Limited Feedback For Cognitive Radio Networks Using Compressed Sensing

Harish Ganapathy[†], Constantine Caramanis[†] and Lei Ying[††]
[†] Department of Electrical and Computer Engineering
The University of Texas, Austin
Austin, TX 78712, USA
[††] Department of Electrical and Computer Engineering,
Iowa State University,
Ames, IA 50011, USA
E-mail: {harishg,caramanis}@mail.utexas.edu,leiying@iastate.edu

Abstract— In this paper, we consider the downlink of a cognitive radio network where a cognitive base station serves multiple cognitive users on the same frequency band as a group of primary transceivers. The cognitive base station uses an orthogonal scheduling scheme (TDMA/FDMA) to serve its users. For this purpose, the base station is interested in acquiring an estimate of the interference (from the primary network) power at each of its cognitive receivers as a measure of channel quality. This can be surely achieved if we allow for the feedback (from the cognitive receivers to the cognitive base station) bandwidth to scale linearly in the number of cognitive receivers, but in densely populated networks, the cost of such an acquisition might be too high. This leads us to the question of whether we can do better in terms of bandwidth efficiency. We observe that in many scenarios – that are common in practice – where the primary network exhibits sparse changes in transmit powers from one scheduling instant to the next, it is possible to acquire this interference state with only a logarithmic scaling in feedback bandwidth. More specifically, in cognitive networks where the channels are solely determined by the positions of nodes, we can use *compressed sensing* to recover the interference state. In addition to being a first application of compressed sensing in the domain of limited feedback, to the best of our knowledge, this paper makes a key mathematical contribution concerning the favourable sensing properties of path-loss matrices that are composed of non-zero mean, dependent random entries. Finally, we numerically study the robustness properties of the least absolute shrinkage and selection operator (LASSO), a popular recovery algorithm, under two error models through simulations. The first model considers a varying amount of error added to all entries of the sensing matrix. The second one, a more adversarial model, considers a large amount of error added to only a fraction of the entries of the sensing matrix that are chosen uniformly at random. Simulation results establish that the LASSO recovery algorithm is robust to imperfect channel knowledge.

I. INTRODUCTION

With the tremendous increase in wireless connectivity over the last decade, the demand for wireless spectrum has never been greater. Traditionally, a portion of spectrum is allocated or licensed for exclusive use by a specific group of users by regulatory agencies. This inherent rigidity coupled with the growing demand for wireless applications has led to a scarcity of spectrum. However, a recent Federal Communications Commission (FCC) study [1] has revealed that large portions of spectrum, though allocated, are significantly under-utilized by the licensees. Cognitive radio technology, introduced by J. Mitola [2], is a promising solution to this problem, one that can potentially increase the spectrum utilization efficiency as recognized by the FCC [3]. This technology calls for cognitive (unlicensed/secondary) users to operate on the same frequency band as the primary licensed users while attempting to access the spectrum seamlessly. In other words, these users adjust their operating parameters to cause only minimal impact to the primary licensed users.

We introduce our problem of interest in the context of cognitive radio networks before presenting a review of related work in the field.

A. The Problem

We have a number of primary transceiver pairs scattered in a given geographical area. There is a cognitive radio network that shares the same geographical area (and spectrum by definition). The cognitive radio network consists of a cognitive base station (cognitive transmitter) and a number of cognitive users or receivers as shown in Fig. 1. The cognitive base station serves its users in an orthogonal manner (e.g. time-division-multiple-access). In order to decide which cognitive user to serve in each scheduling time slot, it is necessary for the cognitive base station to possess information pertaining to the quality of the wireless link between itself and each of its users. Consequently, it is interested in estimating the interference power received by each cognitive node from the primary network. Channel or interference state information in a frequency-division-duplexing (FDD) system is typically acquired through an orthogonal feedback channel between each receiver and the base station[1]. Thus, it is possible to acquire full channel state information by incurring a linear increase of feedback bandwidth in the number of users. But

[1]The feedback channel is necessary since channel reciprocity cannot be exploited in an FDD system as can be done in a time-division-duplexing system. The feedback channel is also assumed to be orthogonal to the primary network.

is this the most bandwidth-efficient method of acquisition? We show that when the primary users exhibit limited changes in their transmitted power levels between adjacent time slots, as is the case in many common wireless applications, it is possible to acquire channel state information with only a logarithmic increase of feedback bandwidth in the number of users.
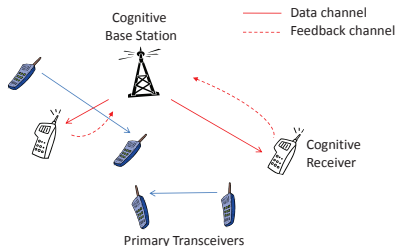

Fig. 1. Downlink of a cognitive radio network that shares the same spectrum as a group of primary transceivers.

B. Literature review

Cognitive radios have been categorized into three operating paradigms based on the type of network side information that is available along with the regulatory constraints imposed by the primary network: *underlay*, *overlay*, and *interweave* [4]. While there has been significant progress in engineering systems under each of these paradigms, we will only review literature that is relevant to the underlay paradigm, which is of interest in this paper. These are systems that allow cognitive users to operate under the condition that the interference caused to the primary users is below an acceptable threshold [5].

Literature on cognitive underlay systems predominantly tackles physical layer design issues. In [6], [7], adaptive power and rate policies are designed for the cognitive transmitters given a constraint on the maximum allowable interference at the primary receivers. Hossain et al. [8] consider power allocation policies for the cognitive users under a minimum performance constraint for each cognitive user in addition to the maximum allowable interference constraint on the primary users.

In contrast to the above literature on cognitive underlay networks, we are interested in the scheduling aspects pertaining to the downlink of a cognitive underlay network. In particular, we focus on the problem of efficient acquisition of channel state information by the cognitive base station in order to facilitate user (cognitive) scheduling. As mentioned earlier, we wish to accomplish this task with a logarithmic increase of feedback bandwidth in the number of cognitive users (as opposed to a linear scaling) using a technique called compressed sensing. This work differs from past work such as Gopalan et al [9] and Ouyang et al. [10] in that we exploit the underlying sparse structure induced by the application to accomplish bandwidth reduction.

Compressed sensing has received tremendous interest in the recent past and it has found many applications in wireless systems. It has been used in the context of multipath channel estimation (see [11]–[13] and references therein). This is possible since a wideband channel typically has a large number of channel taps but with energy concentrated only in a few of them. It has also recently been used to solve spectrum sensing problems in cognitive networks operating in the interweave paradigm (see Yücek et al. [14] for a survey on the topic). In this paradigm, the cognitive radios communicate opportunistically by sensing holes in the spectrum thereby intrinsically relying on intermittent activity patterns of the primary users.

C. Main Contributions and Organization

The main contributions of this paper are as follows:

1) A first (to the best of our knowledge) application of compressed sensing in designing limited feedback protocols for cognitive radio networks.
2) A proof that path-loss matrices satisfy the *null space property*. Such channel matrices allow for efficient acquisition of the interference state using ℓ_1-norm minimization. By efficient, we mean logarithmic scaling in feedback bandwidth. The proof technique is novel since path-loss matrices contain entries that have non-zero mean and are not independent, a scenario that has not been dealt with extensively in past research.
3) A numerical study of the robustness of another popular recovery algorithm to imperfect knowledge of channel state. Numerical evidence establishes favourable robustness properties of the algorithm.

The rest of this paper is organized as follows. In Section 2, we introduce the system model for the cognitive radio network under consideration. In Section III, we prove the "goodness" of path-loss sensing matrices. In Section IV, we study the recovery performance of compressed sensing under imperfect channel knowledge.

Notation: x_{ij} denotes element (i,j) of matrix $\mathbf{X}$ while x_i denotes element i of vector $\mathbf{x}$. $(.)^T$ is the transpose operator. For $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{x}_{\mathcal{A}}$, $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ denotes the vector $\mathbf{x}$ restricted to the entries in $\mathcal{A}$. For matrices $\mathbf{X} \in \mathbb{R}^{k \times N}$, $\mathbf{X}_{\mathcal{A},r}$, $\mathcal{A} \subseteq \{1, 2, \ldots, k\}$ denotes the sub-matrix of $\mathbf{X}$ formed by the rows contained in $\mathcal{A}$. $\mathbf{X}_{\mathcal{A},c}$, $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ denotes the sub-matrix of $\mathbf{X}$ formed by the columns contained in $\mathcal{A}$. Finally, $|| \cdot ||_p$ is the p-norm operator on vectors.

II. SYSTEM MODEL

Each primary transmitter is dropped uniformly on a disc of radius r_p centered at the origin. There are a total of N primary transmitters located at points $\{(X_{p,i}, Y_{p,i})\}_{i=1}^N$ on the disc. There are N cognitive receivers[2] placed on a collection of many circles of radii $\{r_{c,1}, r_{c,2}, \ldots, r_{c,q}\}$ where q is such that $\frac{N}{q} \in \mathbb{Z}^+$ for simplicity and $\mathbb{Z}^+$ represents the

---

[2]We set the number of cognitive receivers equal to the number of primary transmitters for simplicity. The analysis can be extended to cover the general case.

set of non-negative integers. Each circle contains $\frac{N}{q}$ cognitive receivers located at fixed points $\{(X_{c,i}, Y_{c,i})\}_{i=1}^{N}$ that are equally-spaced[3] as shown in Fig. 2. The N corresponding primary receivers that complete the primary network and the cognitive base station that completes the cognitive network can be arbitrarily located anywhere on the xy-plane. The above spatial distribution model is chosen to provide tractability in the analysis that follows. In the simulations section, we will show that the proposed algorithms work even under more regular spatial models such as users scattered uniformly at random on a square area. For the sake of the analysis, we will also partition the cognitive receivers according to the circle they belong to thus creating q partitions $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_q\}$ such that $\bigcup_{i=1}^{q} \mathcal{P}_i = \{1, 2, \ldots, N\}$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for $i \neq j$. The distance between
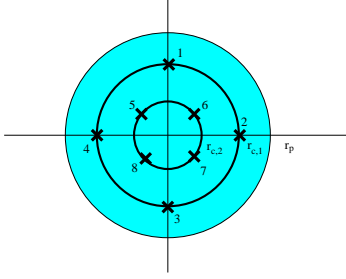

Fig. 2. Network with primary transmitters (not shown) uniformly distributed on the blue disc of radius r_p. There are $N = 8$ cognitive receivers in the network equally-divided across two circles ($q = 2$) of radii $r_{c,1}$ and $r_{c,2}$ respectively. This gives rise to partitions $\mathcal{P}_1 = \{1, 2, 3, 4\}$ and $\mathcal{P}_2 = \{5, 6, 7, 8\}$. The cognitive receivers are equally-spaced on each circle as shown.

primary transmitter i and cognitive receiver j is given by $d_{ij} = \sqrt{(X_{p,i} - X_{c,i})^2 + (Y_{p,i} - Y_{c,i})^2}$. We assume that the channel attenuation between any two nodes in the network is solely determined by a path-loss gain[4]

$$ h_{ij} = \frac{\delta^2}{d_{ij}^2} \mathbb{I}(d_{ij} \geq \delta) + \mathbb{I}(d_{ij} < \delta), \quad (1) $$

where $\delta > 0$ is commonly referred to as the critical distance and $\mathbb{I}(\cdot)$ is the indicator function. Here, we assume a path-loss coefficient of two but this can be generalized to any larger constant quite easily. This is an approximation of the free space path loss model [19] since the latter precludes the placement of nodes within a distance of δ metres. On the other hand, our model in (1) states that within a distance of δ metres, there is no channel attenuation. This assumption affords us analytical tractability while compromising very little on modelling accuracy.

The cognitive base station adopts transmit strategies that respect an upper bound on interference caused to the primary network. We assume that the primary transmitters choose their transmit strategy based on this interference cap (fixed

---

3We note that this would roughly be the case when N becomes large and the users are uniformly distributed.

4This model is accurate for sufficiently large scheduling time scales and/or if all transmitters, receivers and surrounding objects remain stationary for long periods of time.

for all time) thereby not adapting to changes in the transmission policy of the cognitive network. At time t, the primary users transmit collectively at power $\mathbf{p}(t)$, i.e., user i transmits at power $p_i(t)$.

In order to make scheduling decisions, the cognitive base station requires knowledge of the channel quality between itself and each cognitive receiver. The channel quality to a cognitive receiver depends critically on the interference power that this node receives from the primary network. Thus, it is important to estimate (perfectly if possible) the interference power received from the primary network. Let the interference power received by cognitive receiver i at time t be denoted by $p_{int,i}(t)$.

Under the above assumptions, the interference power received by the cognitive receiver i at time t is given by $y_i(t) = \sum_{j=1}^{N} h_{ij} p_j(t)$. Collecting these interference powers across all cognitive receivers, we get the interference state

$$ \mathbf{y}(t) = \mathbf{H}\mathbf{p}(t) \text{ where } \mathbf{y}(t) = [y_1(t) \ y_2(t) \ldots y_N(t)]^T. $$

The rows of $\mathbf{H}$ are ordered according to the partition ordering $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_q\}$. For example, for the system in Fig.2, the rows of $\mathbf{H}$ would be ordered $\{1, 2, 3, 4, 5, 6, 7, 8\}$. The cognitive base station would like to acquire full knowledge of interference state $\mathbf{y}(t)$ in order to perform user scheduling. At the beginning of scheduling instant t, following limited feedback literature (see the survey by Love et al. [15] and references therein), we assume that the cognitive base station has the ability to request observations $y_i(t)$ (interference powers) from a subset I of cognitive receivers. This is done through orthogonal feedback channels from each user, which essentially means that the observations, when transmitted from the users, do not interfere with each other[5]. Of course, the cognitive base station can query all N cognitive receivers, i.e. $I = \{1, 2, \ldots, N\}$ and acquire the interference state but is this the most bandwidth efficient approach? It turns out that this is not the most efficient approach since it does not exploit the underlying structure of the primary network as explained shortly.

A. Sparsity

Since the very motivation for cognitive radio networks is that the primary spectrum is under-utilized, this automatically induces sparsity into the power profile vector $\mathbf{p}(t)$. Moreover, for networks where the dynamics (primary user activity, channel coherence times, cognitive user activity) operate on different time-scales, which is the case with many wireless systems, it is natural for the power difference vector $\mathbf{s}(t) = \mathbf{p}(t) - \mathbf{p}(t-1)$ to be more sparse than the absolute power profile vectors $\mathbf{p}(t)$ and $\mathbf{p}(t-1)$. This is true in particular when the dynamics of the primary user activity operates on a slower time-scale than that of the cognitive network. Consider for instance a scenario where a small subset of primary receivers are mobile. This would require power control on the part of a correspondingly small fraction of

---

5These channels are assumed to be orthogonal to the primary users as well.

primary transmitters in order to maintain the same rate. Another scenario is where only a small subset of primary users have data to transmit at each instant. Under these settings, the vector $\mathbf{s}(t) = \mathbf{p}(t) - \mathbf{p}(t-1)$ is sparse, say S-sparse (it contains at most S non-zero entries) and we can define

$$\mathbf{z}(t) = \mathbf{y}(t) - \mathbf{y}(t-1) = \mathbf{H}\mathbf{s}(t)$$

to be the difference in interference powers between adjacent scheduling time slots. Here, we assume that the channel remains constant for a long period of time, which accurately models a system with slow-moving or stationary users. For this reason, we also assume that the cognitive base station has full knowledge of all (both primary and cognitive) user positions in the network. In the longer version of this paper [31], we will discuss implementation methods that can help alleviate this requirement.

Under such a setting where we have naturally-occurring sparsity in $\mathbf{z}(t)$ and/or $\mathbf{y}(t)$, it is possible to significantly reduce the required feedback bandwidth. Let $\mathbf{y}_I(t)$ denote the observations acquired from the queried users at time t and let $k = |I|$ represent the size of this set. It is of immediate interest to determine the smallest query size k (or feedback bandwidth) that the cognitive base station requires in order to recover $\mathbf{y}_{I^c}(t)$ reliably. We will show that it is possible to recover $\mathbf{s}(t)$ (and hence $\mathbf{y}_{I^c}(t)$) through the technique of *compressed sensing* using only $k = \mathcal{O}(S\log N)$ observations thus resulting in a significant reduction of feedback bandwidth. In this work, we do not consider the number of bits required to communicate $\mathbf{y}_I(t)$ reliably as we are interested primarily in the scaling behaviour of feedback bandwidth.

Compressed sensing The topic of compressed sensing has received tremendous interest in the recent years [16]–[18]. The theory essentially states that one can recover sparse data ($\mathbf{s}(t)$ in our case) exactly, given an underdetermined system of $\mathcal{O}(S\log N)$ equations. The results are attractive from an algorithm perspective as well since it has been shown that recovery is possible using standard computationally-efficient convex optimization techniques such as ℓ_1-norm minimization[6], often called *Basis Pursuit* [23], given by

$$\begin{array}{ll} \text{minimize} & ||\mathbf{x}||_1 \\ \text{subject to} & \mathbf{H}_{I,r}\mathbf{x} = \mathbf{z}_I(t) \end{array}. \qquad (2)$$

Here, $\mathbf{z}_I(t) = \mathbf{y}_I(t) - \mathbf{y}_I(t-1)$ and $\mathbf{H}_{I,r}$ is the sub-matrix of path-loss entries (as defined in (1)) corresponding to the rows in $I \subseteq \{1, 2, \ldots, k\}$. Compressed sensing theory states that it is possible to recover any S-sparse vector if and only if the *sensing matrix* $\mathbf{H}$ satisfies the *Null Space Property* (NSP) [25] of order S. This property will be defined in the next section. Furthermore, the choice of subset I is not important (only the size) for this special class of matrices. In the following section, we will show that path-loss matrices as defined in (1) do indeed satisfy the NSP and hence facilitate compressed sensing. In our setting, this means that we can

[6]ℓ_1-norm minimization can be written as a linear program.

exploit the sparsity structure induced by our wireless application and use compressed sensing techniques to conserve feedback bandwidth. We note that in our application, the sensing matrix is provided by the channel as opposed to traditional compressed sensing where the designer is allowed to choose a convenient sensing mechanism.

Before we conclude this section, we note that a host of other algorithms such as the greedy correlation approach [20], and the least absolute shrinkage and selection operator (LASSO) [22] have all been shown to work well as recovery tools. The LASSO, given by

$$\text{minimize} \quad \tfrac{1}{2}||\mathbf{H}_{I,r}\mathbf{x} - \mathbf{z}_I(t)||^2 + \lambda||\mathbf{x}||_1, \qquad (3)$$

is method of recovery that we will use for the purposes of our simulations in Section IV. The theoretical connections between LASSO and Basis Pursuit have been well-analysed by authors such as Tropp [21].

III. NSP OF PATH-LOSS MATRICES

In this section, we establish that path-loss matrices $\mathbf{H}$ satisfy the NSP (which will be defined shortly) when the feedback bandwidth obeys $k = \mathcal{O}(S\log N)$. Theorem 4 constitutes the main result in this section.

A. Preliminaries

We define the null space property from Gribonval et al. [26]. Given a matrix $\mathbf{M}$, let $\mathcal{N}(\mathbf{M})$ denote its null space. *Definition (Null space Property)*: A matrix $\mathbf{M}$ satisfies the null space property of order S if for all subsets $\mathcal{S} \subseteq \{1, 2, \ldots, N\}$ with $|\mathcal{S}| \leq S$, the following holds

$$||\mathbf{v}_{\mathcal{S}}||_1 \leq ||\mathbf{v}_{\mathcal{S}^c}||_1, \ \forall \mathbf{v} \in \mathcal{N}(\mathbf{M}) \setminus \mathbf{0}.$$

where $\mathcal{S}^c = \{1, 2, \ldots, N\} \setminus \mathcal{S}$. Based on this property, the following recovery result [26] has appeared both implicitly and explicitly in works such as [25], [27]. Let the support set of $\mathbf{s}(t)$ be denoted by $\mathcal{S}$ with $|\mathcal{S}| \leq S$. A vector $\mathbf{s}(t)$ is S-sparse if $\mathcal{S}| \leq S$.

Theorem 1. *Let $\mathbf{M} \in \mathbb{R}^{k \times N}$. Every S-sparse vector $\mathbf{s} \in \mathbb{R}^N$ is the solution to the ℓ_1-norm minimization problem in (2) with $\mathbf{y} = \mathbf{M}\mathbf{s}$ iff $\mathbf{M}$ satisfies the NSP of order S.*

$\square$

The NSP is typically quite difficult to prove directly leading to the development of sufficient conditions that are easier to establish. One such sufficient condition is the *restricted isometry property* [28] that has become quite popular in recent years and is defined below.

Definition (Restricted Isometry Property): A matrix $\mathbf{M}$ satisfies the Restricted Isometry Property (RIP) of order p if there exists $\epsilon_p(\mathbf{M}) \in (0, 1)$ such that

$$(1 - \epsilon_p(\mathbf{M}))||\mathbf{v}_{\mathcal{T}}||_2^2 \leq ||\mathbf{M}_{\mathcal{T},c}\mathbf{v}_{\mathcal{T}}||_2^2 \leq (1 + \epsilon_p(\mathbf{M}))||\mathbf{v}_{\mathcal{T}}||_2^2 \qquad (4)$$

holds for all sets $\mathcal{T}$ with $|\mathcal{T}| \leq p$.

Here, $\epsilon_p(\mathbf{M})$ is called the *restricted isometric constant* of $\mathbf{M}$. The RIP essentially requires that all $k \times |\mathcal{T}|$ sub-matrices of $\mathbf{M}$ be well-conditioned. The following result that connects the RIP and NSP.

Theorem 2. *Suppose the restricted isometric constant $\epsilon_{2p}(\mathbf{M})$ satisfies*

$$\epsilon_{2p} < \frac{1}{3},$$

then $\mathbf{M}$ satisfies the NSP of order p.

□

The approach we use to prove "goodness" of path-loss matrices $\mathbf{H}$ is motivated by the following observation. In general, the null space of a product of two matrices $\mathbf{NM}$ contains the null space of $\mathbf{M}$ and therefore if $\mathbf{NM}$ satisfies the NSP, so does $\mathbf{M}$. This allows us to study the class of *effective* path-loss matrices $\mathbf{A} = \mathbf{BG} = \mathbf{BWH}$ where

$$
\begin{aligned}
\mathbf{W} &= [\mathbf{W}_1^T\ \mathbf{W}_2^T \ldots \mathbf{W}_q^T]^T \\
\mathbf{W}_i^T &= [\mathbf{w}_{i|0}\ \mathbf{w}_{i|1} \ldots \mathbf{w}_{i|\frac{N}{q}-2}] \\
\mathbf{w}_i^T &= \begin{bmatrix} \underbrace{0\ 0\ldots 0}_{(i-1)\frac{N}{q}}\ 1\ -1\ 0\ldots 0 \end{bmatrix}
\end{aligned}
\tag{5}
$$

and $\mathbf{w}_{i|l}$ denotes the vector $\mathbf{w}_i$ cyclically shifted to the right by l positions; $\mathbf{B}$ is defined in (6) with $\beta_i \sim$ Bernoulli $\left(\frac{1}{2}\right)$, $\forall i$ and independent across i. $\mathbf{D} = \text{diag}\{\mathbf{d}\}$ is a diagonal matrix with $\mathbf{d}$ on its diagonal. The Bernoulli random variables have support $\{\pm 1\}$. We focus our attention on establishing the recovery properties of $\mathbf{A}$ rather than $\mathbf{H}$. We will show that $\mathbf{A}$ satisfies the RIP with $k = \mathcal{O}(S\log N)$ observations and then invoke the result in Theorem 2 above. The transformation $\mathbf{W}$ essentially subtracts adjacent rows of $\mathbf{H}$ corresponding to users that belong to the *same* partition. Thus, the dimension of $\mathbf{G}$ is $(k-q) \times k$. The transformation $\mathbf{B}$ weights and adds adjacent rows of $\mathbf{G}$.

According to our spatial distribution model, when conditioned on the positions of the cognitive users, the columns of $\mathbf{H}$ become stochastically independent since each primary transmitter is independently thrown. We will rely heavily on recent results from Vershyin [29] and Adamcyzk et al. [30] that deal with sensing matrices containing independent columns. Before we reproduce the RIP result [29], [30] for matrices with independent columns, we present a primer on sub-gaussian and sub-exponential random variables along with some useful results from non-asymptotic matrix theory.

B. Useful concentration inequalities

We refer the reader to the tutorial paper by Vershynin [29] for a great introduction to non-asymptotic matrix theory. Lemmas 1-6 below are well-known past results that are summarized in this paper [29]. The proofs are not reproduced due to lack of space.

Lemma 1. *Let z be random variable. The following properties are equivalent with parameters $K_i > 0$ differing from each other by at most an absolute constant factor.*
(i) Tails: $Pr(|z| > t) \leq exp(1 - \frac{t^2}{K_2})$ for all $t > 0$,
(ii) Moments: $(\mathbb{E}[|z|^p])^{\frac{1}{p}} \leq K_2\sqrt{p}$ for all $p \geq 1$,
(iii) Super-exponential moment: $\mathbb{E}\left[exp\left(\frac{z^2}{K_3}\right)\right] \leq e$.
Moreover, if $\mathbb{E}[z] = 0$ then properties (i)-(iii) are also

equivalent to the following one:
(iv) Moment generating function: $\mathbb{E}[exp(tz)] \leq exp(t^2 K_4)$ for all $t \in \mathbb{R}$.

□

A random variable that satisfies the above property is called a *sub-gaussian* random variable. Such random variables are often characterized by the ψ_2-norm[7], which is defined as

$$||z||_{\psi_2} = \sup_{p\geq 1} \frac{(\mathbb{E}[|z|^p])^{\frac{1}{p}}}{\sqrt{p}}. \tag{6}$$

It follows that if the ψ_2-norm of z is finite, then z is a sub-gaussian random variable with $||z||_{\psi_2} = K_2$. This is the case for bounded random variables with symmetric distributions.

Lemma 2. *Let z be a symmetrically distributed, bounded random variable with $|z| \leq M$, $M > 0$. Then, $\mathbf{z}$ is a sub-gaussian random variable with $||z||_{\psi_2} \leq cM^2$, $c > 0$.*

□

In higher dimensions, a random vector $\mathbf{z}$ of dimension N is called sub-gaussian if $\mathbf{z}^T\mathbf{x}$ is sub-gaussian for every $\mathbf{x} \in \mathbb{R}^N$.

Lemma 3. *Let $\{z_i\}_{i=1}^M$ be a collection of independent, zero-mean, sub-gaussian random variables. Then, $\mathbf{z}$ is a sub-gaussian random vector with $||\mathbf{z}||_{\psi_2} = C max_i\ ||z_i||_{\psi_2}$.*

□

One can make a similar characterization of sub-exponential random variables through the following lemma.

Lemma 4. *Let z be random variable. The following properties are equivalent with parameters $K_i > 0$ differing from each other by at most an absolute constant factor.*
(i) Tails: $Pr(|z| > t) \leq exp(1 - \frac{t}{K_2})$ for all $t > 0$,
(ii) Moments: $(\mathbb{E}[|z|^p])^{\frac{1}{p}} \leq K_2 p$ for all $p \geq 1$,
(iii) Super-exponential moment: $\mathbb{E}\left[exp\left(\frac{z}{K_3}\right)\right] \leq e$.

□

A random variable that satisfies the above property is called a *sub-exponential* random variable. The ψ_1-norm of z is defined as

$$||z||_{\psi_1} = \sup_{p\geq 1} \frac{(\mathbb{E}[|z|^p])^{\frac{1}{p}}}{p}. \tag{7}$$

This immediately brings us to the next two lemmas which explore the connection between sub-gaussian and sub-exponential random variables.

Lemma 5. *z is a sub-gaussian random variable if and only if z^2 is a sub-exponential random variable. Furthermore, we have that $||z||_{\psi_2}^2 \leq ||z^2||_{\psi_1} \leq 2||z||_{\psi_2}^2$.*

□

The following lemma contains a large-deviations result for a weighted sum of sub-exponential random variables.

---

[7]Alternate definitions of this norm have been adopted (such as in [30]) that are all equivalent to within a constant factor.

$$\mathbf{B} = \operatorname{diag}\left\{ \frac{1}{\sqrt{\operatorname{Var}\{g_{11}\} + \operatorname{Var}\{g_{21}\}}} \cdots \frac{1}{\sqrt{\operatorname{Var}\{g_{(k-q)1}\} + \operatorname{Var}\{g_{11}\}}} \right\} \begin{bmatrix} \beta_1 & \beta_2 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ 0 & \beta_3 & \beta_4 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & \beta_{2(k-q)-3} & \beta_{2(k-q-1)} \\ \beta_{2(k-q)} & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & \beta_{2(k-q)-1} \end{bmatrix} \tag{6}$$

Lemma 6. *Let $\{z_1, z_2, \ldots, z_M\}$ be a collection of independent, zero-mean, sub-exponential random variables and let $\psi_{max,z} = max_i \, ||z_i||_{\psi_1}$. Then,*

$$Pr\left(\left| \sum_{i=1}^{M} z_i \right| > Mt \right) \leq exp\left(-cM min \left\{ \frac{t^2}{\psi_{max,z}^2}, \frac{t}{\psi_{max,z}} \right\} \right).$$

$\square$

We are now ready to prove a concentration bound on the event $\left\{ k(1-\epsilon)||\mathbf{x}||_2^2 \leq ||\mathbf{Ax}||^2 \leq k(1+\epsilon)||\mathbf{x}||_2^2 \right\}$ in order to establish the RIP (hence NSP) for matrix $\mathbf{A}$. Before we move on to this task, we require one more definition. A random vector $\mathbf{m}$ of dimension M is called *isotropic* if $\mathbb{E}[|\mathbf{m}^T\mathbf{x}|^2] = ||\mathbf{x}||^2$ for all $\mathbf{x} \in \mathbb{R}^M$.

C. NSP of effective path-loss matrices $\mathbf{A}$

We reproduce the recent RIP (hence NSP) result [29], [30] concerning matrices with independent columns. We refer the reader to [29], [30] for the proof.

Theorem 3. *Let $\mathbf{M}$ be an $k \times N$ random matrix whose columns are independent, isotropic and sub-gaussian with $\psi_{max,m} = max_i \, ||\mathbf{m}_i||_{\psi_2}$. Furthermore, let the columns satisfy $||\mathbf{m}_i||^2 = k$ almost surely. Then, the normalized matrix $\frac{1}{\sqrt{k}}\mathbf{M}$ is such that if $k \geq C_{\psi_{max,m}} \varepsilon^{-2} S log\left(\frac{eN}{S} \right)$, then*

$$\varepsilon_p\left(\frac{1}{\sqrt{k}}\mathbf{M} \right) \leq \varepsilon \tag{8}$$

with probability at least $1 - 2exp(-c_{\psi_{max,m}}\varepsilon^2 k)$. Here, $c_{\psi_{max,m}}$ and $C_{\psi_{max,m}}$ depend only the worst-case sub-gaussian norm $\psi_{max,m}$.

$\square$

As mentioned earlier, the channel matrix $\mathbf{H}$ contains independent columns since the positions of the cognitive users are fixed. However, each column contains entries that are not centered, not isotropic and that are highly coupled. This is because all entries in $\mathbf{h}_i$ are completely determined by the position of the primary transmitter i. Hence, it is not immediately clear whether the column is sub-gaussian due to this strong coupling.

To prove the NSP of $\mathbf{H}$, our approach will be to suitably left-multiply the channel matrix $\mathbf{H}$ by carefully-chosen matrices so as to meet the sufficient conditions in Theorem 3. The following theorem constitutes the main result of this paper. The proof is involved and hence deferred to a longer version of this paper [31] due to lack of space. Instead, we provide a sketch of the proof.

Theorem 4. *The effective path-loss matrix $\mathbf{A} = \mathbf{BWH}$ of size $(k-q) \times N$ contains independent, isotropic, centered, sub-gaussian columns. Furthermore, we have that $||\mathbf{a}_i||^2 = 1$ almost surely.*

Proof sketch: The channel matrix $\mathbf{H}$ contains elements that are identically distributed by the symmetry of primary transmitters' placement on a disc and the circular placement of the cognitive receivers. Thus, the matrix $\mathbf{G}$ obtained after left multiplication by $\mathbf{W}$ contains zero-mean entries while preserving column independence. A fact that is less obvious is that the entries of $\mathbf{G}$ are symmetric. This follows from the definition of $\mathbf{G}$ in (5), where we have dropped *difference* rows corresponding to users from different partitions, and from the notion of exchangeable random variables. Recall that $\mathbf{G}$ is of size $(k-q) \times N$.

Next, we multiply $\mathbf{G}$ by the random matrix $\mathbf{B}$. Since $\mathbf{G}$ contains entries with symmetric distributions, we will show that the multiplication by $\mathbf{B}$ does not destroy the stochastic independence across the columns of $\mathbf{G}$. In addition, the columns of $\mathbf{A}$ are identically distributed allowing us to focus on the properties of the first column $\mathbf{a}_1$. $\mathbf{a}_1$ is isotropic since the presence of the i.i.d. Bernoulli random variables β_i successfully nulls the pairwise cross-correlation amongst the elements of $\mathbf{a}_1$. The constants in the diagonal matrix

$$\operatorname{diag}\left\{ \frac{1}{\sqrt{\operatorname{Var}\{g_{11}\} + \operatorname{Var}\{g_{21}\}}} \cdots \frac{1}{\sqrt{\operatorname{Var}\{g_{(k-q)1}\} + \operatorname{Var}\{g_{11}\}}} \right\} \text{ are}$$

chosen to make the variance of each a_{i1}, $i = 1, \ldots, (k-d)$, unity.

The column $\mathbf{a}_1$ can be shown to be sub-gaussian as follows. We observe that when conditioned on the location of primary transmitter 1, the only randomness in $\mathbf{a}_1$ resides in the Bernoulli random variables β_i. Thus, when conditioned on the location of primary transmitter 1, $\mathbf{a}_1$ contains symmetrically distributed, bounded, independent random variables bringing into effect the result in Lemma 3.

Similarly, norm-concentration is proved by alluding to the same observation. Again, when conditioned on the location of primary transmitter 1, from Lemma 5, $\{a_{i1}^2\}_{i=1}^{k-q}$ is a collection of independent sub-exponential random variables bringing into effect the result in Lemma 6. $\square$

From Theorems 3 and 4, $\mathbf{A}$ satisfies the RIP and in turn, the NSP for a sufficiently small isometric constant when $(k-q) = \mathcal{O}(S log N)$. This implies that $\mathbf{H}$ satisfies the NSP when $k = \mathcal{O}(S log N)$. Note that the number of circles containing the cognitive receivers appears as an additive constant and hence does not affect the order of the result.

We conclude this section with a comment on some implementation aspects of the proposed limited feedback protocol. We observe almost immediately that the proposed protocol admits a distributed implementation. This would involve each cognitive receiver flipping a coin with a certain probability which guarantees that the number of observations received by the base station is greater than the lower bound in Theorem 4

with high probability. However, such a protocol would work only if we are operating in the setting where the absolute power profile vectors are sparse. This is because, to operate in the sparse dynamics mode, we would require the query set I to remain constant over all time, which would not be guaranteed under a distributed coin-flipping protocol.

IV. STUDY OF ROBUSTNESS

Thus far in this paper, we have studied sparse recovery in the context of path-loss sensing matrices assuming that the cognitive base station has perfect knowledge of the sensing matrix $\mathbf{H}$. In reality though, this is seldom true since the system is subject to errors that could stem from the channel estimation process, quantization, the fact that node mobility that is faster than the tracking ability of channel estimation algorithm, shadowing and other small-scale fading effects. This motivates us to analyse numerically the performance of sparse recovery under imperfect channel knowledge, which forms the subject of this section.

We consider two perturbation models in our simulations: *(i)* (E1) Small error introduced into all entries of the sensing matrix *(ii)* (E2) Large error introduced into a small fraction of entries, say $f \in [0, 1]$, chosen uniformly at random. Under both models E1 and E2, the error is multiplicative and random, i.e. the perturbed matrix entry is $\hat{h}_{ij} = (1 + e_{ij})h_{ij}$ where $e_{ij} \sim U[-p, p]$ for some $p \in [0, 1]$. This multiplicative per-entry perturbation allows us to introduce error in a systematic fashion since the parameter p now represents the maximum percentage error per entry. Under both perturbation models E1 and E2, the performance metric is average distortion per channel measurement. More formally, let $\hat{\mathbf{s}}$ represent the output of ℓ_1 minimization under channel imperfections and $\mathbf{s}$ denote the true vector. Then, average distortion per channel measurement when fkN entries, chosen uniformly at random, are corrupted by multiplicative noise with a maximum magnitude of $100p\%$ while consuming a feedback bandwidth of k observations is given by

$$D_k(p, f) = \frac{1}{N}\mathbb{E}_{\mathbf{E}(p), \mathbf{H}, \mathbf{L}(f)}\left[||\mathbf{y} - \hat{\mathbf{y}}||_1\right] \qquad (9)$$

where $\hat{\mathbf{y}} = \hat{\mathbf{H}}\hat{\mathbf{s}}$, $\mathbf{y} = \mathbf{H}\mathbf{s}$ and $\mathbf{L}(f) \in \{0, 1\}^{k \times N}$ is a random matrix that models the locations of the error. This matrix contains fkN ones chosen uniformly at random. The remaining $kN - fkN$ entries are zero.

Under the first error model E1, we conduct experiments in order to measure $D_{\theta S\log N}(p, 1)$ for $p \in [0, 1]$ and $\theta \in \{1, 1.25, 1.5, 1.75, 2\}$. The results are presented in Fig. 3(a). Under the second error model E2, we measure $D_{\theta S\log N}(1, f)$ for $f \in [0, 1]$ and $\theta \in \{1, 1.25, 1.5, 1.75, 2\}$ and present the results in Fig. 3(b). The recovery algorithm of choice is the LASSO as defined in (3). Primary user activity is modelled as ON-OFF for the sake of simplicity. This means that if primary user i is ON, then $p_i(t) = 10$; else $p_i(t) = 0$. This models scenarios where the primary users' packet arrivals are bursty thus requiring intermittent transmissions followed by periods of inactivity. Note that in this scenario, the power vector $\mathbf{p}(t)$ at each instant is

sparse (which of course automatically implies that $\mathbf{s}(t)$, the difference in powers, is sparse). Hence, we can focus on recovering $\mathbf{p}(t)$ instead of $\mathbf{s}(t)$. We further assume that the cognitive base station knows *a priori* the value of the primary transmit power thereby requiring only support recovery. The simulation parameters are provided below in Table I.

TABLE I
SIMULATION PARAMETERS

| Description of parameter	Associated variable	Simulation Value
Cell size	d	2000 metres
Critical distance	δ	100 metres
No. of users (cognitive and primary)	N	100
Sparsity	S	5
LASSO penalty parameter	λ	0.0005
Primary transmit power vector	$\mathbf{p}(t)$	$\left[\underbrace{10 \ldots 10}_{5} \underbrace{0\ 0 \ldots 0}_{95}\right]^{T}$ Watts
Iterations to estimate distortion		500

We make a few remarks on the simulation results. Note that each figure contains four curves corresponding to $k \in \{24, 29, 35, 41, 47\}$.

Remarks:
(i) Under both error models, we see from Fig. 3 that the system is significantly robust to imperfect channel knowledge given the range of the distortion values.
(ii) Under error model E1, when all entries are corrupted by factor of 15%, we see from Fig. 3(a) that a 25% increase in the number of observations or feedback bandwidth results in a 30% decrease in distortion. As we move from $k = S\log N$ to $k = 1.25S\log N$ however, the reduction in distortion is now less than 30%. This leads to believe that the distortion exhibits *diminishing returns* in the feedback bandwidth, which is a reasonable in the context of wireless systems.
(iii) Under error model E1, if we keep the distortion fixed at 0.6, we see from Fig. 3(a) that a 60% increase in feedback bandwidth from $k = 1.25S\log N$ to $k = 2S\log N$ results in a 230% increase in error tolerance.
(iv) Under error model E2, when 10% of the entries are corrupted by an error of magnitude at most 100%, we see that a 25% increase in the number of observations or feedback bandwidth results in a 27% decrease in distortion. As in the earlier case, we also see evidence of diminishing returns.
(v) Under error model E2, if we keep the distortion fixed at 0.6, we see from Fig. 3(a) that a 30% increase in feedback bandwidth from $k = 1.5S\log N$ to $k = 2S\log N$ results in a 150% increase in error tolerance.

REFERENCES

[1] National Telecommunications and Information Administration (NTIA), "FCC frequency allocation chart", 2003. Download
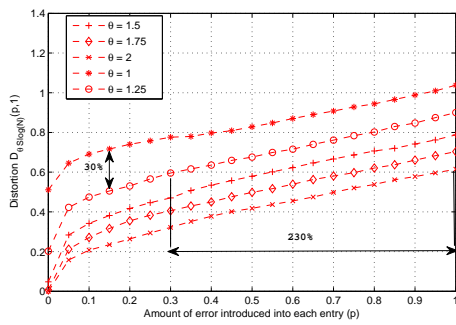
(a) Robustness study under error model E1.
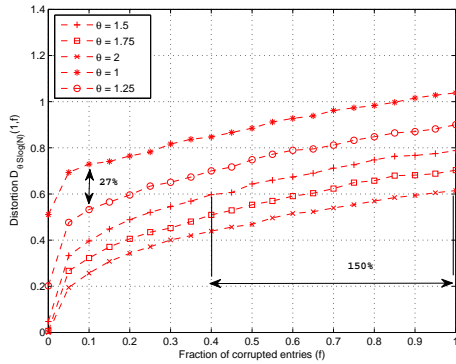
(b) Robustness study under error model E2.

Fig. 3. Study of the robustness properties of LASSO under error models E1 and E2 with $N = 100$ users, Sparsity $S = 5$ and $k = \lceil \theta S \log N \rceil$ with $\theta \in \{1, 1.25, 1.5, 1.75, 2\}$.

available at www.ntia.doc.gov/osmhome/allochrt.pdf.

[2] J. Mitola and G. Q. Maguire, "Cognitive radios: making software radios more personal", *IEEE Personal Communications.*, vol. 6, pp. 1318, Aug. 1999.

[3] Federal Communications Commission Spectrum Policy Task Force, "Report of the spectrum efficiency working group", Technical Report 02-135, Nov. 2002.

[4] P. J. Kolodzy, "Cognitive radio fundamentals", *SDR Forum*, Singapore, Apr. 2005.

[5] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective", *Proc. of the IEEE*, vol. 97, pp. 894-914, May 2009.

[6] A. Attar, M. Reza Nakhai and A. H. Aghvami, "Cognitive radio game: a framework for efficiency, fairness and QoS guarantee", *Proc. of IEEE ICC 2008*, pp. 4170 - 4174, 2008.

[7] N. Gatsis, A. G. Marques and G. B. Giannakis, "Utility based power control for peer-to peer cognitive radio networks with heterogeneous constraints", *IEEE ICASSP 2008*.

[8] L. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks", *IEEE Trans. Wireless Commun.*, vol. 7, Dec. 2008.

[9] A. Gopalan, C. Caramanis and S. Shakkotai, "On wireless scheduling with partial channel-state information", *Proc. of the Allerton Conf. on Communication, Control, and Computing*, Monticello, IN, Oct. 2007.

[10] M. Ouyang and L. Ying, "On scheduling in multi-channel wireless downlink networks with limited feedback", *Proc. of the Allerton Conference on Communication, Control, and Computing*, Monticello, IN, Oct. 2009.

[11] N. Kalouptsidis, G. Mileounis, B. Babadi, V. Tarokh, "Adaptive algorithms for sparse nonlinear channel estimation", *Proc. IEEE Workshop on Statistical Signal Processing*, Cardiff, Wales, Sep. 2009.

[12] S.F. Cotter and B.D. Rao, "Sparse channel estimation via matching pursuit with application to equalization", *IEEE Trans. Commun.*, vol. 50, Mar. 2002.

[13] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels", *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sep. 2008.

[14] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications", *IEEE Comm. Surveys And Tutorials*, vol. 11, First Quarter 2009.

[15] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao and M. Andrews, "An overview of limited feedback in wireless communication systems", *IEEE Journ. Sel. Commun.*, vol. 26, pp. 1341-1365, Oct. 2008

[16] E. Candes and M. Wakin, "An introduction to compressive sampling", *IEEE Sig. Proc. Magazine*, vol. 25, pp. 21 - 30, Mar. 2008.

[17] R. Baraniuk, "Compressive sensing", *IEEE Sig. Proc. Magazine*, vol. 24, pp. 118-121, July 2007.

[18] D. Donoho, "Compressed sensing", *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289-1306, Apr. 2006.

[19] Rappaport, S. T., "Wireless communications principles and practice", *Prentice-hall*, Chp. 4, pp. 138-139, 2002.

[20] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit", *IEEE Trans. Inform. Theory*, vol. 53, pp. 4655-4666, 2007.

[21] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals", *IEEE Trans. Info. Theory*, vol. 51, pp. 1030-1051, Mar. 2006.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. Roy. Statist. Soc. Ser. B*, pp. 267-288, 1996.

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

[24] H. Rauhut, "Compressive sensing and structured random matrices", *in Theoretical Foundations and Numerical Methods for Sparse Recovery, ser. Radon Series Comp. Appl. Math. deGruyter*, in preparation.

[25] A. Cohen, W. Dahmen and R. DeVore, "Compressed sensing and best k-term approximation", *J. Amer. Math. Soc.*, vol. 22, pp. 211-231, 2009.

[26] R. Grobinval and M. Nielson, "Sparse representations in unions of bases", *IEEE Trans. Info. Theory*, vol. 49, pp. 3320-3325, Dec. 2003.

[27] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition", *IEEE Trans. Inform. Theory*, vol. 47, pp. 28452862, 2001

[28] E. J. Candes, T. Tao, "Decoding by linear programming", *IEEE Trans. Inform. Theory*, vol. 51, pp. 4203-4215, Dec. 2005.

[29] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices", *In Compressed sensing: theory and applications*, Y. Eldar and G. Kutyniok, editors, Cambridge University Press, Submitted.

[30] S. Mendelson, A. Pajor and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators", arXiv:math/0506239v1 [math.FA], 2005.

[31] H. Ganapathy, C. Caramanis and L. Ying, "Exploiting sparse dynamics for controlling cognitive wireless networks", *Technical Report* available at www.ece.utexas.edu/~cmcaram, The University of Texas, Austin, Oct. 2010.