NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Hyder, Md Mashud; Mahata, Kaushik David "A sparse recovery method for initial uplink synchronization in OFDMA systems". Published in IEEE Transactions on Communications Vol. 64, Issue 1, p. 377-386 (2016)

Available from: <http://dx.doi.org/10.1109/TCOMM.2015.2497232>

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Accessed from: <http://hdl.handle.net/1959.13/1329244>

A sparse recovery method for initial uplink synchronization in OFDMA systems

Md Mashud Hyder and Kaushik Mahata

Abstract—Initial uplink synchronization is an integral part of the wireless communication systems. It enables the base station (BS) to detect new subscriber stations (SS) willing to commence communication. It also enables the BS to estimate the uplink channel parameters of these SSs. Accurate estimation of channel parameters is crucial, as they ensure the uplink signals from all the SSs arrive at the BS synchronously at similar power levels. However, this detection and estimation problem turns out to be very challenging when multiple users initiate the synchronization procedure at the same time. We address this issue by exploiting the underlying sparsity of the estimation problem. We propose a fast sparse signal recovery approach that shows a clear improvement in detection and estimation performance compared to other state of the art methods. The proposed method can be integrated into any OFDM based standard.

Index Terms—Initial ranging, random access, code detection, sparse representation, OFDMA.

I. INTRODUCTION

A. Background

The orthogonal frequency-division multiple access (OFDMA) has been adopted in most of the modern wireless communication standards [1], [2]. To avoid multiple-access interference (MAI), the OFDMA systems must ensure the uplink signals arriving at the BS are aligned with the local time and frequency references. This is accomplished by an initial uplink synchronization (IUS) procedure, called initial ranging (IR) in the IEEE 802.16 standards [1], and random access (RA) in the long-term evolution (LTE) [1]–[3]. The IUS starts with the allocation of a pre-defined set of subcarriers by the BS in some pre-specified time slots. This is called an “uplink synchronization opportunity”. The SSs willing to commence communication, referred to as the synchronizing terminals (STs), use this opportunity by modulating a randomly selected IUS code onto the allocated subcarriers. Due to different positions of STs within the radio coverage area, their signals arrive at the BS with different time delays. At the receiving side, the BS needs to detect the transmitted IUS-codes, and extract the timing and power information for each detected code.

The frequency domain correlation technique for IUS [4] estimates the time delays via corresponding phase shifts in frequency domain. The time domain correlation method [5] is an iterative process, where in each iteration the signal of the strongest active ST is removed from the received signal, and

the resulting residual signal is used in the succeeding iteration. This idea has been improved in [6] so that one additional IUS code is detected at each iteration. In [7] it is shown that the frequency-domain correlation based approach outperforms its time domain counterpart. Lee [8] replaces the IUS codes by a set of generalized chirp-like polyphase sequences to get a more accurate timing estimate. The IUS resource allocation scheme in [9] requires the STs to transmit their codes on disjoint sets of subcarriers, resulting in a minimum level of MAI. This scheme reduces the number of effective subcarriers for each user, resulting some degradation of timing estimation performance [5]. The IUS scheme in [10] needs that the uplink signals from STs are transmitted over disjoint subcarriers, and the receivers use filter banks to separate multiuser codes. The generalized likelihood ratio test (GLRT) based methods have been proposed in [3], [11], which are used in LTE and IEEE 802.16 based networks respectively, estimate the timing error and the channel response jointly via the maximum-likelihood (ML) method. GLRT performs very well with one or two STs. However, its performance appears to degrade quickly with an increase in the number of STs.

B. Contributions

Sparse signal representation [12], [13] has been proved useful in numerous applications [14]. Here we exploit sparsity to solve the IUS problem. The specific contributions are as follows:

Sparse signal model for IUS problem: We show that the sparsity arises in IUS from two directions: (i) the number of STs at a particular time under a BS is much smaller than the total number of available IUS codes; (ii) only a few channel taps have some noticeable magnitudes. We develop a signal model that allows us to exploit the above facts and pose the IUS problem as a sparse recovery problem. To the best of our knowledge, resolving IUS problem by exploiting the inherent sparse structure has not been addressed before.

A sparse recovery method for IUS problem: We found that the standard sparse recovery methods are too slow compared to the latency requirements in a modern communication system. Hence we propose a new algorithm to solve the sparse-IUS problem. Our approach combines two different types of sparse recovery paradigms. In particular, we apply a non-convex sparse recovery algorithm called ISL0 [15], which converges quickly when initialized sufficiently close to the final solution. We develop a new ℓ_1 method tailor-made to generate the above initial guess for ISL0 in a few iterations.

Authors are with the Department of Electrical Engineering, University of Newcastle, Callaghan, NSW-2308, Australia (e-mail: md.hyder@uon.edu.au; Kaushik.Mahata@newcastle.edu.au).

Research is supported by the Australian research council.

Acceleration via FFT: We show how the DFT based architecture of OFDM can be exploited to accomplish a significant acceleration of the sparse recovery algorithm by intelligent use of FFTs in different computation steps.

Statistically sound code detection method: Since the measured signals are noisy, the output of the sparse recovery algorithm is also noisy, and thus is not strictly sparse. We develop a systematic hypothesis test to detect the codes and the associated timing errors from the noisy output of the sparse recovery algorithm.

The simulation results demonstrate that the proposed algorithm outperforms other state of the art algorithms by a significant margin. This performance gain is more prominent when the number of STs increases.

The rest of the paper is organized as follows. The signal model and system description are provided in Section-II. In Section-III, we will develop a sparse signal recovery framework for the IUS problem and present a fast sparse recovery algorithm to resolve the sparse-IUS problem. Section-IV validates the performance of the proposed algorithm through numerical simulations. Conclusions are given in Section-V.

The following notation is used herein. The k -th component of a vector $\mathbf{x}$ is denoted by $[x]_k$. $(\mathbf{A})^*$ and $(\mathbf{A})^\top$ denote complex conjugate transpose and transpose of $\mathbf{A}$, respectively. $\lceil \cdot \rceil$ denotes the ceiling operator. A diagonal matrix using the components of the vector $\mathbf{x}$ is written as $\text{diag}(\mathbf{x})$. $\mathbf{0}_n$ and $\mathbf{1}_n$ are column vectors of length n with all 0's and 1's, respectively. The ℓ_p norm ($0 < p < \infty$) of a vector is defined as $\|\mathbf{v}\|_p = (\sum_t |\mathbf{v}_t|^p)^{1/p}$.

II. SIGNAL MODEL

A. Single ST

Consider an OFDMA system with N subcarriers. The system identifies M specific uplink subcarriers as 'IUS subchannels', and we denote their indices by $\{j_m : m = 1, 2, \dots, M\}$. An ST chooses a column $\mathbf{c}_\ell$ of a pre-specified $M \times G$ IUS code matrix

$$\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_G]$$

uniformly at random, and sends it via the IUS subchannels during an IUS opportunity. In the IUS opportunity, the ST transmits $N + 2N_1$ number of time domain channel symbols¹ which are constructed from $\mathbf{c}_\ell$. Here N_1 is the length of IUS specific cyclic prefix [3], [9] such that $N_1 > P + D$ where P is the duration of the maximum multipath delay spread (normalized by OFDM sampling period) i.e., the maximum channel length and D is the largest propagation delay for an ST located within the cell. Let $\mathbf{T}$ be a typical ST for which we denote these $N + 2N_1$ channel symbols by $\{u(k)\}_{k=-N_1}^{N+N_1-1}$. To get $\{u(k)\}_{k=-N_1}^{N+N_1-1}$ from the code $\mathbf{c}_\ell$, first calculate

$$s(q) = \frac{1}{\sqrt{N}} \sum_{m=1}^M c_{m,\ell} \exp\{i2\pi j_m q/N\}, \quad q = \mathcal{I}, \quad (1)$$

¹The concept of transmitting $N + 2N_1$ number of IUS channel symbols is adopted in the LTE system [3]. This scheme is also consistent with the generic structure described in [9]. The structure of IUS channel symbols in the WiMAX context [5], [11] is slightly different. However, the proposed approach can be adopted easily in the WiMAX context (see [16] for details).

where $\mathcal{I} := \{0, 1, 2, \dots, N-1\}$, and $c_{m,\ell}$ denotes the m th component of $\mathbf{c}_\ell$. The operation (1) can be seen as the process of modulating the j_m th subcarrier by $c_{m,\ell}$, and modulating the non-IUS subcarriers with 0. Subsequently, $\{u(k)\}_{k=-N_1}^{N+N_1-1}$ are constructed from $\{s(q)\}_{q=0}^{N-1}$ as

$$u(k) = \begin{cases} s(k \bmod N), & -N_1 \leq k \leq N-1, \\ 0, & N \leq k \leq N+N_1-1, \end{cases} \quad (2)$$

where as usual, $(k \bmod N) := k - N \cdot \lfloor k/N \rfloor$, with $\lfloor r \rfloor$ denoting the largest integer less than or equal to r . Note that for any integer k positive or negative, $(k \bmod N) \in \mathcal{I}$. Note that the last N_1 number of channel symbols in (2) are zero valued guard symbols [3], [9].

Suppose $h(p)$, $p \in \mathcal{I}$ are the channel impulse response coefficients between the transmitter T and BS. Let $\{v(k)\}_{k=-N_1}^{N+N_1-1}$ be the contribution of T in the symbols received by the base station during the IUS opportunity. At the receiving side, the base station discards the first N_1 channel symbols as it only contains incomplete part of the transmitted code [11]. In particular, BS uses $v(0), v(1), \dots, v(N-1)$ to extract the IUS information. These are delayed and convoluted version of the transmitted symbols:

$$v(k) = \sum_{p=0}^{N-1} h(p) u(k-p-d), \quad k \in \mathcal{I}. \quad (3)$$

Here d represents the propagation delay from T to BS. In particular, $d < D$. One purpose of the IUS process is to estimate d so that the transmitter can align its transmission with the frame boundaries of the BS. Now (2) and (3) imply

$$v(k) = \sum_{p=0}^{N-1} h(p) s\{(k-p-d) \bmod N\}, \quad k \in \mathcal{I}. \quad (4)$$

Thus $\{v(k)\}_{k=0}^{N-1}$ is the result of length N circular convolution of h , and s circularly shifted by d places. The BS calculates N point discrete Fourier transform (DFT) of v to obtain

$$V(n) = \sum_{k=0}^{N-1} v(k) \exp(-i2\pi kn/N), \quad n \in \mathcal{I}. \quad (5)$$

Likewise, let S and H denote the N point DFTs of s and h , respectively. It is well known [17] that in DFT domain the circular convolution in (4) is equivalent to

$$V(n) = H(n)S(n) \exp(-i2\pi dn/N). \quad (6)$$

After calculating V in (5), the base station forms the vector

$$\tilde{\mathbf{v}} = [V(j_1) \ V(j_2) \ \dots \ V(j_M)]^\top, \quad (7)$$

by taking the data received in the IUS subchannels with indices $\{j_m\}_{m=1}^M$. This vector is processed further to extract the IUS information.

Next we use (1) and (6) to express $\tilde{\mathbf{v}}$ in a convenient matrix form. For this we need the $N \times N$ DFT matrix $\mathbf{F}$ defined element-wise as

$$[\mathbf{F}]_{km} = \exp\{-i2\pi(k-1)(m-1)/N\}/\sqrt{N}.$$

Equation (1) is compactly given in matrix form as

$$[s(0) \ s(1) \ \dots \ s(N-1)]^\top = \mathbf{F}^* \mathbf{\Theta}^\top \mathbf{c}_\ell, \quad (8)$$

where Θ is an $M \times N$ row selector matrix such that the m th row of Θ is the j_m th row of the $N \times N$ identity matrix. Since S is the DFT of s , and $\mathbf{F}\mathbf{F}^* = \mathbf{I}$, we get from (8) that

$$\begin{bmatrix} S(0) & S(1) & \cdots & S(N-1) \end{bmatrix}^\top \\ = \mathbf{F} \begin{bmatrix} s(0) & s(1) & \cdots & s(N-1) \end{bmatrix}^\top = \Theta^\top \mathbf{c}_\ell. \quad (9)$$

Also from the theory of DFT, it is well known that $H(n)e^{-i2\pi dn/N}$ is the DFT of h circularly shifted by d places [17], which is written compactly as

$$\begin{bmatrix} H(0) & H(1)e^{-i2\pi d/N} & \cdots & H(N-1)e^{-i2\pi d(N-1)/N} \end{bmatrix}^\top \\ = \mathbf{F} \mathbf{h}_{\downarrow(d)}, \quad (10)$$

where $\mathbf{h}_{\downarrow(d)}$ is the circularly shifted version of the channel impulse response by d places expressed as a vector:

$$\mathbf{h}_{\downarrow(d)} := \begin{bmatrix} h(N-d) \cdots h(N-1) & h(0) & h(1) & \cdots & h(N-d-1) \end{bmatrix}^\top. \quad (11)$$

Also using the definition of Θ we write (7) as

$$\tilde{\mathbf{v}} = \Theta \begin{bmatrix} V(0) & V(1) & \cdots & V(N-1) \end{bmatrix}^\top.$$

Hence (6), (9) and (10) imply

$$\tilde{\mathbf{v}} = \mathbf{E}_\ell \mathbf{h}_{\downarrow(d)}, \quad \mathbf{E}_\ell = \Theta \text{diag}(\Theta^\top \mathbf{c}_\ell) \mathbf{F}. \quad (12)$$

Recall that P is assumed to be the maximum delay spread of a channel. Hence $h(k) = 0$ for $k \geq P$. Since $\mathbf{h}_{\downarrow(d)}$ is a circularly shifted version of channel impulse response by d places, all possible nonzero components of $\mathbf{h}_{\downarrow(d)}$ are concentrated in the rows with indices in the set $\{d+1, d+2, \dots, d+P\}$. At this point we recall that $d+P \leq D+P < N_1$. This also means that all rows of $\mathbf{h}_{\downarrow(d)}$ with indices are larger than N_1 are zeros. Hence it is fine to truncate $\mathbf{h}_{\downarrow(d)}$ to a N_1 dimensional vector, and thus it is enough to work with only first N_1 columns of $\mathbf{E}_\ell$.

B. Multiple STs

Let $\tilde{N}_\ell$ be the number of STs transmitting code $\mathbf{c}_\ell$. In practice, $\tilde{N}_\ell = 0$ for most values of ℓ , and the probability that $\tilde{N}_\ell > 1$ is very small. Let $\mathbf{h}_\ell^{(k)}$ and $d_{\ell,k}$, $k = 1, 2, \dots, \tilde{N}_\ell$ denote the channel impulse response vector and the delay of the k th ST transmitting the code $\mathbf{c}_\ell$. Let

$$\mathbf{h}_\ell = \begin{cases} \sum_{k=1}^{\tilde{N}_\ell} [\mathbf{h}_\ell^{(k)}]_{\downarrow(d_{\ell,k})}, & \tilde{N}_\ell > 0, \\ 0, & \tilde{N}_\ell = 0, \end{cases} \quad (13)$$

be the combined channel vector for all the STs transmitting the code $\mathbf{c}_\ell$. By the principle of superposition the single ST equation (12) is generalized for multiple STs. The data vector $\mathbf{y}$ received by the BS at the IUS subchannels is given by

$$\mathbf{y} = \sum_{\ell=1}^G \mathbf{E}_\ell \mathbf{h}_\ell + \mathbf{e}, \quad (14)$$

where $\mathbf{e}$ is complex Gaussian receiver noise with zero mean and covariance matrix $\sigma_e^2 \mathbf{I}$. The power received by the BS corresponding to the code $\mathbf{c}_\ell$ is given by [5]:

$$\Gamma_\ell = \mathbf{h}_\ell^* \mathbf{h}_\ell. \quad (15)$$

III. ESTIMATION OF IUS INFORMATION

A. Formal problem Statement

Given $\mathbf{y}$ the BS needs to i) find the set $\mathcal{L} = \{\ell : \Gamma_\ell \neq 0\}$; and ii) for every $\ell \in \mathcal{L}$ find Γ_ℓ and $d_{\ell,1}$ assuming $\tilde{N}_\ell = 1$.

Recall that the first d components of $\mathbf{h}_{\downarrow(d)}$ are zero, see (11). Hence by construction of $\mathbf{h}_\ell$ in (13), the index of the first nonzero component of $\mathbf{h}_\ell$ is $1 + \bar{d}_\ell$, where

$$\bar{d}_\ell = \min_{k \in \{1, 2, \dots, \tilde{N}_\ell\}} d_{\ell,k}, \quad \ell \in \mathcal{L}.$$

Clearly, if $\tilde{N}_\ell = 1$, then $\bar{d}_\ell = d_{\ell,1}$. For this reason we propose to estimate $\bar{d}_\ell$ as the timing error corresponding to each $\ell \in \mathcal{L}$. When $\tilde{N}_\ell = 1$ this estimate is consistent with our requirements. On the other hand, if $\tilde{N}_\ell > 1$, this estimate will have no practical relevance. This is because the transmissions made by the terminals involved will collide, refraining the BS from allocating any bandwidth to these stations [18], [19].

B. Sparse recovery framework

Since all rows of $\mathbf{h}_\ell$ with indices are larger than N_1 are zeros,

$$\mathbf{E}_\ell \mathbf{h}_\ell = \mathbf{E}_\ell(:, 1 : N_1) \mathbf{h}_\ell(1 : N_1). \quad (16)$$

Here $\mathbf{E}_\ell(:, 1 : N_1)$ denotes the submatrix of $\mathbf{E}_\ell$ formed by taking its first N_1 columns. Then (14) implies

$$\begin{aligned} \mathbf{y} &= \mathbf{A} \mathbf{x} + \mathbf{e}, \\ \mathbf{x} &:= [\mathbf{h}_1^\top(1 : N_1) \quad \mathbf{h}_2^\top(1 : N_1) \quad \cdots \quad \mathbf{h}_G^\top(1 : N_1)]^\top, \\ \mathbf{A} &= [\mathbf{E}_1(:, 1 : N_1) \quad \mathbf{E}_2(:, 1 : N_1) \quad \cdots \quad \mathbf{E}_G(:, 1 : N_1)]. \end{aligned} \quad (17)$$

Here $\mathbf{h}_\ell(1 : N_1)$ denotes the sub-vector of $\mathbf{h}_\ell$ consisting of its first N_1 components. By construction, $\mathbf{A} \in \mathbb{C}^{M \times G \cdot N_1}$ is a known matrix. On the other hand $\mathbf{x}$ and $\mathbf{e}$ are unknowns. Typically, the total number of STs $K = \sum_{\ell=1}^G \tilde{N}_\ell \ll G$, implying $\tilde{N}_\ell = 0$ (and therefore $\mathbf{h}_\ell = 0$) for a vast majority of the values $\ell \in \{1, 2, \dots, G\}^2$. This makes $\mathbf{x}$ very sparse. This motivates a sparse recovery framework for solving the IUS problem.

From a sparse $\tilde{\mathbf{x}}$ estimate of $\mathbf{x}$, BS can extract the IUS information as follows. Partition $\tilde{\mathbf{x}}$ into G sub-vectors:

$$\tilde{\mathbf{x}} = [\tilde{\mathbf{h}}_1^\top \quad \tilde{\mathbf{h}}_2^\top \quad \cdots \quad \tilde{\mathbf{h}}_G^\top]^\top,$$

where each $\tilde{\mathbf{h}}_\ell$ is of length N_1 . Then we declare $\ell \in \mathcal{L}$ only if $\|\tilde{\mathbf{h}}_\ell\| \neq 0$ and the index of the first nonzero component of $\tilde{\mathbf{h}}_\ell$ leads to an estimate of $\bar{d}_\ell$. In the following, we propose an algorithm to obtain a sparse estimate $\mathbf{x}$ from (17).

C. Background on sparse recovery methods

Since $\mathbf{A}$ is an $M \times G \cdot N_1$ matrix with $M < G \cdot N_1$, (17) is underdetermined, and has infinitely many solutions of $\mathbf{x}$, even when $\mathbf{e} = 0$. With $\mathbf{e} \neq 0$ we have more possible combinations of $\mathbf{x}$ and $\mathbf{e}$ satisfying (17). To solve $\mathbf{x}$ we

²To prevent a large number of STs from initiating IUS simultaneously, different approaches have been considered by the Third Generation Partnership Project [18]. For example, the collision resolution algorithms such as binary exponential back-off is used to ensure that K is small [19].

need prior knowledge to identify the desired solution from the solution set. This is often accomplished by solving

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{x}\|_p, \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \quad (18)$$

Here $p \geq 0$ should be chosen appropriately to incorporate the prior knowledge, which in our case is that $\mathbf{x}$ is sparse. It is well known that a sparsity constraint requires $p \leq 1$ [12]. The solution for $p = 0$ (commonly referred to as the ℓ_0 optimization method) has the smallest number of non-zero components. However, this requires solving a combinatorial problem [20]. Hence researchers either suggest using $0 < p < 1$ [21], or use some smoothed approximation of the zero norm [15]. The later is often referred to as the ℓ_0 approximation algorithms. Taking $p = 1$ gives a convex problem popularly known as Basis Pursuit (BP) [13], [20]. It is well known that under certain moderate assumptions [13], [20], [22] BP yields the unique sparse solution. We note that $p = 2$ has been a very common choice for regularization, particularly in maximum a posteriori estimation and optimal filtering [23]. However, it is not recommended for sparse recovery [22] (see Figure-2).

The existing BP algorithms typically don't produce a satisfactory solution within the time-frame available to solve the IUS problem. In the sequel we propose a new approach to overcome this hurdle. Certain ℓ_0 approximation methods [15], [21], [24] with iterative re-weighted least square (IRLS) type architecture are known to offer super-quadratic convergence when initialized sufficiently close to the final solution [21]. However, these methods being non-convex, may produce a local optimum. The ℓ_1 methods, being convex, yield the global solution, but typically need a large number of iterations. Hence, we aim to start with an ℓ_1 approach, and run it until it produces a rough solution. Then we handover to an IRLS type ℓ_0 approach called ISL0 with this rough solution as the initial guess. In particular we propose a new ℓ_1 optimization strategy to produce this rough estimate in a few iterations.

D. The proposed ℓ_1 optimization strategy

Our ℓ_1 optimization strategy, which is based on the Theorem-1, is designed to inherit the IRLS type architecture [21], and thereby produces a rough solution in a few iterations. Since the ℓ_1 algorithm will be used for initialization of ISL0, we shall call it Initialization ℓ_1 (INL-1).

Theorem 1: Let $\mathbf{a}_i$ be the i th column of $\mathbf{A}$, and $\mathbf{1}$ denote the $G.N_1$ dimensional vector of all ones. Consider the optimization problem in $\mathbf{g}$:

$$\underset{\mathbf{g}}{\text{maximize}} \quad (\mathbf{g}^* \mathbf{y} + \mathbf{y}^* \mathbf{g})/2 \quad (19a)$$

$$\text{subject to } \mathbf{g}^* \mathbf{a}_i \mathbf{a}_i^* \mathbf{g} \leq 1, \quad i = 1, \dots, G.N_1. \quad (19b)$$

The Lagrangian dual of (19) is

$$\underset{\mathbf{z}}{\text{minimize}} \quad \{\mathbf{1}^T \mathbf{z} + \mathbf{y}^* [\mathbf{A} \text{diag}(\mathbf{z}) \mathbf{A}^*]^{-1} \mathbf{y}\}/2 \quad (20a)$$

$$\text{subject to } \mathbf{z} \geq 0, \quad (20b)$$

where $\mathbf{z}$ is the variable of optimization. Let $\mathbf{z}_*$ be the optimal solution to (20), and $\mathbf{g}_*$ be the optimal solution to (19). Then $\text{diag}(\mathbf{z}_*) \mathbf{A}^* \mathbf{g}_*$ the optimal solution to the problem

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{x}\|_1, \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \quad (21)$$

TABLE I
PRIMAL-DUAL ALGORITHM FOR INITIALIZATION ℓ_1 (INL-1)

Initialization: Set $\mathbf{g} = \mathbf{0}$, $\mathbf{z} = \mathbf{0}$, and $\mu, 0 < \alpha \leq 1$.
repeat
1. Compute primal-dual search directions: $\Delta \mathbf{g} = (\mathbf{A} \text{diag}(\mathbf{z}) \mathbf{A}^* - \mathbf{J}^* \mathbf{S} \mathbf{J})^{-1} (-\mathbf{y} + \mu \mathbf{J}^* \mathbf{b})$ $\Delta \mathbf{z} = -(\mathbf{z} + \mu \mathbf{b} + \mathbf{S} \mathbf{J} \Delta \mathbf{g})$; where, $\mathbf{S} = [\text{diag}\{f(\mathbf{g})\}]^{-1} \text{diag}(\mathbf{z})$.
2. Find $0 < s \leq 1$ such that: $f_i(\mathbf{g} + s \Delta \mathbf{g}) \leq 0, \mathbf{z}(i) + s \Delta \mathbf{z}(i) \geq 0; \forall i$.
3. Set $\mathbf{g} = \mathbf{g} + s \Delta \mathbf{g}, \mathbf{z} = \mathbf{z} + s \Delta \mathbf{z}, \mu = \alpha \mu$.
4. Compute $\mathbf{x} = \text{diag}(\mathbf{z}) \mathbf{A}^* \mathbf{g}$.
until (A rough estimate of optimal $\mathbf{x}$ is obtained)
Output: $\mathbf{x}^{(1)} = \mathbf{x}$.

Proof: See Appendix A. ■

We use a primal-dual algorithm to solve the primal-dual pair (19) and (20) concurrently. In this case the primal-dual path following method [25] reduces to the form outlined in Table-I, where we write $f(\mathbf{g}) = [f_1(\mathbf{g}) \ f_2(\mathbf{g}) \ \dots \ f_{G.N_1}(\mathbf{g})]^T$, with

$$f_i(\mathbf{g}) = \mathbf{g}^* \mathbf{a}_i \mathbf{a}_i^* \mathbf{g} - 1, \quad \text{for } i = 1, 2, \dots, G.N_1. \quad (22)$$

In addition we define the vector $\mathbf{b}$ such that

$$[\mathbf{b}]_i = 1/f_i(\mathbf{g}), \quad \text{for } i = 1, \dots, G.N_1. \quad (23)$$

It can be shown that the Jacobean matrix of $f(\mathbf{g})$ is

$$\mathbf{J} = [\mathbf{a}_1 \mathbf{a}_1^* \mathbf{g} \quad \mathbf{a}_2 \mathbf{a}_2^* \mathbf{g} \quad \dots \quad \mathbf{a}_{G.N_1} \mathbf{a}_{G.N_1}^* \mathbf{g}]^T. \quad (24)$$

In Table I, μ controls the surrogate duality gap, and α controls the step length in backtracking line search. As suggested in [25], we take $\mu = 0.5$ and $\alpha = 0.1$.

We terminate the algorithm when a sparse enough $\mathbf{x}$ is produced. The measure of sparsity relies on the unique representation principle: a sparse solution can be recovered only if it has at most $M/2$ non-zero components [21]. After we compute $\mathbf{x}$ in Step-4 of Table-I, we form its thresholded version $\hat{\mathbf{x}}$ by retaining its $M/2$ most significant components, while setting others to zero. Then we compute

$$\kappa = \|\hat{\mathbf{x}}\|_2^2 / \|\mathbf{x}\|_2^2. \quad (25)$$

Clearly $\kappa < 1$. Moreover, $\kappa \approx 1$ indicates that $\mathbf{x}$ is very close to its optimal value. Since our aim is to produce only a rough estimate of the optimal $\mathbf{x}$, we stop well before κ reaches 1. The terminating value of κ being an interesting design parameter, will be studied in detail in Section-IV.

E. Smoothed ℓ_0 norm minimization [15]

We use the rough solution produced by the primal-dual method to initialize the improved smoothed ℓ_0 (ISL0) algorithm [15], [24]. ISL0 approximates $\|\mathbf{x}\|_0$ by the sum of Gaussian functions [15], [24]. In effect, it minimizes

$$L_\sigma(\mathbf{x}) := - \sum_{j=1}^{G.N_1} e^{-\frac{|\mathbf{a}_j^T \mathbf{x}|^2}{2\sigma^2}} + \frac{\lambda}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad (26)$$

to estimate $\mathbf{x}$ from (17). Here σ is a small real number and $\lambda > 0$. In principle, one should take $\sigma \rightarrow 0$ to achieve a close

TABLE II
ISL0 ALGORITHM

Input: $\mathbf{x}, \sigma_{st}, \lambda, \sigma_0$.
Initialization: $\sigma = \sigma_{st}$, and $\rho, \eta, \gamma \in [0, 1], i = 0, \beta = 1$.
repeat
1. while $L_\sigma\{\beta\zeta_\sigma(\mathbf{x}) + (1 - \beta)\mathbf{x}\} > L_\sigma(\mathbf{x})$
$\beta = \gamma\beta$.
end
2. $\mathbf{x}_o = \mathbf{x}$.
3. $\mathbf{x} = \beta\zeta_\sigma(\mathbf{x}) + (1 - \beta)\mathbf{x}$. Set $\beta = 1$.
4. If $\ \mathbf{x} - \mathbf{x}_o\ _2 < \eta\sigma$ then $\sigma = \rho\sigma$.
while $\sigma \geq \sigma_0$.
Output: $\hat{\mathbf{x}} = \mathbf{x}$.

approximation of the $\|\mathbf{x}\|_0$, but that makes L_σ highly non-smooth. To get around this difficulty ISL0 adopts the graduated non-convexity (GNC) strategy [26], where it uses σ to control the degree of non-convexity. ISL0 constructs a sequence $\sigma_n > \sigma_{n-1} > \dots > \sigma_0$ with $\sigma_j - \sigma_{j-1}$ being a small positive number for each j . Thereby, the minimizer of $L_{\sigma_{j-1}}$ is quite close to the minimizer of L_{σ_j} . If ISL0 finds out the minimizer of L_{σ_j} then it can be used to initialize the solver for minimizing $L_{\sigma_{j-1}}$. This procedure of successive optimization continues until the terminating value of σ i.e., σ_0 is reached. The value of σ_0 , can be chosen using a procedure outlined in [15].

A Gauss-Newton type convex-concave procedure is used by ISL0 to minimize L_σ for a fixed σ . This algorithm uses the fact that $L_\sigma(\mathbf{x})$ is decreasing along $\zeta_\sigma(\mathbf{x}) - \mathbf{x}$ [24], where

$$\zeta_\sigma(\mathbf{x}) = \lambda [W_\sigma(\mathbf{x})/\sigma^2 + \lambda \mathbf{A}^* \mathbf{A}]^{-1} \mathbf{A}^* \mathbf{y}, \quad (27)$$

$$\text{and } W_\sigma(\mathbf{x}) = \text{diag} \left\{ e^{-\frac{\|\mathbf{a}_1\|^2}{2\sigma^2}}, \dots, e^{-\frac{\|\mathbf{a}_{G \cdot N_1}\|^2}{2\sigma^2}} \right\}. \quad (28)$$

Furthermore, at the minimum point $\mathbf{x}_*$ of $L_\sigma(\mathbf{x})$ it holds that

$$\mathbf{x}_* = \zeta_\sigma(\mathbf{x}_*). \quad (29)$$

The ISL0 algorithm is given in Table-II. The procedure for finding the starting value of σ , denoted σ_{st} , will be described in the next section. The value of λ depends on the noise level. Following the recommendation in [27] we take $\lambda = c/\sqrt{2\sigma_e^2 M \log(G \cdot N_1)}$, where c depends on $\mathbf{A}$ [27]. In this work, we set $c = 100$ [27]. The ρ, η and γ are standard tuning parameters. We follow the recommendation in [15], [24] and set $\rho = 0.3, \eta = 0.5, \gamma = 0.5$ and $\sigma_0 = 0.001$.

F. The Handover process

Recall that the output $\mathbf{x}^{(1)}$ of the ℓ_1 -optimization in Table-I is used to initialize ISL0. For the GNC strategy of ISL0 to work well, we must choose the starting value of $\sigma_{st} = \sigma_n$ carefully. We cannot allow $\mathbf{x}^{(1)}$ to be far from the minimum point of $L_{\sigma_{st}}$. In particular, if we like to ensure that $\mathbf{x}^{(1)}$ is the minimizer of $L_{\sigma_{st}}$, then (27) and (29) require

$$\mathbf{x}^{(1)} = \lambda \left[W_{\sigma_{st}}(\mathbf{x}^{(1)})/\sigma_{st}^2 + \lambda \mathbf{A}^* \mathbf{A} \right]^{-1} \mathbf{A}^* \mathbf{y}. \quad (30)$$

We take σ_{st} as the solution to the least squares problem

$$\sigma_{st}^2 = \arg \min_{\sigma^2} \left\| \frac{W_\sigma(\mathbf{x}^{(1)})}{\sigma^2} \mathbf{x}^{(1)} - \lambda \mathbf{A}^* (\mathbf{y} - \mathbf{A} \mathbf{x}^{(1)}) \right\|_2^2, \quad (31)$$

induced by (30). The problem (31) being one dimensional, can be solved reliably by an interior trust region algorithm [28].

G. Acceleration via FFT

In this section, we explore how ISL0 can be accelerated by using FFT to handle the computationally demanding steps. We have applied the same procedure to accelerate the ℓ_1 algorithm as well³. ISL0 spends most of its time computing $\zeta_\sigma(\mathbf{x})$ via (27). By the matrix inversion lemma, (27) is equivalent to

$$\zeta_\sigma(\mathbf{x}) = W_\sigma^{-1}(\mathbf{x}) \mathbf{A}^* [\mathbf{I}/(\lambda\sigma^2) + \mathbf{A} W_\sigma^{-1}(\mathbf{x}) \mathbf{A}^*]^{-1} \mathbf{y}. \quad (32)$$

To compute $\zeta_\sigma(\mathbf{x})$ efficiently using FFT, rewrite (32) as

$$\zeta_\sigma(\mathbf{x}) = W_\sigma^{-1}(\mathbf{x}) \mathbf{A}^* \hat{\mathbf{z}} \quad (33)$$

$$\text{where, } \mathbf{y} = [\mathbf{R} + \mathbf{I}/(\lambda\sigma^2)] \hat{\mathbf{z}}, \quad \mathbf{R} = \mathbf{A} W_\sigma^{-1}(\mathbf{x}) \mathbf{A}^*.$$

Now partition $\mathbf{x}$ into G number of sub-vectors: $\mathbf{x} = [\mathbf{x}_1^\top \mathbf{x}_2^\top \dots \mathbf{x}_G^\top]^\top$ where length of each $\mathbf{x}_i$ is N_1 . Construct the matrix $\hat{\mathbf{F}}$ by extracting first N_1 columns of the Fourier matrix $\mathbf{F}$ in (8). We calculate (33) by using the following steps.

Step 1: Since $W_\sigma(\mathbf{x})$ is a diagonal matrix, it follows using (17) and (12) that

$$\mathbf{R} = \mathbf{A} W_\sigma^{-1}(\mathbf{x}) \mathbf{A}^* = \sum_{i=1}^{N_1} \mathbf{R}_i,$$

$$\mathbf{R}_i = [\Theta \text{diag}(\Theta^\top \mathbf{c}_i)] \hat{\mathbf{F}} W_\sigma^{-1}(\mathbf{x}_i) \hat{\mathbf{F}}^* [\Theta \text{diag}(\Theta^\top \mathbf{c}_i)]^\top,$$

$$[\hat{\mathbf{F}} W_\sigma^{-1}(\mathbf{x}_i) \hat{\mathbf{F}}^*]_{k,\ell} = \sum_{j=-N_1}^{N_1} \mathbf{w}_i e^{-i2\pi(k-\ell)j/N} = [\hat{\mathbf{F}} \mathbf{w}_i]_{k-\ell}$$

where the vector $\mathbf{w}_i$ is constructed by vectorizing the diagonal elements of $W_\sigma^{-1}(\mathbf{x}_i)$. Since $\mathbf{w}_i$ is real valued, we compute $\hat{\mathbf{F}} \mathbf{w}_i$ via FFT using $2.5N \log_2(N)$ floating point operations.

Since Θ is a row selector matrix, constructing $[\Theta \text{diag}(\Theta^\top \mathbf{c}_i)]$ does not require any floating point operations. By exploiting the symmetric structure of $\hat{\mathbf{F}} W_\sigma^{-1}(\mathbf{x}_i) \hat{\mathbf{F}}^*$, it can be verified that the construction of $\mathbf{R}_i$ requires extracting a small block of $\hat{\mathbf{F}} W_\sigma^{-1}(\mathbf{x}_i) \hat{\mathbf{F}}^*$ followed by $M(M+1)/2$ complex multiplications which requires $3M(M+1)$ flops. Finally, we compute $\mathbf{R} = \sum_{i=1}^G \mathbf{R}_i$ which requires $(G-1)(M+1)M$ flops. Hence this step requires $G\{2.5N \log_2(N) + 4M(M+1)\} - M(M+1)$ flops in total.

Step 2: Calculate $\hat{\mathbf{z}}$ by solving $[\mathbf{R} + \mathbf{I}/(\lambda\sigma^2)] \hat{\mathbf{z}} = \mathbf{y}$. By using Cholesky factorization, this $M \times M$ positive definite system of equations needs $4M^3/3$ flops to compute $\hat{\mathbf{z}}$.

Step 3: Next form $\mathbf{A}^* \hat{\mathbf{z}}$. Partition $\mathbf{A}^* \hat{\mathbf{z}} = [\tilde{\mathbf{z}}_1 \tilde{\mathbf{z}}_2 \dots \tilde{\mathbf{z}}_G]$, where each $\tilde{\mathbf{z}}_i = \hat{\mathbf{F}}^* [\Theta \text{diag}(\Theta^\top \mathbf{c}_i)]^\top \hat{\mathbf{z}}$ can be computed using the IFFT of $[\Theta \text{diag}(\Theta^\top \mathbf{c}_i)]^\top \hat{\mathbf{z}}$. Since forming $[\Theta \text{diag}(\Theta^\top \mathbf{c}_i)]^\top \hat{\mathbf{z}}$ takes $6M$ floating point operations, we need $G(5N \log_2(N) + 6M)$ floating point operations at this step.

³Recall that ISL0 and ℓ_1 carry out very similar calculations.

Step 4: As $W_\sigma^{-1}(\mathbf{x})$ is diagonal, we need $6GN_1$ more flops to compute $\zeta_\sigma(\mathbf{x}) = W_\sigma^{-1}(\mathbf{x})\mathbf{A}^*\hat{\mathbf{z}}$.

Thus in total we need $7.5GN \log_2(N) + 4GM(M+3) + 6GN_1 + 4M^3/3 - M^2$ flops to compute $\zeta_\sigma(\mathbf{x})$.

H. Thresholding the recovered signal from ISLO

The output of the ISLO algorithm $\bar{\mathbf{x}}$ is somewhat noisy because $\mathbf{y}$ is noise contaminated. We present a thresholding procedure to detect the transmitted codes from the noisy $\bar{\mathbf{x}}$. Denote the estimation error by $\mathbf{q} = \bar{\mathbf{x}} - \hat{\mathbf{x}}$, with $\hat{\mathbf{x}}$ denoting the true value of $\mathbf{x}$, i.e. $\hat{\mathbf{x}} := [\mathbf{h}_1^T(1:N_1) \quad \mathbf{h}_2^T(1:N_1) \quad \dots \quad \mathbf{h}_G^T(1:N_1)]^T$. Our thresholding procedure exploits some statistical property of $\mathbf{q}$. We have shown in Appendix B for small $\|\mathbf{q}\|_2$ that $\mathbf{q} = \mathbf{D}\mathbf{e}$, where

$$\mathbf{D} = \left[\frac{W_\sigma(\hat{\mathbf{x}})}{\lambda\sigma^2} \left\{ \mathbf{I} - \left[\text{diag} \left(\frac{\hat{\mathbf{x}}}{\sigma} \right) \right]^2 \right\} + \mathbf{A}^*\mathbf{A} \right]^{-1} \mathbf{A}^*. \quad (34)$$

We need $\hat{\mathbf{x}}$ to compute $\mathbf{D}$. Since $\hat{\mathbf{x}}$ is unknown, we replace $\hat{\mathbf{x}}$ in (34) by $\bar{\mathbf{x}}$. Partition $\mathbf{D}^T = [\mathbf{D}^{(1)} \quad \mathbf{D}^{(1)} \dots \mathbf{D}^{(G)}]$ such that each $\mathbf{D}^{(i)} \in \mathbb{C}^{M \times N_1}$. Also partition $\mathbf{q} = [\mathbf{q}_1^T \quad \mathbf{q}_2^T \dots \mathbf{q}_G^T]^T$. Since, the entries of $\mathbf{D}$ are independent of $\mathbf{e}$, and $M < N_1$, we know $\|\mathbf{q}_i\|_2^2 = \|[\mathbf{D}^{(i)}]^T \mathbf{e}\|_2^2$ has a generalized chi-square distribution of order M [29]. The procedure for computing the cumulative distribution function (CDF) of a generalized chi-square distribution has been described in [29]. The CDF of $\|\mathbf{q}_i\|_2^2$ at t is denoted by $\chi_i(t)$. The function χ_i depends on the noise variance σ_e^2 , and the singular values of $\mathbf{D}^{(i)}$.

Partition $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1^T \quad \bar{\mathbf{x}}_2^T \dots \bar{\mathbf{x}}_G^T]^T$ such that every $\bar{\mathbf{x}}_i \in \mathbb{C}^{N_1}$. Recall that $\mathcal{L} = \{i : \|\bar{\mathbf{x}}_i\|_2 \neq 0\}$. Given $\bar{\mathbf{x}}_i$ we consider the two hypotheses: $[\mathcal{H}_0 : i \notin \mathcal{L}; \quad \mathcal{H}_1 : i \in \mathcal{L}]$. The task here is to choose the positive numbers $\{t_i\}_{i=1}^G$ to be used as thresholds. In particular, we decide $\mathcal{H}_1$, i.e. $i \in \mathcal{L}_1$ if $\|\bar{\mathbf{x}}_i\|_2^2 > t_i$. Otherwise we decide $i \notin \mathcal{L}$ or $\mathcal{H}_0$. The value of t_i controls the false alarm probability $\psi_i = 1 - \chi_i(t_i)$ associated with the code $\mathbf{c}_i$. It is natural to take $\{t_i\}_{i=1}^G$ so that the false alarm probabilities are the same for all i , i.e.

$$\psi = 1 - \chi_i(t_i), \quad \forall i. \quad (35)$$

In practice we start from a desired overall false alarm probability P_{fa} , and compute ψ by solving $P_{fa} = 1 - (1 - \psi)^G$. From ψ we get $t_i = \chi_i^{-1}(1 - \psi)$.

IV. SIMULATION RESULTS

In this section we present some simulation results using an LTE network [2]. Similar results for WiMAX can be found in [16]. We simulate a 7.68 MHz LTE system [2], with $N = 6144$ subcarriers. The subcarrier frequency spacing is 1.25 kHz, and the sampling interval $T_s = 130$ ns. The modulation pulse is a root-raised-cosine function with a roll-off 0.22, and duration $10T_s$. IUS subchannels consist of 839 adjacent subcarriers. The wireless cell radius is 1.5 km, which corresponds to $D = 80$. The wireless channels are modeled according to a mixed channel model specified by the ITU IMT-2000 standards: Ped-A, Ped-B, and Veh-A. For each ST the simulator selects one of the above channel models uniformly at random. The mobile speed varies in the interval $[0, 5]$ m/s

TABLE III
THE VALUE OF P_s FOR DIFFERENT VALUES OF $\bar{\kappa}$, SNR, AND THE NUMBER OF USERS K .

K	SNR (dB)	P_s				
		$\bar{\kappa} = 0.6$	$\bar{\kappa} = 0.7$	$\bar{\kappa} = 0.8$	$\bar{\kappa} = 0.9$	$\bar{\kappa} = 0.99$
6	10	0.94	0.95	0.95	0.95	0.95
6	4	0.91	0.93	0.93	0.935	0.935
10	10	0.905	0.905	0.91	0.91	0.91
10	4	0.82	0.85	0.86	0.86	0.86

for Ped-A, Ped-B channels, and $[5, 20]$ m/s for Veh-A. The channel impulse response of the STs have a maximum of $P_{\max} = 30$ taps [3]. Similar to [5], [11], we assume that BS has an approximate knowledge about $P_{\max}$, and we set $N_1 = P_{\max} + D$ in (16). The IUS codes are the Zadoff-Chu (ZC) sequences of length 839. The k -th element $Z^u(k)$ of the u -th root ZC sequence is given by

$$Z^u(k) = e^{-i\pi u k(k+1)/M}, \quad 0 \leq k < M, \quad (36)$$

where u is an integer with $0 < u < M$. Different IUS codes can be obtained by cyclic shifting the u -th root ZC sequence. For example, the components of $\ell + 1$ -th column of $\mathbf{C}$ can be generated from the u -th root ZC sequence as

$$c_{k+1, \ell+1} = Z^u((k + \ell n_{CS}) \bmod M), \quad 0 \leq k < M, \quad (37)$$

where the value of n_{CS} depends on the wireless cell radius. For a cell radius 1.5 km, we can use $n_{CS} = 15$ [30]. Then for a given value of u , we can generate maximum $\lfloor M/n_{CS} \rfloor = 55$ IUS codes [19], i.e., we set $G = 55$. For performance comparison, we consider the SMUD [5], SRMD [6], and RA-GLRT [3] algorithms. The signal to noise ratio (SNR) is defined as $\text{SNR} = 10 \log_{10}(\sigma_h^2/\sigma_e^2)$, where σ_h^2 is the variance of a channel tap, and σ_e^2 is the variance of a component of $\mathbf{e}$, respectively [11]. The code detection capability of different algorithms is assessed in terms of probability of success denoted by P_s [31]. Recall that the set of active IUS code indices is $\mathcal{L}$. Let $\hat{\mathcal{L}}$ be the set of code indices detected by an algorithm. Then P_s is the probability that $\mathcal{L} = \hat{\mathcal{L}}$. The following results are based on 200 independent Monte-Carlo simulations.

Recall that the terminating value $\bar{\kappa}$ of κ (see (25)) determines when we stop iterations of INL1. We must find a good $\bar{\kappa}$ to avoid unnecessary iterations in generating the rough estimate (to be used as the initial guess by ISLO) obtained via INL1. In Table III we tabulate the IUS code detection performance of Handover approach for different values of $\bar{\kappa}$. Note that the performance of Handover remains almost same for $\bar{\kappa} \geq 0.8$. Hence we use $\bar{\kappa} = 0.8$ in the subsequent simulations.

We have also compared the iterations needed by INL1 (Table-I) with that needed by a popular ℓ_1 -norm minimization algorithm called ℓ_1 -magic [32]. From a wide range of simulations with different number of STs and SNR conditions we have found that INL1 requires 4 – 5 iterations to produce $\kappa \geq 0.80$, while ℓ_1 -magic takes 12 – 18 iterations in average. It should also be noted that the number of flops required for INL1 in every iteration is smaller than ℓ_1 -magic.

TABLE IV
 P_{md} FOR HANDOVER FOR DIFFERENT VALUES OF P_{fa} AND THE NUMBER OF USERS K . SNR= 4 dB.

K	P_{md}					
	$P_{fa} = 0.3$	0.2	0.15	0.1	0.01	0.001
6	0.06	0.06	0.065	0.065	0.065	0.08
10	0.12	0.14	0.14	0.14	0.15	0.16

Recall from (35) that Handover uses the thresholds $\{t_i\}_{i=1}^G$ to detect the active codes. The values of $\{t_i\}_{i=1}^G$ are determined by the desired false alarm probability P_{fa} . A false alarm occurs when the algorithm includes the index of an inactive code into $\hat{\mathcal{L}}$. To obtain an operating value of P_{fa} we construct the receiver operating characteristic (ROC) [33]. The ROC shows how the mis-detection probability P_{md} varies with P_{fa} . A mis-detection occurs when the algorithm fails to include the index of an active code into $\hat{\mathcal{L}}$. If we like to have a smaller P_{fa} , we need to increase the values of $\{t_i\}_{i=1}^G$. But that would increase P_{md} . We examine the ROC to choose an operating P_{fa} that optimally balances the conflicting objectives of making both P_{fa} and P_{md} as small as possible. In Table IV, we tabulate the values of P_{md} obtained from numerical simulations for different values of P_{fa} . As can be seen in Table IV, regardless of the number of IUS users, the values of P_{md} remain almost similar for $P_{fa} \geq 0.1$. But as we decrease P_{fa} below 0.1, the value of P_{md} starts increasing quickly. Hence to make the optimal trade-off, we use $P_{fa} = 0.1$ in the subsequent simulations.

Fig. 1. Code detection performance of Handover for different values of τ . The value of $P_{fa} = 1e-1$.

Recall that 839 adjacent subcarriers are reserved for IUS purpose. Thus, $\mathbf{y} \in \mathbb{C}^{839}$. The neighbouring components of $\mathbf{y}$ are highly correlated, and contain redundant information. According to the theory of compressive sensing [12], [13], when the components of $\mathbf{y}$ are highly correlated then the sparse recovery algorithm can yield near optimum estimation performance by using an appropriately chosen small sized sub-vector of $\mathbf{y}$. By working with a sub-vector we can reduce the complexity of the algorithm significantly. The size of the sub-vector depends on the dimension of $\mathbf{x}$ in (17) i.e., value of

$G.N_1$ and the sparsity level of $\mathbf{x}$, i.e., the number of active users K . In particular, it has been shown in [12], [13] that one way to choose this sub-vector is to pick $\lceil \tau K \ln(G.N_1) \rceil$ number of components of $\mathbf{y}$ randomly, where τ is problem dependent. As we show shortly, the Handover algorithm can detect $K \leq 11$ number of STs with high probability. Hence we set $K = 11$. Then we find an appropriate value of τ by examining the code detection performance of the Handover algorithm for different values of τ . The results are shown in Figure-1. Note that for all K and SNR settings P_s is fairly constant for $\tau \geq 2.5$. But if we decrease τ below 2.5, then P_s starts decreasing noticeably. Since a larger τ implies higher computational complexity of Handover, we use $\tau = 2.5$ in the following simulations. It should be noted that when we work with a sub-vector of $\mathbf{y}$, we must adjust the row selector matrix Θ accordingly to select the rows of $\mathbf{y}$ used to build the subvector. This is the only adjustment needed in the Handover algorithm.

In Section-III-C we stated that ℓ_1 -norm regularization is better than ℓ_2 -norm regularization for imposing a sparsity prior in (18). This is illustrated in Figure-2 using a typical simulation result from an arbitrary realization. Here we plot the typical channel impulse response (CIR) spectra for 6 active users and SNR =10 dB. The red curves in Figure-2 represent the true CIR spectra. Form the ℓ_2 -norm regularized CIR spectrum in Figure-2(a) we are unable to detect two STs. Figure-3 compares the value of P_s achieved by different algorithms as a function of K at different values of SNR. Handover performs the best, while RA-GLRT performs the worst. For instance, the value of P_s for Handover at 4 dB SNR is significantly higher than that achieved by other algorithms even at 10 dB SNR. Thus the gain in detection performance offered by Handover is significant when compared to the other algorithms. Also note that Handover is robust to noise. Its performance at 4 dB SNR is just slightly worse than that at 10 dB SNR with as many as 9 users. We point out that RA-GLRT performs poorly as low SNR, hence the corresponding performance figures are not shown here.

In Figure-4 we plot the mean squared error associated with the estimate of the power Γ as a function of K for different values of SNR. As expected, the MSE of power estimate increases with K and decrease in SNR, while the Handover algorithm outperforms the other algorithms by a respectable margin. This margin is significantly big when we consider the similar plots for timing error estimate in Figure-4(b). Table-V compares the computational complexity of different algorithms. RA-GLRT requires $96N \log_2(N) = 7.4 \times 10^6$ flops [3]. SMUD and Handover algorithms being iterative, their complexities are obtained by multiplying the flops-per-iteration by the average number of iterations taken to produce the final result. Note that the complexity of Handover is about 3-4 times higher than SMUD. To the best of our knowledge, the complexity figures for SRMD are not available. However, according to our experience in simulations, SRMD appears to have a complexity 2.5-3 times higher than SMUD⁴. Note

⁴In our simulations we have implemented an accelerated version of the original SRMD algorithm [6] using FFT.

Fig. 2. Estimation of sparse CIR by using ℓ_2 -norm and ℓ_1 -norm optimization (see (18)). Total number of STs is $K = 6$ and SNR = 10 dB, $P_{fa} = 0.1$.

Fig. 3. Active code detection probability by different algorithms. The value of $P_{fa} = 1e-1$ for Handover algorithm.

TABLE V
COMPUTATIONAL COMPLEXITY COMPARISON. SNR=10 dB.

K	Flops (in 10^8)		K	Flops (in 10^8)	
	Handover	SMUD		Handover	SMUD
3	6.63	2.29	6	9.95	3.1

in Figures 3-4 that SRMD is the second best in estimation and detection performance, and Handover outperforms SRMD by a big margin. This enhancement in performance comes in expense of only about 35% increase in computational load. We have observed that this increase in computation load is much smaller for WiMAX [16].

V. CONCLUSIONS

In this work, we have proposed a novel sparse signal recovery framework for the OFDMA initial uplink synchronization (IUS) problem. We have presented a generic signal model applicable to any OFDMA system. In addition, we have given an algorithm to implement our idea. The algorithm has been tested on a simulated LTE environment. In comparison to other well known methods, the proposed algorithm shows a significant improvement in the multiuser code detection

performance, and is robust to noise. Furthermore, it provides superior quality estimates of timing error and channel power.

Initial uplink synchronization is considered as a bottleneck in realizing the future “network of things” [34]. The current methods for the IUS problem are considered inadequate to handle the expected IUS load in these future networks. In order to handle the dramatically increased random access load in a 5G system, some significant changes are being made. One of them is MIMO random access (see [35] and references therein). Given the success of the OFDMA technology in modern communication systems, the proposed sparse recovery framework appears to be a credible candidate for handling the increasing IUS load in future technologies. This viewpoint is also shared by the authors of in the context of 5G architecture. This is also envisaged by the overview paper [35].

APPENDIX A PROOF OF THEOREM-1

Given $\mathbf{u} \in \mathbb{C}^{G \cdot N_1}$ we define the infinite norm

$$\|\mathbf{u}\|_{\infty} = \max_{k \in \{1, 2, \dots, G \cdot N_1\}} |[\mathbf{u}]_k|.$$

write $\langle \mathbf{v}, \mathbf{u} \rangle := (\mathbf{u}^* \mathbf{v} + \mathbf{v}^* \mathbf{u})/2$. Hölder’s inequality gives

$$\langle \mathbf{v}, \mathbf{u} \rangle \leq \|\mathbf{v}\|_1 \|\mathbf{u}\|_{\infty}, \quad (38)$$

provided that both $\|\mathbf{v}\|_1$ and $\|\mathbf{u}\|_{\infty}$ exist. The following fact is well known, see e.g. [36]:

Proposition 1: Let $\mathcal{K} = \{k : |[\mathbf{u}]_k| = \|\mathbf{u}\|_{\infty}\}$. Then $\langle \mathbf{v}, \mathbf{u} \rangle = \|\mathbf{v}\|_1 \|\mathbf{u}\|_{\infty}$ only if

- $[\mathbf{v}]_k = 0$, for all $k \notin \mathcal{K}$;
- For all $k \in \mathcal{K}$ it holds that $[\mathbf{v}]_k = \mu_k \text{conj}([\mathbf{u}]_k)/|[\mathbf{u}]_k|$ for some non-negative number $\mu_k \in \mathbb{R}$. ■

By applying a key result [36] on norm minimization in Banach spaces we get

$$\min_{\mathbf{x} \in \mathbb{V}} \|\mathbf{x}\|_1 = \max_{\mathbf{g} \in \mathbb{U}} \frac{1}{2} (\mathbf{y}^* \mathbf{g} + \mathbf{g}^* \mathbf{y}), \quad (39)$$

where the sets $\mathbb{V}$ and $\mathbb{U}$ are defined as

$$\mathbb{V} = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{y}\}, \quad \mathbb{U} = \{\mathbf{g} : \|\mathbf{A}^* \mathbf{g}\|_{\infty} \leq 1\}.$$

Fig. 4. MSE of initial uplink synchronization parameter estimations. (a) MSE of estimated channel power versus the number of active STs, (b) MSE of estimated timing error versus the number of active STs.

In addition, let $\mathbf{x}_*$ be the solution to the optimization problem in the left hand side of (39), and let $\mathbf{g}_*$ be the solution to the optimization problem in the right hand side of (39). Then

$$\langle \mathbf{x}_*, \mathbf{A}^* \mathbf{g}_* \rangle = \|\mathbf{x}_*\|_1 \|\mathbf{A}^* \mathbf{g}_*\|_\infty. \quad (40)$$

Note that problem in the right hand side of (39) is same as (19). Next we show that the Lagrangian dual of (19) is (20). The Lagrangian associated with (19) is

$$L(\mathbf{g}, \mathbf{z}) = \frac{1}{2} \{ \mathbf{g}^* \mathbf{y} + \mathbf{y}^* \mathbf{g} - \mathbf{g}^* \mathbf{A} \text{diag}(\mathbf{z}) \mathbf{A}^* \mathbf{g} + \mathbf{1}^T \mathbf{z} \} \quad (41)$$

where $\mathbf{1}$ is a $G.N_1$ dimensional vector of all ones, and $\mathbf{z}/2$ is the real-valued vector of Lagrange multipliers. Each component of $\mathbf{z}$ must be non-negative, and we denote it by $\mathbf{z} \geq 0$. From the theory of least squares we know

$$\arg \max_{\mathbf{g}} L(\mathbf{g}, \mathbf{z}) = [\mathbf{A} \text{diag}(\mathbf{z}) \mathbf{A}^*]^{-1} \mathbf{y}, \quad (42)$$

$$\max_{\mathbf{g}} L(\mathbf{g}, \mathbf{z}) = \frac{1}{2} \{ \mathbf{1}^T \mathbf{z} + \mathbf{y}^* [\mathbf{A} \text{diag}(\mathbf{z}) \mathbf{A}^*]^{-1} \mathbf{y} \}. \quad (43)$$

This implies that the Lagrangian dual of (19) is (20).

Recall the $\mathbf{z}_*$ denote the solution to (20). It remains to show $\mathbf{x}_* = \text{diag}(\mathbf{z}_*) \mathbf{A}_* \mathbf{g}_*$. Note that by setting $\mathbf{x}_* = \text{diag}(\mathbf{z}_*) \mathbf{A}_* \mathbf{g}_*$ we do satisfy $\mathbf{y} = \mathbf{A} \mathbf{x}_*$. It remains to verify the condition (40) is also satisfied when we set $\mathbf{x}_* = \text{diag}(\mathbf{z}_*) \mathbf{A}_* \mathbf{g}_*$.

Now, it must hold that $\|\mathbf{A}^* \mathbf{g}_*\|_\infty = 1$. This is because if $\|\mathbf{A}^* \mathbf{g}_*\|_\infty < 1$, then we could always multiply $\mathbf{g}_*$ by a suitable real scalar $\kappa > 1$ such that $\|\mathbf{A}^* (\kappa \mathbf{g}_*)\|_\infty = 1$, and

$$\mathbf{g}_*^* \mathbf{y} + \mathbf{y}^* \mathbf{g}_* < (\kappa \mathbf{g}_*)^* \mathbf{y} + \mathbf{y}^* (\kappa \mathbf{g}_*),$$

leading to a contradiction. This means that the set

$$\mathcal{K} = \{i : \mathbf{g}_*^* \mathbf{a}_i \mathbf{a}_i^* \mathbf{g}_* = 1\}.$$

is nonempty. However, the complementary slackness condition associated with the primal-dual pair (19) and (20) implies that implies that $[\mathbf{z}_*]_i = 0$ for all $i \notin \mathcal{K}$. Hence

$$\|\mathbf{x}_*\|_1 = \sum_{i \in \mathcal{K}} [\mathbf{z}_*]_i |\mathbf{a}_i^* \mathbf{g}_*| = \sum_{i \in \mathcal{K}} [\mathbf{z}_*]_i. \quad (44)$$

The last equality follows because $|\mathbf{a}_i^* \mathbf{g}_*| = 1, \forall i \in \mathcal{K}$. Then

$$\mathbf{g}_*^* \mathbf{A} \mathbf{x}_*(1) = \sum_{i \in \mathcal{K}} \mathbf{g}_*^* \mathbf{a}_i \mathbf{a}_i^* \mathbf{g}_* [\mathbf{z}_*]_i = \sum_{i \in \mathcal{K}} [\mathbf{z}_*]_i = \|\mathbf{x}_*\|_1,$$

and consequently

$$\begin{aligned} \langle \mathbf{x}_*, \mathbf{A}^* \mathbf{g}_* \rangle &= \frac{1}{2} (\mathbf{g}_*^* \mathbf{A} \mathbf{x}_* + \mathbf{x}_*^* \mathbf{A}^* \mathbf{g}_*) = 1 \times \|\mathbf{x}_*\|_1 \\ &= \|\mathbf{A}^* \mathbf{g}_*\|_\infty \|\mathbf{x}_*\|_1, \end{aligned}$$

and thus (40) is verified.

APPENDIX B PROOF OF THE RELATION IN (34)

Using (27) and (29), we can write

$$\bar{\mathbf{x}} = [\mathbf{W}_\sigma(\bar{\mathbf{x}})/(\lambda\sigma^2) + \mathbf{A}^* \mathbf{A}]^{-1} \mathbf{A}^* \mathbf{y}. \quad (45)$$

Using the definition of $\mathbf{W}_\sigma$, and assuming $\|\mathbf{q}\|_2$ is small, the first order Taylor series expansion of $\mathbf{W}_\sigma(\bar{\mathbf{x}})$ around $\dot{\mathbf{x}}$ is

$$\mathbf{W}_\sigma(\bar{\mathbf{x}}) = \mathbf{W}_\sigma(\dot{\mathbf{x}}) - \mathbf{W}_\sigma(\dot{\mathbf{x}}) \text{diag}(\dot{\mathbf{x}}/\sigma^2) \text{diag}(\mathbf{q}). \quad (46)$$

In the following we write $\mathbf{W} = \mathbf{W}_\sigma(\dot{\mathbf{x}})/(\lambda\sigma^2)$ for simplicity. Using (46) in (45),

$$\mathbf{A}^* \mathbf{y} = \{\mathbf{W} - \mathbf{W} \text{diag}(\dot{\mathbf{x}}/\sigma^2) \text{diag}(\mathbf{q}) + \mathbf{A}^* \mathbf{A}\} (\dot{\mathbf{x}} + \mathbf{q})$$

Using $\mathbf{y} = \mathbf{A} \dot{\mathbf{x}} + \mathbf{e}$ in above we have

$$\begin{aligned} \mathbf{A}^* \mathbf{e} &= \{\mathbf{W} - \mathbf{W} \text{diag}(\dot{\mathbf{x}}/\sigma^2) \text{diag}(\mathbf{q})\} \dot{\mathbf{x}} \\ &\quad + \{\mathbf{W} - \mathbf{W} \text{diag}(\dot{\mathbf{x}}/\sigma^2) \text{diag}(\mathbf{q}) + \mathbf{A}^* \mathbf{A}\} \mathbf{q}. \end{aligned}$$

For a small σ , we can neglect $\mathbf{W}_\sigma(\dot{\mathbf{x}}) \dot{\mathbf{x}}/\sigma^2$. Hence by neglecting second order terms in $\mathbf{q}$ we have

$$\mathbf{A}^* \mathbf{e} = \mathbf{W} \{\mathbf{I} - [\text{diag}(\dot{\mathbf{x}}/\sigma)]^2 + \mathbf{A}^* \mathbf{A}\} \mathbf{q},$$

which is same as (34)

ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous referees for their insightful comments and suggestions.

REFERENCES

- [1] "IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems," *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004)*, pp. 1–2080, 2009.
- [2] "3rd Generation Partnership Project; technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 10)," *European Telecommunications Standards Institute*, 2011.
- [3] L. Sanguinetti, M. Morelli, and L. Marchetti, "A random access algorithm for LTE systems," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 1, pp. 49–58, 2013. [Online]. Available: <http://dx.doi.org/10.1002/ett.2575>
- [4] J. K. et al, "Comments on OFDMA ranging scheme described in IEEE 802.16ab-01/01r1," *IEEE 802.16abc-01/24*, 2001.
- [5] M. Ruan, M. Reed, and Z. Shi, "Successive multiuser detection and interference cancellation for contention based OFDMA ranging channel," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 481–487, 2010.
- [6] C.-L. Lin and S.-L. Su, "A robust ranging detection with MAI cancellation for OFDMA systems," in *Advanced Communication Technology (ICACT), 2011 13th International Conference on*, 2011, pp. 937–941.
- [7] Y. Zhou, Z. Zhang, and X. Zhou, "OFDMA initial ranging for IEEE 802.16e based on time-domain and frequency-domain approaches," in *Communication Technology, 2006. ICCT '06. International Conference on*, 2006, pp. 1–5.
- [8] D. H. Lee, "OFDMA uplink ranging for IEEE 802.16e using modified generalized chirp-like polyphase sequences," in *Internet, 2005. The First IEEE and IFIP International Conference in Central Asia on*, 2005, pp. 1–5.
- [9] X. Fu, Y. Li, and H. Minn, "A new ranging method for OFDMA systems," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 2, pp. 659–669, 2007.
- [10] S. Barbarossa, M. Pompili, and G. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 2, pp. 474–486, 2002.
- [11] L. Sanguinetti and M. Morelli, "An initial ranging scheme for the IEEE 802.16 OFDMA uplink," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 9, pp. 3204–3215, 2012.
- [12] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, mar 2008.
- [13] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb. 2006.
- [14] "Compressive sensing resources: Rice University," <http://dsp.rice.edu/cs>.
- [15] M. M. Hyder and K. Mahata, "An improved smoothed ℓ^0 approximation algorithm for sparse representation," *IEEE transactions on Signal Processing*, vol. 58, no. 4, pp. 2194–2205, april 2010.
- [16] —, "A sparse recovery method for initial ranging in IEEE 802.16 OFDMA systems," *CoRR*, 2015. [Online]. Available: <http://arxiv.org/abs/1505.05935>
- [17] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.
- [18] "3GPP TR 37.868 V11.0.0: Third generation partnership project; technical specification group radio access network; study on RAN improvements for machine-type communications," *European Telecommunications Standards Institute*, 2011.
- [19] J. Kim, D. Munir, S. Hasan, and M. Chung, "Enhancement of LTE RACH through extended random access process," *Electronics Letters*, vol. 50, no. 19, pp. 1399–1400, September 2014.
- [20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.
- [21] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, pp. 600–616, Mar 1997.
- [22] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 14–20, mar 2008.
- [23] T. Kailath, *Linear Estimation*. Prentice-Hall, Inc., 2000.
- [24] M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [26] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
- [27] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (LASSO)," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2183–2202, May 2009.
- [28] T. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.
- [29] R. B. Davies, "Algorithm as 155: The distribution of a linear combination of χ^2 random variables," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 3, pp. 323–333, 1980.
- [30] H. S. Jang, S. M. Kim, K. S. Ko, J. Cha, and D. K. Sung, "Spatial group based random access for M2M communications," *Communications Letters, IEEE*, vol. 18, no. 6, pp. 961–964, June 2014.
- [31] Y. Xie, Y. Eldar, and A. Goldsmith, "Reduced-dimension multiuser detection," *Information Theory, IEEE Transactions on*, vol. 59, no. 6, pp. 3858–3874, June 2013.
- [32] E. J. Candès and J. Romberg. (2005) ℓ^1 -magic: Recovery of sparse signals via convex programming. [Online]. Available: <http://www.acm.caltech.edu/l1magic/downloads/l1magic.pdf>
- [33] S. M. Kay, *Fundamentals of statistical signal processing, V. II: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1998.
- [34] N. Himayat, S. Talwar, K. Johnsson, S. Andreev, O. Galinina, and A. Turlikov, "Proposed IEEE 802.16p performance requirements for network entry by large number of devices," *IEEE 802.16 Broadband Wireless Access Working Group*, pp. 1–11, 2011.
- [35] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *Access, IEEE*, vol. 3, pp. 195–208, 2015.
- [36] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1997.