problem dimension K. Likewise, for general M-ary alphabets the complexity incurred by (5) is $\mathcal{O}((M+1)^K)$. On the other hand, many (sub-) optimal alternatives are available in the MUD literature; see e.g., [17, Ch. 5-7]. Similarly here, we will develop different (sub-) optimal algorithms to trade off complexity for probability of error performance in sparsity-exploiting MUD alternatives.

Since the exponential complexity of MUD stems from the finite-alphabet constraint, one reduced-complexity approach is to solve the unconstrained convex problem, and then quantize the resultant soft decision to the nearest point in the alphabet. This approach includes the sub-optimal linear MUD algorithms (decorrelating and minimum mean-square error (MMSE) detectors). Another approach is to search over (possibly a subset of) the alphabet lattice directly as in the decision-directed detectors or the sphere decoders [18]. Likewise, it is possible to devise (sub-) optimal algorithms for solving the S-MAP MUD problem (7) along these two categories. First, we will present the sparsity-exploiting MUD algorithms by relaxing the finite-alphabet constraint.

IV. Relaxed S-MAP Detectors

In addition to offering an S-MAP detector for constant modulus constellations, the cost in (7) is convex. Thus, by relaxing the combinatorial constraint, the optimization problem (7) can be solved efficiently by capitalizing on convexity. As mentioned earlier, this problem is similar to the penalty-augmented LS criterion that is used for VS in linear regression, where the choice of p is important for controlling the shrinking effect, that is the degree of sparsity in the solution. Next, we will develop detectors for two choices of p, and compare them in terms of complexity and performance.

A. Linear Ridge MUD

The choice $p = 2$ is a popular one in statistics, well-known as *Ridge regression*. Its popularity is mainly due to the fact that it can regularize the LS solution while retaining its closed-form expression as a linear function of the data $\mathbf{y}$. A relaxed detection algorithm for S-MAP MUD can be developed accordingly with $p = 2$, what we term Ridge detector (RD). Ignoring the finite-alphabet constraint, the optimal solution of (7) for $p = 2$ takes a linear form

$$\begin{aligned}\mathbf{b}^{\mathrm{RD}} &= \arg\min_{\mathbf{b}} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_2^2 \\ &= (\mathbf{H}^T\mathbf{H}+2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}. \end{aligned} \tag{8}$$

In addition to its simplicity, and different from LS, the existence of the inverse in (8) is ensured even for ill-posed or under-determined problems; i.e., when $\mathbf{H}$ is rank deficient or fat ($N < K$). Notice that $\mathbf{b}^{\mathrm{RD}}$ takes a form similar to the linear MMSE multiuser detector, with the parameter λ replacing the noise variance, and connecting the activity factor with the degree of regularization applied to the matrix $\mathbf{H}^T\mathbf{H}$.

Upon quantizing each entry of the soft decision $\mathbf{b}^{\mathrm{RD}}$ with the operator

$$\mathcal{Q}_\theta(x) := \mathrm{sign}(x)\mathbb{1}(|x| \geq \theta) \tag{9}$$

where $\theta > 0$, $\mathrm{sign}(x) = 1(-1)$ with $x > (<)0$, and $\mathbb{1}$ denoting the indicator function, the hard RD is

$$\hat{\mathbf{b}}^{\mathrm{RD}} = \mathcal{Q}_\theta(\mathbf{b}^{\mathrm{RD}}) = \mathcal{Q}_\theta\left((\mathbf{H}^T\mathbf{H}+2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}\right). \tag{10}$$

Because the detector in (8) is linear, it is possible to express linearly its soft output $\mathbf{b}^{\mathrm{RD}}$ with respect to (w.r.t.) the input symbol vector $\mathbf{b}$. Based on this relationship, one can subsequently derive the symbol error rate (SER) of the hard detected symbols in $\hat{\mathbf{b}}^{\mathrm{RD}}$ as a function of the quantization threshold θ. These steps will be followed next to obtain the performance of the RD.

1) Performance Analysis: Letting $\check{\mathbf{b}}$ denote the vector transmitted, and substituting $\mathbf{y} = \mathbf{H}\check{\mathbf{b}} + \mathbf{w}$ into (8) yields

$$\mathbf{b}^{\mathrm{RD}} = (\mathbf{H}^T\mathbf{H}+2\lambda\mathbf{I})^{-1}\mathbf{H}^T(\mathbf{H}\check{\mathbf{b}}+\mathbf{w}) = \mathbf{G}\check{\mathbf{b}} + \mathbf{w}' \tag{11}$$

where $\mathbf{G} := \mathbf{I} - 2\lambda(\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-1}$, and the colored noise $\mathbf{w}' := (\mathbf{H}^T\mathbf{H}+2\lambda\mathbf{I})^{-1}\mathbf{H}^T\mathbf{w}$ is zero-mean Gaussian with covariance matrix $\mathbf{\Sigma}_{w'} := E\{\mathbf{w}'(\mathbf{w}')^T\} = (\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})^{-2}\mathbf{H}^T\mathbf{H}$.

It follows readily from (11) that the k-th entry of $\mathbf{b}^{\mathrm{RD}}$ satisfies

$$b_k^{\mathrm{RD}} = G_{kk}\check{b}_k + \sum_{\ell\neq k} G_{k\ell}\check{b}_\ell + w'_k \tag{12}$$

where $G_{k\ell}$ and w'_k are the (k,ℓ)-th and k-th entries of $\mathbf{G}$ and $\mathbf{w}'$, respectively. The last two terms in the right-hand side of (12) capture the multiuser interference-plus-noise effect, which has variance

$$\sigma_k^2 = \mathrm{var}\left\{\sum_{\ell\neq k} G_{k\ell}\check{b}_\ell + w'_k\right\} = \sum_{\ell\neq k} G_{k\ell}^2 p_a + \Sigma_{w',kk} \tag{13}$$

where $\Sigma_{w',kk}$ denotes the (k,k)-th entry of $\mathbf{\Sigma}_{w'}$.

With the interference-plus-noise term being approximately Gaussian distributed, deciphering $\check{b}_k$ from (12) amounts to detecting a ternary deterministic signal in the presence of zero-mean, Gaussian noise of known variance. Hence, the symbol error rate (SER) for the k-th entry using the quantization rule in (10) entry-wise can be analytically obtained as

$$P_{e,k}^{\mathrm{RD}} = 2(1-p_a)Q\left(\frac{\theta}{\sigma_k}\right) + p_a Q\left(\frac{G_{kk}-\theta}{\sigma_k}\right) \tag{14}$$

where $Q(\mu) := (1/\sqrt{2\pi})\int_\mu^\infty \exp(-\nu^2/2)d\nu$ denotes the Gaussian tail function.

The SER in (14) is a convex function of the threshold θ. Thus, taking the first-order derivative w.r.t. θ and setting it equal to zero yields the optimal threshold for the k-th entry as

$$\hat{\theta}_k = \frac{G_{kk}}{2} + \frac{\sigma_k^2}{G_{kk}}\lambda\,. \tag{15}$$

The corresponding minimum SER becomes [cf. (14) and (15)]

$$\hat{P}_{e,k}^{\mathrm{RD}} = 2(1-p_a)Q\left(\frac{G_{kk}}{2\sigma_k} + \frac{\lambda\sigma_k}{G_{kk}}\right) + p_a Q\left(\frac{G_{kk}}{2\sigma_k} - \frac{\lambda\sigma_k}{G_{kk}}\right). \tag{16}$$

As the CDMA system signal-to-noise ratio (SNR) goes to infinity, asymptotically we have $\mathbf{G} \to \mathbf{I}$ and $\mathbf{\Sigma}_{w'} \to \mathbf{0}$, so the optimal threshold in (15) approaches 0.5. The numerical tests in Section VII will also confirm that selecting $\theta_k = 0.5$ approaches the minimum SER $\hat{P}_{e,k}^{\mathrm{RD}}$ over the range of SNR values encountered in most practical settings.

The clear advantage of RD-MUD is its simplicity as a linear detector. However, using the ℓ_2 norm for regularization, the RD-MUD inherently introduces a Gaussian prior for the unconstrained symbol vector and is thus not effecting sparsity in $\mathbf{b}^{\mathrm{RD}}$; see also [15]. This renders the performance of RD dependent on the quantization threshold θ – a fact also corroborated by the simulations in Section VII. These considerations motivate the ensuing development of an alternative relaxed S-MAP algorithm, which accounts for the sparsity present in $\mathbf{b}$.

B. Lasso-based MUD

Another popular regression method is the Lasso one, which regularizes the LS cost with the ℓ_1 norm. In the Bayesian formulation, regularization with the ℓ_1 norm corresponds to adopting a Laplacian prior for $\mathbf{b}$ [15]. The nice feature of Lasso-based regression is that it ensures sparsity in the resultant estimates. The degree of sparsity depends on the value of λ, which is selected here using the *a priori* information available on the activity factor [cf. (4)]. The optimal solution of (7) for $p = 1$ without the finite-alphabet constraint yields the Lasso detector (LD) as

$$\mathbf{b}^{\mathrm{LD}} = \arg\min_{\mathbf{b}} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1. \tag{17}$$

While a closed-form solution is impossible for general $\mathbf{H}$, the minimization in (17) is a quadratic programming (QP) problem that can be readily accomplished using available QP solvers, such as SeDuMi [13]. Upon slicing the solution in (17), we obtain the detection result as

$$\hat{\mathbf{b}}^{\mathrm{LD}} = \mathcal{Q}_\theta(\mathbf{b}^{\mathrm{LD}}). \tag{18}$$

The larger λ is, the more sparse $\mathbf{b}^{\text{LD}}$ becomes [cf. (4)]. This is intuitively reasonable, because λ is inversely proportional to the activity factor p_a. Since the Lasso approach (17) yields sparse estimates systematically, and can be obtained via QP solvers in polynomial time, LD is a competitive MUD alternative. Lack of a closed-form solution prevents analytical evaluation of the SER, which will be tested using simulations in Section IV.

Remark 3: So far, we assumed $\mathcal{A} = \{\pm 1\}$ to ensure equivalence of the ℓ_p-norm regularized S-MAP detector (7) with the more general one in (5). However, the sub-optimal algorithms of this section ignore the finite-alphabet constraint, and just rely on the convexity of the cost function in (7) to offer MUD schemes that can be implemented efficiently, either in linear closed-form or through quadratic programming. In fact, starting from (7) and forgoing its equivalence with (5), the RD and LD relaxations of (7) apply also for general M-ary alphabets for any $M > 2$. Of course, the quantization thresholds required for slicing the soft symbol estimates in order to obtain hard symbol estimates must be modified in accordance with the corresponding constellation. For non-constant modulus transmissions, the cost in (7) favors low-energy (close to the origin) constellation points, but this effect is mitigated by the judicious selection of the quantization thresholds.

Note that forgoing the equivalence of (7) with (5) is less of an issue for RD and LD because the major limitation of these simple relaxation-based algorithms is their sub-optimality, which emerges because they do not account for the finite-alphabet symbol constraints. Next, MUD algorithms are developed to minimize the S-MAP cost while adhering to the constraint in (5) explicitly.

V. S-MAP Detectors with Lattice Search

User symbols in this section are drawn from an M-ary PAM alphabet $\mathcal{A} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, with M even. Consider also reformulating the S-MAP problem in (5) using the QR decomposition of the matrix $\mathbf{H}$ (assumed here to be square or tall with full column rank) as $\mathbf{H} = \mathbf{QR}$, where $\mathbf{R}$ is a $K \times K$ upper triangular matrix, and $\mathbf{Q}$ is an $N \times K$ unitary matrix. Substituting this QR decomposition into (5), and left multiplying with the unitary $\mathbf{Q}$ inside the LS cost, the S-MAP detector becomes: $\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^k} \frac{1}{2}\left\|\mathbf{Q}^T\mathbf{y} - \mathbf{Q}^T(\mathbf{QR})\mathbf{b}\right\|_2^2 + \lambda\|\mathbf{b}\|_0 = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\mathbf{y}' - \mathbf{Rb}\|_2^2 + \lambda\|\mathbf{b}\|_0$, where $\mathbf{y}' := \mathbf{Q}^T\mathbf{y}$; or, after using the definitions of the norms,

$$\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \sum_{k=1}^{K} \left\{ \frac{1}{2}\left(y'_k - \sum_{\ell=k}^{K} R_{k\ell} b_\ell \right)^2 + \lambda |b_k|_0 \right\} . \tag{19}$$

Although the optimal solution of (19) still incurs exponential complexity, the upper triangular form of $\mathbf{R}$ enables decomposition of (19) into sub-problems involving only scalar variables. As it will be seen later in Section V-A, the S-MAP problem accepts a neat closed-form solution in the scalar case ($K = 1$). This is instrumental for the development of efficient (near-) optimal algorithms searching over the finite-alphabet induced lattice. One such sub-optimal MUD algorithm is described next.

A. Sparsity-Exploiting Decision-Directed MUD

Close look at (19) reveals that once the estimates $\{\hat{b}_\ell\}_{\ell=k+1}^K$ are available, the optimal $\hat{b}_k$ can be obtained by minimizing the cost corresponding to the k-th summand of (19). This leads to the per-symbol optimal decision-directed detector (DDD), which following related schemes in different contexts, could be also called *successive interference cancellation* or *decision-feedback decoding*, see e.g., [11, Sec. 9.4] and [17, Ch. 7]. The main difference here is that $\mathbf{b}$ is sparse.

The DDD algorithm relies on back substitution to decompose the overall S-MAP cost into K sub-costs each dependent on a single scalar variable and accepting a closed-form solution. Specifically, supposing the symbols $\{\hat{b}_\ell^{\text{DDD}}\}_{\ell=k+1}^K$ have been already detected, the DDD algorithm detects the k-th symbol as

$$\hat{b}_k^{\text{DDD}} = \arg\min_{b_k\in\mathcal{A}_a} \frac{1}{2}\left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\text{DDD}} - R_{kk}b_k \right)^2 + \lambda|b_k|_0 . \tag{20}$$

This minimization problem entails only one scalar variable taking one of $(M+1)$ possible points in $\mathcal{A}_a$. Thus, the minimum is found after comparing the costs corresponding to these $(M+1)$ values. Appendix A proves that this optimal solution can be found in closed form as

$$\hat{b}_k^{\text{DDD}} = \lfloor b_k^{\text{LS}} \rceil \mathbb{1}\left(2b_k^{\text{LS}}\lfloor b_k^{\text{LS}}\rceil - \lfloor b_k^{\text{LS}}\rceil^2 - 2\lambda/R_{kk}^2 > 0\right) \tag{21}$$

where $b_k^{\text{LS}} := (y'_k - \sum_{\ell=k+1}^K R_{k\ell}\hat{b}_\ell^{\text{DDD}})/R_{kk}$, and $\lfloor\cdot\rceil$ quantizes to the nearest point in $\mathcal{A}$. The simple implementation steps are tabulated as Algorithm 1.

Algorithm 1 (DDD): Input $\mathbf{y}'$, $\mathbf{R}$, λ, and M even. Output $\hat{\mathbf{b}}^{\text{DDD}}$.

```
1: for k = K, K−1, ..., 1 do
2:    (Back substitution) Compute the unconstrained LS solution
      b_k^LS := (y'_k − Σ_{ℓ=k+1}^K R_kℓ b̂_ℓ^DDD)/R_kk.
3:    (Quantize to A) Set b̂_k^DDD := ⌊b_k^LS⌉.
4:    (Compare with 0) If 2 b_k^LS ⌊b_k^LS⌉ − ⌊b_k^LS⌉^2 − 2λ/R_kk^2 ≤ 0, then
      set b̂_k^DDD := 0.
5: end for
```

When $\mathbf{R}$ is diagonal, Algorithm 1 yields the optimal S-MAP detection result; i.e., $\hat{\mathbf{b}}^{\text{MAP}} = \hat{\mathbf{b}}^{\text{DDD}}$ for this case. However, since the DDD detects symbols sequentially, it is prone to error propagation, especially at low SNR values. The error propagation can be mitigated by preprocessing and ordering methods [17, Ch. 7]. Also similar to all related detectors that rely on back substitution, error performance of the sparse DDD will be analyzed assuming there is no error propagation. For the special case of $M = 2$, Appendix B shows that under this assumption, the SER for (21) becomes

$$P_{e,k}^{\text{DDD}} = 2(1-p_a)Q\left(\frac{|R_{kk}|}{2} + \frac{\lambda}{|R_{kk}|}\right) + p_a Q\left(\frac{|R_{kk}|}{2} - \frac{\lambda}{|R_{kk}|}\right) . \tag{22}$$

For a general M-ary constellation, it is also possible to approximate the SER using the union bound.

Because it accounts for the finite-alphabet constraint, sparse DDD outperforms the relaxed detectors of the previous section – a fact that will be confirmed also by simulations. However, the error propagation emerging at medium-low SNR degrades the sparse DDD performance when compared to the optimal but computationally complex S-MAP detector. As a compromise, a branch-and-bound type of MUD algorithm is developed next, to attain (near-) optimal performance by exploiting the finite-alphabet and sparsity constraints, at the price of increased complexity compared to DDD.

B. Sparsity-Exploiting Sphere Decoding-based MUD

Sphere decoding (SD) algorithms have been widely used for maximum-likelihood (ML) demodulation of multiuser and/or multiple-input multiple-output (MIMO) transmissions. Given a PAM or QAM alphabet, SD yields (near-) ML performance at polynomial average complexity; see e.g., [11, Sec. 5.2]. However, different from the ML-optimal SD that minimizes an LS cost, the S-MAP problem (19) entails also a regularization term to account for sparsity in $\mathbf{b}$. Although the resultant algorithm will be termed sparse sphere decoder (SSD), it searches in fact within an "ℓ_0-*norm regularized* sphere," which is not a sphere but a hyper-solid.

The goal is to find the unknown $K \times 1$ vector $\mathbf{b} \in \mathcal{A}_a^K$, which minimizes the distance metric [cf. (19)]

$$D_1^K(\mathbf{b}) := \sum_{k=1}^{K} \left\{ \frac{1}{2} \left(y'_k - \sum_{\ell=k}^{K} R_{k\ell} b_\ell \right)^2 + \lambda |b_k|_0 \right\}. \quad (23)$$

For a large enough threshold τ, candidate vectors (and thus the minimizer of D_1^K too) satisfy[2]

$$D_1^K(\mathbf{b}) < \tau \quad (24)$$

that specifies a hyper-solid inside which the wanted minimizer must lie. Define now [cf. (21)]

$$\rho_k := \left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell} b_\ell \right) \Big/ R_{kk}, \quad k = K, K-1 \cdots 1 \quad (25)$$

where $\rho_K := y'_K / R_{KK}$. Note that ρ_k depends on $\{b_\ell\}_{\ell=k+1}^K$.

Using (25), the hyper-solid in (24) can be expressed as

$$D_1^K(\mathbf{b}) = \sum_{k=1}^{K} \left\{ \frac{R_{kk}^2}{2} (\rho_k - b_k)^2 + \lambda |b_k|_0 \right\} < \tau \quad (26)$$

or, in a more compact form as $D_1^K(\mathbf{b}) = \sum_{k=1}^K d_k(b_k) < \tau$, where $d_k(b_k) := (R_{kk}^2/2)(\rho_k - b_k)^2 + \lambda |b_k|_0$. In addition to the overall metric D_1^K assessing a candidate vector $\mathbf{b} \in \mathcal{A}_a^K$, as well as the per entry metric d_k for each candidate symbol $b_k \in \mathcal{A}_a$, it will be useful to define the *accumulated* metric $D_k^K := \sum_{\ell=k}^K d_\ell(b_\ell)$ corresponding to the $K-k+1$ candidate symbols from entry K down to entry k.

[2]At initialization, τ is set equal to ∞ so that (24) is always satisfied.

Per entry k, which subsequently will be referred to as level k, eq. (26) implies a set of inequalities:

$$\text{Level } k: \quad d_k(b_k) < \tau - D_{k+1}^K, \qquad \text{for } k = K, K-1 \ldots 1, \quad (27)$$

with $D_{K+1}^K := 0$. SSD relies on the Schnorr-Euchner (SE) enumeration, see e.g., [9], properly adapted here to account for the ℓ_0-norm regularization. SE capitalizes on the inequalities (27) to search efficiently over all possible vectors $\mathbf{b}$ with entries belonging to $\mathcal{A}_a$. Any candidate $\mathbf{b} \in \mathcal{A}_a^K$ obeying the K inequalities in (27) for a given τ, will be termed *admissible*. Threshold τ is reduced after each admissible $\mathbf{b}$ is identified, as will be detailed soon. The SE-based SSD amounts to a *depth-first* tree search, which seeks and checks candidate vectors starting from entry (level) K and working downwards to entry 1 per candidate vector.

At level K, SE search chooses $b_K = \lfloor \rho_K \rceil \mathbb{1}(2\rho_K \lfloor \rho_K \rceil - \lfloor \rho_K \rceil^2 - 2\lambda / R_{KK}^2 > 0)$, which we know from (21) is the constellation point yielding the smallest d_K. If this choice of b_K does not satisfy the inequality (27) with $k = K$, no other constellation point will satisfy it either, and the minimizer of D_1^K in (23) must lie outside[3] the hyper-solid postulated by (24). If this choice of b_K satisfies (27), SE proceeds to level $K-1$ in which (25) is used first with $k = K-1$ to find ρ_{K-1} that depends on the chosen b_K from level K; subsequently, b_{K-1} is selected as $b_{K-1} = \lfloor \rho_{K-1} \rceil \mathbb{1}(2\rho_{K-1} \lfloor \rho_{K-1} \rceil - \lfloor \rho_{K-1} \rceil^2 - 2\lambda / R_{K-1,K-1}^2 > 0)$. If this choice of b_{K-1} does not satisfy (27) with $k = K-1$, then we move back to level K, and select b_K equal to the constellation point yielding the *second* smallest d_K, and so on; otherwise, we proceed to level $K-2$. Continuing this procedure down to level 1, yields the first candidate vector $\hat{\mathbf{b}}$, which is deemed admissible since it has entries belonging to $\mathcal{A}_a$ and also satisfying (24). This candidate is stored, and the threshold is updated to $\tau := D_1^K(\hat{\mathbf{b}})$.

Then, the search proceeds looking for a better candidate. Now at level 1, we move up to level 2 and choose b_2 equal to the constellation point yielding the *second* smallest cost d_2. If this b_2 satisfies (27) at level 2 with the current τ, we move down to level 1 to update the value of b_1 (note that b_2 has just been updated and $\{b_\ell\}_{\ell=3}^K$ are equal to the corresponding entries in $\hat{\mathbf{b}}$). If (27) at level 2 is not satisfied with the current τ, we move up to level 3 to update the value of b_3, and so on.

Finally, when it fails to find any other admissible candidate satisfying (27), the search stops, and the latest admissible candidate $\hat{\mathbf{b}}$ is the optimal $\hat{\mathbf{b}}^{\mathrm{MAP}}$ solution sought. With $\tau = \infty$, the first found admissible candidate $\hat{\mathbf{b}}$ is the $\hat{\mathbf{b}}^{\mathrm{DDD}}$ solution of Section V-A.

Before summarizing the SSD steps, it is prudent to elaborate on the *ordered* enumeration of the constellation points *per level*, which in fact constitutes the main difference of SSD relative to SD. In lieu of the 0 constellation point and the ℓ_0 norm, SE in SD enumerates the PAM symbols per level k in the order of increasing cost as: $\{b_k, b_k + 2\Delta_k, b_k - 2\Delta_k, b_k + 4\Delta_k, b_k - 4\Delta_k, \cdots\} \bigcap \mathcal{A}$, with $b_k := \lfloor \rho_k \rceil$ and $\Delta_k := \mathrm{sign}(\rho_k - b_k)$. (With $\lfloor \rho_k \rceil$ yielding the smallest d_k,

[3]This will never happen with $\tau = \infty$ in (24).

if $\Delta_k = 1$, then $\lfloor \rho_k \rceil + 2$ yields the second smallest d_k and $\lfloor \rho_k \rceil - 2$ the third; and the other way around, if $\Delta_k = -1$.) SD effects such an ordered enumeration by alternately updating $b_k = b_k + 2\Delta_k$ and $\Delta_k = -\Delta_k - \text{sign}(\Delta_k)$ [9]. To demonstrate how SSD further accounts for the ℓ_0-norm regularization and the augmented alphabet of S-MAP which includes 0, let $b_k^{(i)} \in \mathcal{A}_a$ denote the symbol for level k incurring the i-th smallest ($i = 1, 2, \ldots, M+1$) cost d_k. If $b_k^{(i)} \in \mathcal{A}$, then $b_k^{(i+1)}$ will be either 0 or $b_k^{(i)} + 2\Delta_k$. If $d_k(0) < d_k(b_k^{(i)} + 2\Delta_k)$, then the next symbol in the ordered enumeration should be $b_k^{(i+1)} = 0$, and an auxiliary variable $b_k^{(c)}$ is used to cache the subsequent symbol in the order as $b_k^{(i+2)} = b_k^{(i)} + 2\Delta_k$. With $b_k^{(i+1)} = 0$, the auxiliary variable allows the wanted $b_k^{(i+2)}$ at the next enumeration step to be retrieved from $b_k^{(c)}$.

Similar to SD, the ordered enumeration pursued by SSD per level implies a corresponding order in considering all $\mathbf{b} \in \mathcal{A}_a^K$, which leads to a repetition-free and exhaustive search of all admissible candidate vectors. At the same time, the hyper-solid postulated by (24) shrinks as τ decreases, until no other admissible vector can be found. This guarantees that the SSD outputs the vector with the smallest D_1^K, and thus the optimal solution to (19). The SSD algorithm can be summarized in the following six steps 1–6 tabulated as Algorithm 2.

Remark 4: SSD inherits all the attractive features of SD [9]. Specifically, during the search one basically needs to store D_k^K per level k. Its *in place update* for each b_k candidate implies that SSD *memory requirements* are only linear in K. In addition, the *computational efficiency* of SSD (relative to that of ML which is $\mathcal{O}((M+1)^K)$) stems from four factors: (i) the DDD solution provides an admissible initialization reducing the search space at the outset; (ii) the recursive search enabled by the QR decomposition gives rise to the causally dependent inequalities (27), which restrict admissible candidate entries to choices that even decrease over successive depth-first passes of the search; (iii) ordering per level increases the likelihood of finding "near-optimal admissible" candidates early, which means quick and sizeable shrinkage of the hyper-solid, and thus fast convergence to the S-MAP optimal solution; and (iv) metrics involved in the search can be efficiently reused since children of the same level in the tree share the already computed accumulated metric of the "partial path" from this level to the root.

Compared to other sub-optimal detection schemes proposed in previous sections, the SSD algorithm can return the S-MAP optimal solution possibly at exponential complexity, unless one stops the search at the affordable complexity – case in which the solution is only ensured to be near-optimal. Fortunately, at medium-high SNR, both SD and SSD return the optimal solution at average complexity which is polynomial (typically cubic). Moreover, SSD can be generalized to provide symbol-by-symbol soft output with approximate *a posteriori* probabilities, as is the case with the SD; see e.g., [11, Chapter 5].

Algorithm 2 (SSD): Input $\tau = \infty$, $\mathbf{y}'$, $\mathbf{R}$, λ, and M even. Output the solution $\hat{\mathbf{b}}^{\text{MAP}} := \hat{\mathbf{b}}^{\text{SSD}}$ to (19).

1: (Initialization) Set $k := K$, $D_{K+1}^K := 0$.

2: Compute ρ_k as in (25), set $b_k := \lfloor \rho_k \rceil$, $b_k^{(c)} := 0$, and $\Delta_k := \text{sign}(\rho_k - b_k)$.
  If $2\rho_k b_k - b_k^2 - 2\lambda/R_{kk}^2 < 0$, then
    `// symbol 0 yields smaller` d_k `than` $\lfloor \rho_k \rceil$
    Set $b_k^{(c)} := b_k$, and $b_k := 0$.
  End if and go to Step 3.

3: **If** $d_k(b_k) := (R_{kk}^2/2)(\rho_k - b_k)^2 + \lambda |b_k|_0 \geq \tau - D_{k+1}^K$, then
    go to Step 4. `// outside hyper-solid in (24)`
  Else if $|b_k| > M - 1$, then
    go to Step 6. `// inside hyper-solid in (24), but outside` $\mathcal{A}_a$
  Else if $k > 1$, then
    compute $D_k^K := D_{k+1}^K + d_k(b_k)$; set $k := k - 1$, and go to Step 2. `// go the next level (deeper in the tree)`
  Else go to Step 5. `//` $k = 1$`, at the tree's bottom`
  End if

4: **If** $k = K$, then terminate
  Else set $k := k + 1$, go to Step 6.
  End if

5: (An admissible $\mathbf{b}$ is found)
  Set $\tau := D_2^K + d_1(b_1)$, $\hat{\mathbf{b}}^{\text{SSD}} := \mathbf{b}$, and $k := k + 1$; then go to Step 6.

6: (Enumeration at level k proceeds to the candidate symbol next in the order)
  If $b_k = 0$, then
    Retrieve the next (based on cost d_k ordering) symbol $b_k := b_k^{(c)}$, and set $b_k^{(c)} := FLAG$.
  Else set $b_k := b_k + 2\Delta_k$, and $\Delta_k := -\Delta_k - \text{sign}(\Delta_k)$.
    If $b_k^{(c)} \neq FLAG$ and $2\rho_k b_k - b_k^2 - 2\lambda/R_{kk}^2 < 0$, then
      `// 0 yields smaller` d_k `than` b_k
      Set $b_k^{(c)} := b_k$, and $b_k := 0$.
    End if
  End if and go to Step 3.

VI. Generalizations of S-MAP Detectors

Up to now, four sparsity-exploiting MUD algorithms have been developed to solve the integer program associated with the linear model in (1). The present section will present interesting generalizations to account for correlated user (in)activity across symbols, and under-determined CDMA systems.

A. Exploiting user (in)activity across symbols

Sparsity-aware detectors for the linear model in (1) were so far developed on a symbol-by-symbol basis, which does not account for the fact that user (in)activity typically persists across multiple symbols. To this end, user activity across time can for instance be thought of as a Markov chain with two states (active-inactive). Once a user terminal starts transmitting to the AP, it becomes more likely to stay active for the next symbol slot too; and likewise, inactive once it stops transmitting. In this model, the state transition probability from either one state to the other is relatively much smaller than that of staying unchanged, and this manifests itself to the said dependence of user (in)activities across time.

Admittedly, MUD schemes accounting for this dependence must process the aggregation of data vectors $\mathbf{y}$ obeying (1) across slots. With N_s denoting the number of slots, the number of unknowns (KN_s) can grow prohibitively with N_s. One approach to cope with this "curse of dimensionality" is via dynamic programming, which can take advantage of the fact that correlation is only present between two consecutive slots; see e.g., [17, Sec. 4.2]. However, for M-ary alphabets, the resultant sequential detector requires evaluating per symbol slot the path weights of all $(M+1)^K$ possible symbol vectors. This high computational burden is impractical for real-time implementations.

The proposed alternative to bypass this challenge stems from the observation that for a given slot t the influence of all previous slots $\{t'\}_{t'=0}^{t-1}$ on the S-MAP detection rule is reflected only in the prior probability of each user being active at time t; i.e., user k's current (and time-varying) activity factor $p_{a,k}(t)$. The natural means to capture this influence online is to track each user's activity factor using the recursive LS (RLS) estimator [12, Ch. 12], based on activity factors from previous slots; that is,

$$\hat{p}_{a,k}(t) = \arg\min_p \sum_{t'=0}^{t-1} \beta_k^{t,t'} (p - |\hat{b}_k(t')|_0)^2, \quad t = 1, \ldots \tag{28}$$

where $\hat{b}_k(t')$ denotes user k's detected symbol at time t', and the so-called "forgetting-factor" $\beta_k^{t,t'}$ describes the effective memory (data windowing). A popular choice is the exponentially decaying window, for which $\beta_k^{t,t'} := \beta_k^{t-t'}$ for some $0 \ll \beta_k < 1$. Accordingly, (28) can expressed in closed form, recursively as

$$\begin{aligned}\hat{p}_{a,k}(t) &= \frac{1-\beta_k}{\beta_k}\left(\sum_{t'=0}^{t-1} \beta_k^{t-t'} |\hat{b}_k(t')|_0\right) / (1-\beta_k^t) \\ &= \frac{\beta_k - \beta_k^t}{1-\beta_k^t}\hat{p}_{a,k}(t-1) + \frac{1-\beta_k}{1-\beta_k^t}|\hat{b}_k(t-1)|_0, \; t = 1, \ldots\end{aligned} \tag{29}$$

where the last equality comes from back substitution of $\hat{p}_{a,k}(t-1)$. The choice of β_k critically depends on the (in)activity correlation between consecutive slots. In the extreme case where user (in)activities across slots are independent, the infinite-memory window ($\beta_k = 1$) is optimal, and (29) reduces to the simple online time-averaging estimate $\hat{p}_{a,k}(t) = \frac{1}{t}\sum_{t'=0}^{t-1} |\hat{b}_k(t')|_0$.

Adapting the user activity factors allows one to weigh entries of ℓ_0-norm regularization which in turn affects the prior probability in the S-MAP detector (5) through the coefficient $\lambda_k(t)$ corresponding to $\hat{p}_{a,k}(t)$ (cf. Remark 2). Note that when the correlation across time is strong, it is possible that $\hat{p}_{a,k}(t)$ can approach 1, case in which $\lambda_k(t)$ is not guaranteed to stay positive. This will cause problems to the relaxed S-MAP detectors of Section IV, as those schemes rely on the convexity of the cost function in (7). However, the DDD and SSD algorithms will remain operational, because they rely on enumeration per symbol (in DDD) or per group of symbols within a sphere (in SSD). Note also that with $\lambda < 0$ the regularization term in the minimization of (19) is non-positive; hence, $\lambda < 0$ encourages searching over the non-zero constellation points (and thus discourages sparsity), whereas $\lambda > 0$ promotes sparsity.

B. Under-determined CDMA Systems

Minimal-size spreading sequences, even smaller than the number of users, is well motivated for bandwidth and power savings. Without finite-alphabet constraints on the wanted vector, results available in the CS literature guarantee recovery of sparse signals from a few observations; see e.g., [4], [5] and references therein. Specifically, [4] shows that if the vector of interest is sparse (or compressible) over a known basis, then it is possible to reconstruct it with very high accuracy from a small number of random linear projections at least in the ideal noise-free case. For non-ideal observations corrupted with unknown noise of bounded perturbation, [5] provides an upper bound on the reconstruction error, which is proportional to the noise variance for a sufficiently sparse signal. However, CS theory pertains to sparse analog-valued signals. Moreover, the noise considered in a practical communication system is typically Gaussian, or generally drawn from a distribution having possibly unbounded support. Therefore, existing results from the CS literature do not carry over to the present context.

Nevertheless, it is still interesting to consider extensions of all the (sub-)optimal sparsity-exploiting MUD algorithms to an under-determined CDMA system with $N < K$, where the observation matrix $\mathbf{H}$ becomes fat. Consider first the two types of relaxed S-MAP detectors. The RD-MUD in (10) clearly works when $N < K$, since the $2\lambda\mathbf{I}$ term inside the inversion renders the overall matrix full rank. However, since the RD is a linear detector, it is expected to lose identifiability in the under-determined case, similar to the MMSE detectors for the sparsity-agnostic MUD schemes. Similarly, the LD problem (17) is also solvable for a fat $\mathbf{H}$ matrix, as the Lasso problem in CS. However, neither of them accounts for the augmented finite-alphabet constraint present in the original S-MAP problem (7).

The S-MAP detectors with lattice search are challenging to implement when $N < K$. The main obstacle is the QR decomposition of the fat matrix $\mathbf{H}$, which yields the upper triangular matrix $\mathbf{R}$ of the same dimension $N \times K$. Instead of a single unknown symbol, now the sparse DDD must optimize over the last $(N - K + 1)$ symbols in $\mathbf{b}$. However, apart from exhaustive search there is no low-complexity method to solve the aforementioned problem involving $(N - K + 1)$ variables, because sub-optimal alternatives introduce severe error propagation.

The same problem appears also with the SSD. To tackle the under-determined case, the generalized SD in [9] fixes the last $(N - K)$ symbols of $\mathbf{b}$ and relies on the standard SSD to detect the remaining K symbols that minimize a cost similar to the one in (23). Repeating this search for every choice of the last $(N - K)$ symbols, yields eventually the overall optimum vector. The complexity of the latter is exponential in $(N - K)$, regardless of the SNR. Recently, an alternative SD approach to avoid this exponential complexity has been developed for the under-determined case [8]. This algorithm takes advantage of the fact that for constant-modulus constellations the usual LS cost can be modified without affecting optimality, by adding the ℓ_2-norm $\mathbf{b}^T\mathbf{b}$ constant for every vector in the alphabet. This extra term allows one to obtain an equivalent full-rank system on which the standard SD algorithm can be applied. This efficient method can be readily extended to handle non-constant modulus constellations.

Interestingly, for our S-MAP detectors in (7) with lattice search of binary transmitted symbols, the norm term needed for regularization comes naturally from the Bernoulli prior. Specifically, with $p = 2$ the reformulated S-MAP detectors in (7) can be equivalently written as

$$\hat{\mathbf{b}}^{\text{MAP}} = \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}[\mathbf{b}^T(\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I})\mathbf{b} - 2\mathbf{y}^T\mathbf{H}\mathbf{b}]$$
$$= \arg\min_{\mathbf{b}\in\mathcal{A}_a^K} \frac{1}{2}\|\tilde{\mathbf{y}}' - \tilde{\mathbf{R}}\mathbf{b}\|_2^2 \qquad (30)$$

where $\tilde{\mathbf{R}}$ is the full rank $K \times K$ upper triangular matrix such that $\tilde{\mathbf{R}}^T\tilde{\mathbf{R}} = \mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{I}$, and $\tilde{\mathbf{y}}' := \tilde{\mathbf{R}}^{-T}\mathbf{H}^T\mathbf{y}$. Utilizing the metric of (30), the back substitution of DDD and the lattice point search of SSD can be implemented easily. Hence, these S-MAP detectors can be readily extended to under-determined systems. In this way, all the (sub-)optimal S-MAP detectors are applicable with less observations than unknowns in a CDMA system with low activity factor, but their SER performance will certainly be affected. In Section VII, we will provide simulated performance comparisons of the different optimal and sub-optimal S-MAP algorithms proposed, for a variable number of observations.

C. Group Lassoing block activity

The last generalization considered pertains to user (in)activity in a (quasi-)synchronous block fashion, where during a block of N_s symbol slots, user k remains (in)active independently from other users and across blocks. Concatenate the K user symbols across N_s time slots in the $K \times N_s$ matrix $\mathbf{B} := [\mathbf{b}(1) \ldots \mathbf{b}(N_s)]$, where $\mathbf{b}(t)$ collects the symbols of all K users at slot t, and likewise for the receive-data matrix $\mathbf{Y}$ as well as the noise matrix $\mathbf{W}$, both of size $N \times N_s$. With these definitions, the counterpart of (1) for this block model is $\mathbf{Y} = \mathbf{H}\mathbf{B} + \mathbf{W}$. Letting the $N_s \times 1$ vector $\breve{\mathbf{b}}_k := [b_k(1), \ldots, b_k(N_s)]^T$ collect the N_s symbols of user k, it is useful to consider it drawn from an augmented (due to possible inactivity) block alphabet $\mathcal{A}_{a,N_s} := \mathcal{A}^{N_s} \bigcup\{\mathbf{0}_{N_s}\}$. Assuming again binary transmissions, the S-MAP *block* detector will now yield

$$\hat{\mathbf{B}}^{\text{MAP}} = \arg\min_{\breve{\mathbf{b}}_k\in\mathcal{A}_{a,N_s}} \frac{1}{2}\|\mathbf{Y} - \mathbf{H}\mathbf{B}\|_F^2 + \sum_{k=1}^{K} \frac{\lambda_b}{\sqrt{N_s}}\|\breve{\mathbf{b}}_k\|_0$$
$$= \arg\min_{\breve{\mathbf{b}}_k\in\mathcal{A}_{a,N_s}} \frac{1}{2}\|\mathbf{Y} - \mathbf{H}\mathbf{B}\|_F^2 + \sum_{k=1}^{K} \lambda_b\|\breve{\mathbf{b}}_k\|_2 \quad (31)$$

where $\lambda_b := \frac{1}{\sqrt{N_s}} \ln\left(\frac{1-p_a}{p_a/2^{N_s}}\right)$.

Similar to (7), the convex reformulation of the cost in (31) will lead to what is referred to in statistics as *Group Lasso* [19], which effects group sparsity on a block of symbols ($\breve{\mathbf{b}}_k$ in our case). This Group-Lasso based formulation is particularly handy for under-determined CDMA systems. In fact, the unconstrained version of (31) can be solved first to unveil the nonzero rows (i.e., the support) of $\mathbf{B}$, with improved reliability as N_s increases. Subsequently, standard sparsity-agnostic MUD schemes can be run on the estimated set of active users. Note that such a two-step approach works in the under-determined case, and also reduces number of symbols to be detected per time slot.

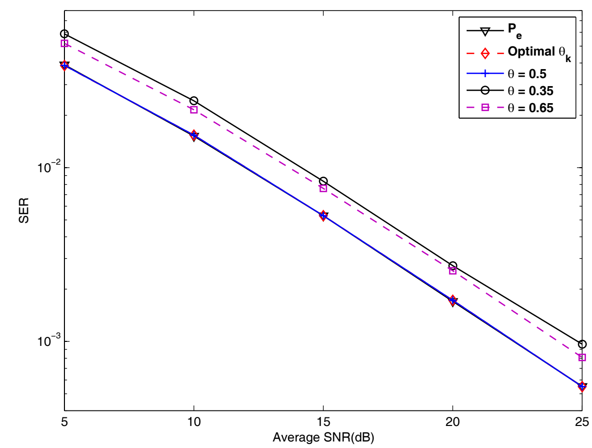

Fig. 2. SER vs. average SNR (in dB) for RD-MUD with $N = 32$ and $K = 20$ and different quantization threshold θ's.

VII. Simulations

We simulate a CDMA system with $K = 20$ users, each with activity factor $p_a = 0.3$. Random sequences with length $N = 32$ are used for spreading. We consider non-dispersive independent Rayleigh fading channels between AP and users, where the channel gain g_k of the k-th user is Rayleigh distributed with variance $E[g_k^2] = \sigma^2$. Thus, the average system SNR is set to be σ^2 since the AWGN $\mathbf{w}$ has unit variance.

Test Case 1 (Quantization thresholds for RD). First, we test the RD scheme with different quantization thresholds θ in (9). The optimum threshold for the k-th symbol is obtained as in (15) per channel realization $\mathbf{H}$. The resulting SER is compared for four choices of θ: $0.5,\ 0.35,\ 0.65$, and $\hat{\theta}_k$. The theoretical minimum SER $\hat{P}_{e,k}^{\text{RD}}$ in (16) using the optimum θ is also added for comparison. Fig. 2 shows that the SER curve with $\theta = 0.5$ comes very close to the one of the optimal $\hat{\theta}_k$, and thus constitutes a near-optimal choice in practice. Moreover, those two curves also coincide with the analytical SER formulation corresponding to $\hat{P}_{e,k}^{\text{RD}}$, thus corroborating the closed-form expression in (16).

Test Case 2 (S-MAP MUD algorithms). Next, the RD, LD, DDD, and SSD MUD algorithms are all tested for both BPSK and 4-PAM constellations, and their SER performance is compared. For LD, the quadratic program in (17) is solved using the general-purpose SeDuMi toolbox [13]. The quantization rule chooses the nearest point in $\mathcal{A}_a$ for both RD and LD. For comparison, we also include the ordinary LS detector, which corresponds to the RD solution in (10) with $\lambda = 0$.

Fig. 3(a) shows that the LS detector exhibits the worst performance. This is intuitive since it neither exploits the finite alphabet nor the sparsity present in $\mathbf{b}$. The SSD exhibits the best performance at the price of highest complexity. The LD outperforms the RD algorithm, as predicted. It is interesting to observe that even at low SNR region the DDD algorithm is surprisingly competitive, especially in view of its low complexity that grows only linearly in the number of symbols K. The diversity orders for those detectors are basically the same.

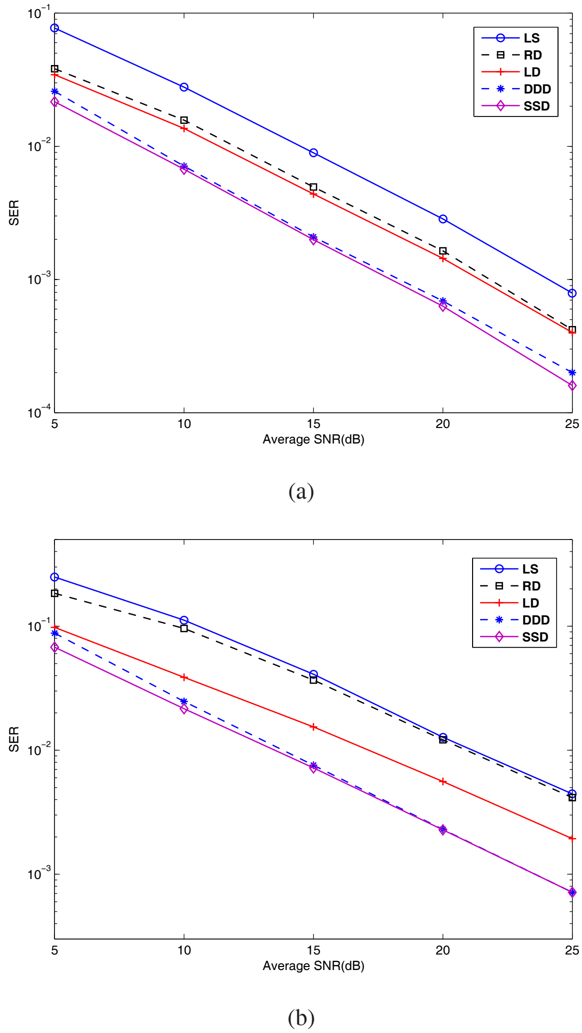

Fig. 3. SER vs. average SNR (in dB) of sparsity-exploiting MUD algorithms with $N = 32$ and $K = 20$ for (a) BPSK, and (b) 4-PAM alphabets.

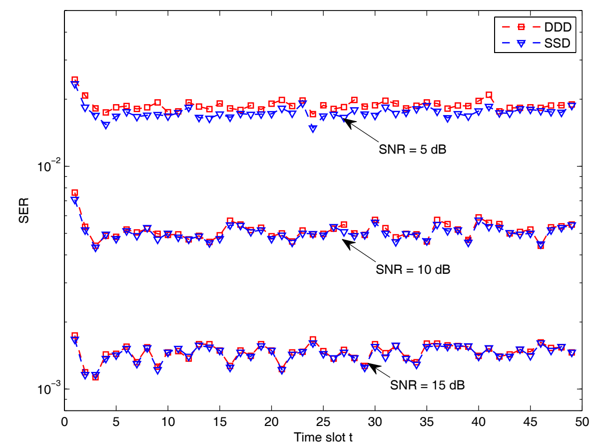

Fig. 4. SER vs. time t of sparsity-exploiting MUD algorithms with RLS estimation of the activity factors.

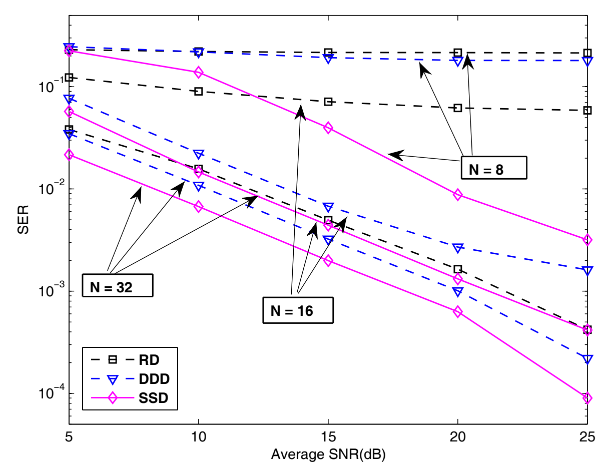

Fig. 5. SER vs. average SNR (in dB) of sparsity-exploiting MUD algorithms with $N = 32, 16$, or 8 and $K = 20$.

This is reasonable since independent Rayleigh fading channels between AP and users were simulated here. The corresponding curves for 4-PAM depicted in Fig. 3(b) follow the same trend. However, compared to Fig. 3(a), the RD algorithm degrades noticeably as its SER approaches the LS one. This is because choices of quantization thresholds become more influential as the constellation size increases. As expected, the LD exhibits resilience to this influence. The DDD and SSD algorithms have almost identical performance in high-SNR region.

Test Case 3 ((In)activity across symbols). In this case, the user (in)activity is correlated across time slots. We model this random (in)activity process as a two-state (active-inactive) stationary Markov chain. The state transition matrix is $\mathbf{P} = [a\ (1-a);\ b\ (1-b)]$, where a is uniformly distributed over $[0.8\ 0.85]$, and b over $[0.05\ 0.1]$, for each user. For this model, the expected number of successive active slots is $1/(1-a)$, and $1/b$ for the inactive ones. Also, the limiting probability for the "active" state becomes $b/(1-a+b)$, taking values from the interval $[0.2\ 0.4]$. Note that the activity factor over time is still quite low. We use the RLS approach to estimate $\hat{p}_{a,k}(t)$ as in (29) using $\beta = 0.5$, and test both the DDD and SSD algorithms in solving the resultant S-MAP detection problem. The empirical SER is plotted in Fig. 4 across time for different SNR values. Clearly, the proposed scheme is effective in tracking the evolving time-correlated activity. For the same SNR value, it yields SER performance similar to the independent case in Fig. 3(a).

Test Case 4 (Under-determined CDMA systems). We also test the S-MAP MUD algorithms for under-determined systems, by varying N from 32 to 16 and 8. The results are depicted in Fig. 5. Since the RD is a simple linear detector, it is expected that once $N < K$, it will lose identifiability, and exhibits a considerably flat SER curve. At the same time, DDD still enjoys almost full diversity with a moderate choice of $N = 16$. Being the optimum detector, the SSD collects the full diversity even if $N = 8$; however, the other two kinds of detectors exhibit flat SER curves, as expected.

The Group Lasso scheme for recovering block activity is also included for the under-determined case. Fig. 6 illustrates the activity recovery error rate for different values of N and N_s. The number of observations N affects the diversity order,

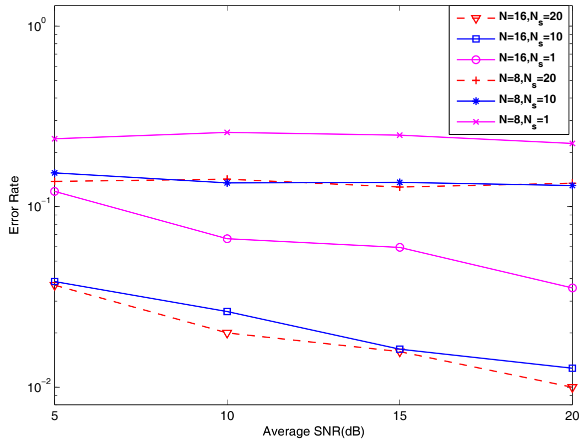

Fig. 6. Error rate for block activity vs. average SNR (in dB) of Group Lasso algorithm with $N = 16, 8$ and $N_s = 20, 10, 1$.

while the block size N_s influences the recovery accuracy.

VIII. Conclusions

The MUD problem of sparse symbol vectors emerging with CDMA systems having low-activity factor was considered. Viewing user inactivity as augmenting the underlying alphabet, the *a priori* available sparsity information was exploited in the optimal S-MAP detector. The exponential complexity associated with the S-MAP detector was reduced by resorting to (sub-) optimal algorithms. Relaxed S-MAP detectors (RD and LD) come with low complexity but sacrifice optimality, because they ignore the finite-alphabet constraint. The second kind of detectors (DDD and SSD), searches over a subset of the alphabet and exhibits improved performance at increased complexity. The performance was analyzed for the RD and DDD algorithms, and closed-form expressions were derived for the corresponding symbol error rates.

S-MAP detectors were further generalized to deal with correlated user (in)activity across symbols by recursively estimating each user's activity factor online; and also with under-determined (a.k.a. over-saturated CDMA) settings emerging when the spreading gain is smaller than the potential number of users. Coping with the latter becomes possible through regularization with the ℓ_0 norm prior, or, with the Group Lasso-based recovery of the active user set. The numerical tests corroborated our analytical findings and quantified the relative performance of the novel sparsity-exploiting MUD algorithms.

Our future research agenda includes analytical performance evaluation in the under-determined case, which has been available for the recovery of sparse signals but without finite-alphabet constraints. In addition, it will be interesting to develop S-MAP detectors based on lattice search for higher-order constellations in the under-determined case.

Appendix A
Proof of (21)

To solve the optimization problem (20), consider first the unconstrained solution to the LS part of the cost, which can be written as

$$b_k^{\text{LS}} := \left(y'_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_\ell^{\text{DDD}}\right)/R_{kk}. \quad (32)$$

The detected symbol in (20) can be equivalently expressed as

$$\hat{b}_k^{\text{DDD}} = \arg\min_{b_k\in\mathcal{A}_a} f(b_k),$$
$$f(b_k) := \left(b_k^{\text{LS}} - b_k\right)^2 + (2\lambda/R_{kk}^2)|b_k|_0. \quad (33)$$

The solution $\hat{b}_k^{\text{DDD}}$ can be obtained by comparing $f(0)$ with $\min_{b_k\in\mathcal{A}} f(b_k)$. Specifically, as the cost $f(b_k)$ is quadratic for $b_k \in \mathcal{A}$, the minimum is achieved at $f(\lfloor b_k^{\text{LS}}\rceil)$, by quantizing b_k^{LS} to the nearest point in $\mathcal{A}$. Thus, $\hat{b}_k^{\text{DDD}} = 0$ only if $f(0) \le f(\lfloor b_k^{\text{LS}}\rceil)$, or equivalently, after using the definition of $f(\cdot)$, if $2b_k^{\text{LS}}\lfloor b_k^{\text{LS}}\rceil - \lfloor b_k^{\text{LS}}\rceil^2 - 2\lambda/R_{kk}^2 \le 0$. This completes the proof of (21).

Appendix B
Proof of (22)

When $M = 2$, the DDD solution (21) reduces to

$$\hat{b}_k^{\text{DDD}} = \text{sign}(b_k^{\text{LS}})\mathbb{1}(2|b_k^{\text{LS}}| - 1 - 2\lambda/R_{kk}^2 > 0) \quad (34)$$

due to the fact that $\lfloor b_k^{\text{LS}}\rceil = \text{sign}(b_k^{\text{LS}}) \in \{\pm 1\}$.

Recalling that $\check{\mathbf{b}}$ denotes the transmitted vector, and substituting $\mathbf{y} = \mathbf{H}\check{\mathbf{b}} + \mathbf{w}$ yields

$$\mathbf{y}' := \mathbf{Q}^T\mathbf{y} = \mathbf{R}\check{\mathbf{b}} + \mathbf{u} \quad (35)$$

where $\mathbf{u} := \mathbf{Q}^T\mathbf{w}$ is zero-mean Gaussian with identity covariance matrix. Supposing that there is no error propagation, the b_k^{LS} term in (34) becomes [cf. (32)]

$$b_k^{\text{LS}} = \left(\sum_{\ell=k}^{K} R_{k\ell}\check{b}_\ell + u_k - \sum_{\ell=k+1}^{K} R_{k\ell}\hat{b}_k^{\text{DDD}}\right)\Big/ R_{kk}$$
$$= \check{b}_\ell + u_k/R_{kk} \quad (36)$$

where u_k denotes the k-th entry of $\mathbf{u}$.

To analyze the error probability for the detector in (34) consider the following three cases.

a) $\check{b}_k = 0$ is sent: under this case, $b_k^{\text{LS}} = u_k/R_{kk}$ and a detection error emerges when $\hat{b}_k^{\text{DDD}} = 1$ or -1. With the closed-form DDD detector (34) in mind, such an error occurs only if both $b_k^{\text{LS}} \ne 0$ and $2|b_k^{\text{LS}}| - 2\lambda/R_{kk}^2 - 1 > 0$ hold. The first case corresponds to $u_k \ne 0$ and the second one is equivalent to $|u_k| > |R_{kk}|/2 + \lambda/|R_{kk}|$, which is included in the event $u_k \ne 0$. Hence, to evaluate the error probability it suffices to consider only the case $|u_k| > |R_{kk}|/2 + \lambda/|R_{kk}|$.

b) $\check{b}_k = 1$ is sent: following the analysis in a), an error occurs if $\text{sign}(R_{kk})u_k \le -|R_{kk}|/2 + \lambda/|R_{kk}|$.

c) $\check{b}_k = -1$ is sent: following the analysis in a), an error occurs if $\text{sign}(R_{kk})u_k \ge |R_{kk}|/2 - \lambda/|R_{kk}|$.

Given that the Gaussian distributed u_k has variance 1, the overall SER for the k-th entry b_k becomes

$$P_{e,k}^{\text{DDD}} = \sum_{i=0,\pm1} P(\text{error}|\check{b}_k = i)P(\check{b}_k = i)$$
$$= 2(1-p_a)Q\left(\frac{|R_{kk}|}{2} + \frac{\lambda}{|R_{kk}|}\right)$$
$$+ p_a Q\left(\frac{|R_{kk}|}{2} - \frac{\lambda}{|R_{kk}|}\right). \quad (37)$$

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their helpful feedback which improved exposition of this paper.

REFERENCES

[1] D. Angelosante, E. Grossi, G. B. Giannakis, and M. Lops, "Sparsity-aware estimation of CDMA system parameters," *EURASIP J. Advances Signal Process.*, June 2010, article ID 417981.

[2] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, pp. 118-121, July 2007.

[3] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708-1721, Mar. 2010.

[4] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406-5425, Dec. 2006.

[5] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21-30, Mar. 2008.

[6] M. Cetin and B. M. Sadler, "Semi-blind sparse channel estimation with constant modulus symbols," in *Proc. Intl. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, Mar. 2005.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comput.*, vol. 20, pp. 33-61, 1998.

[8] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Commun. Lett.*, vol. 9, pp. 423-425, May 2005.

[9] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2389-2402, Oct. 2003.

[10] A. K. Fletcher, S. Rangan, and V. K. Goyal, "A sparsity detection framework for on-off random access channels," in *IEEE Proc. Intl. Symp. Inf. Theory*, pp. 169-173, Seoul, South Korea, June 2009.

[11] G. B. Giannakis, Z. Liu, X. Ma, and S. Zhou, *Space-Time Coding for Broadband Wireless Communications.* John Wiley & Sons, Inc., 2007.

[12] A. H. Sayed, *Fundamentals of Adaptive Filtering.* John Wiley & Sons, 2003.

[13] J. F. Sturm "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods Software*, vol.11-12, pp. 625-653, 1999. [Online], Available: http://sedumi.mcmaster.ca.

[14] Z. Tian, G. Leus, and V. Lottici, "Detection of sparse signals under finite-alphabet constraints," in *Proc. Intl. Conf. Acoust., Speech Signal Process.*, pp. 2349-2352, Taipei, Apr. 2009.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statistical Soceity: Series B*, vol. 58, pp. 267-288, 1996.

[16] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Research*, vol. 1, pp. 211-244, 2001.

[17] S. Verdu, *Multiuser Detection.* Cambridge, UK: Cambridge Univ. Press, 1998.

[18] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory* , vol. 45, pp. 1639-1642, July 1999.

[19] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statistical Soceity: Series B*, vol. 68, pp. 49-67, 2006.

[20] H. Zhu and G. B. Giannakis, "Sparsity-embracing multiuser detection for CDMA systems with low activity factor," in *Proc. IEEE Intl. Symp. Inf. Theory*, pp. 164-168, Seoul, South Korea, June 2009.

Hao Zhu (S'07) received her Bachelor degree from Tsinghua University, Beijing, China in 2006 and the M.Sc. degree from the University of Minnesota, Minneapolis in 2009, both in electrical engineering.

From 2005 to 2006, she was working as a research assistant with the Intelligent Transportation information Systems (ITiS) Laboratory, Tsinghua University. Since September 2006, she has been working towards her Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota. Her research interests include sparse signal recovery and blind source separation in signal processing.

Georgios B. Giannakis (Fellow'97) received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. Since 1999 he has been a professor with the Univ. of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the ECE Department, and serves as director of the Digital Technology Center.

His general interests span the areas of communications, networking and statistical signal processing, subjects on which he has published more than 300 journal papers, 500 conference papers, two edited books and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile ad hoc networks, wireless sensor and social networks. He is the (co-) inventor of 18 patents issued, and the (co-) recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.