
Article:
Nan, Y, Sun, X and Zhang, LX (2016) Joint Channel Estimation Algorithm via Weighted Homotopy for Massive MIMO OFDM System. Digital Signal Processing, 50. pp. 34-42. ISSN: 1051-2004

https://doi.org/10.1016/j.dsp.2015.11.010

Joint Channel Estimation Algorithm via Weighted Homotopy for Massive MIMO OFDM System

Yang Nan[a,1,*], Xin Sun[a,1], Li Zhang[b,1]

[a] *School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China*
[b] *School of Electronic and Electrical Engineering, University of Leeds, Leeds, the United Kingdom*

Abstract

Massive (or large-scale) multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) system is widely acknowledged as a key technology for future communication. One main challenge to implement this system in practice is the high dimensional channel estimation, where the large number of channel matrix entries requires prohibitively high computational complexity. To solve this problem efficiently, a channel estimation approach using few number of pilots is necessary. In this paper, we propose a weighted Homotopy based channel estimation approach which utilizes the sparse nature in MIMO channels to achieve a decent channel estimation performance with much less pilot overhead. Moreover, inspired by the fact that MIMO channels are observed to have approximately common support in a neighborhood, an information exchange strategy based on the proposed approach is developed to further improve the estimation accuracy and reduce the required number of pilots through joint channel estimation. Compared with the traditional sparse channel estimation methods, the proposed approach can achieve more than 2dB gain in terms of mean square error (MSE) with the same number of pilots, or

*Corresponding author
Email addresses: ynanbjtu@hotmail.com (Yang Nan), xsun@bjtu.edu.cn (Xin Sun), l.x.zhang@Leeds.ac.uk (Li Zhang)

[1]The authors would like to thank the supports by "the Fundamental Research Funds for the Central Universities (2014YJS006)" and "National Natural Science Foundation of China (61201199)".

achieve the same performance with much less pilots.

Keywords: channel estimation, Massive MIMO OFDM, sparse recovery, complex Homotopy

1. Introduction

The reliable high-speed broadband wireless links are expected to be on a huge development prospective due to the foreseen rapidly increases in the number of users, amount of data traffic and number of applications [1]. To meet these demands, it is expected that future communication systems, i.e., beyond 4G or 5G systems, will reach a faster data rate at gigabit-scale over the next few years. One of the recent proposed technologies for the future communication system is the massive (or large-scale) multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) which is based on the same concepts of the classical MIMO OFDM but with much larger number of antenna arrays on each side of the link. As such, the massive MIMO techniques could provide unprecedented spectral efficiency and array gain that potentially meet the rapidly growing demand for high data rates [2]. For instance, NTT DoCoMo has performed the field experiment of a massive MIMO system with 12×12 antennas, which could reach the spectral efficiency of 50 bps/Hz and a data rate of 4.92 Gbps on the wireless channel with 100 MHz bandwidth [3], while currently a 16×16 MIMO configuration is considered by the evolution of WiFi standard called IEEE 802.11 ac [4].

However, the very large number of channel matrix entries make the traditional channel estimation strategy infeasible for frequency division duplex (FDD) protocol. Therefore, most researches on massive MIMO systems suggest the use of time division duplex (TDD), where the channel state information (CSI) can be acquired at the base station (BS) side and then utilized at both transmission directions based on the assumption of channel reciprocity [5]. However, the inaccurate CSI acquired in the uplink can lead to significant performance degradation. Thus, accurate estimation of the high-dimensional MIMO channel

matrix in the uplink is critical for the deployment of massive MIMO.

Basically, there are two categories of channel estimation techniques for MIMO OFDM systems: blind estimation and pilots-based estimation. In [5], a blind channel estimation technique based on the eigenvalue decomposition (EVD) has been proposed recently. Though this method could approach the near maximum-likelihood performance in theory, it has two shortcomings. Firstly, it utilizes the sample covariance matrix as a substitution of the actual covariance matrix to estimate the channel coefficients. Secondly, it assumes that the number of BS antennas is infinite. For these reasons, the EVD based method suffers from severe mean square error (MSE) performance penalty in practical dimensional massive MIMO systems. Unlike the blind estimation techniques, several pilot-based channel estimation schemes such as the Bayesian MMSE estimator and minimum variance unbiased (MVU) have been adopted in MIMO system [6][7]. However, the pilot overhead of those estimation schemes increases sharply in massive MIMO system where the number of antennas is very large. To solve this problem, some recent works have exploited the sparse nature of the multipath channel and used compressive sensing (CS) based estimators to reconstruct the channel perfectly with relatively less pilots [8][9]. Some CS algorithms, e.g., orthogonal matching pursuit (OMP) and basis pursuit (BP), have been already used in channel estimation for MIMO systems [9, 10, 11, 12]. However, neither of these algorithms can attain accurate estimation performance and low complexity at the same time. More recently, one quadratic semi-definite programming (SDP) method has been discussed in [1], where the author demonstrated that SDP solver can be stable and provide accurate channel estimation, as long as the degree of freedom (DoF) of the channel matrix is much smaller than the size of channel matrix (i.e., total number of elements in the channel matrix). However, experimental results have shown that the convex optimization solver runs slowly in the large-scale applications since it requires explicit operations on the large matrix [13]. Therefore, scholars have abandoned the convex optimization based estimators and turned their attentions to the fast iterative methods.

The Homotopy algorithm has been originally proposed to solve noisy overde-

termined ℓ_1-penalized least squares problems, which becomes popular quickly since it can achieve the estimation accuracy as good as convex optimization schemes with the estimation speed as fast as OMP [12]. Existing literature includes several improvements proposed for the standard Homotopy algorithm [14, 15, 16, 17, 18]. One of the latest researches is the weighted Homotopy in [15], which improves the original Homotopy by replacing its ℓ_1 term with a weighted ℓ_1 term. However, this method can only work in the real domain, which restricts its applications in the complex domain.

In this paper, we propose a set of algorithms for channel estimation in massive MIMO systems. Firstly, we extend the conventional weighted Homotopy to the complex field and adopt it to estimate the sparse channel on each BS antenna. By adjusting the weight separately according to the channel coefficients, this approach improves the channel estimation performance with faster convergence rate. In addition, inspired by the fact that neighboring BS antennas observe similar channel support (i.e., sparsity pattern) [19], we propose a simple information sharing method between the BS antennas to further improve the channel estimation performance. Compared with the conventional sparse channel estimators, the required pilots in the proposed method is significantly reduced, leading to higher spectral efficiency, while its computational complexity is proved to be much lower than the convex optimization solvers.

The rest of the paper proceeds as follows. We first describe the system model and analyse the channel properties in Section 2. Then we propose the improved Homotopy-based channel estimator in Section 3. A simple joint channel estimation method is presented in Section 4 while the performance analysis is provided in Section 5. Numerical experiments are presented in Section 6. Finally, Section 7 concludes the paper.

Notations: Throughout this paper, matrices and vectors are denoted in bold letters while the signals in frequency and time domains are represented by upper and lowercase characters, respectively. $diag\{x\}$ is the diagonal matrix with x on its main diagonal and $\min(\cdot)_+$ means that the minimum is taken over only positive arguments. Operators ^{T} and * represent transpose and complex

conjugate transpose. $|\cdot|$, $\|\cdot\|_p$ and $sgn(\cdot)$ denote absolute value, ℓ_p-norm and sign function respectively. $\boldsymbol{F}_N$, $\mathbb{C}$, $\mathbb{R}$, $\boldsymbol{I}_L$, $\Re$, $\Im$ represent the $N \times N$ normalized discrete Fourier transform (DFT) matrix, the set of complex number, the set of real number, the identity matrix with dimension L, the real part and the imaginary part, respectively.

2. System model and problem formulation

2.1. System model

Considering the uplink of a massive MIMO OFDM system where the BS (receiver) is equipped with a large antenna array consisting of $N_r = M_r \times M_c$ antennas distributed across M_r rows and M_c columns* [19]. The BS simultaneously serves several UTs. Accordingly, for a certain UT the ith OFDM symbol is composed of N_{CP}-length cyclic prefix (CP) and N-length data block $\boldsymbol{x}_i = [x_{i,0}, \ldots, x_{i,1}, \ldots, x_{i,N-1}]^T$ among which N_p positions are randomly collected to transmit the pilots [20]. Let $\boldsymbol{p} = [P_1, \cdots, P_j, \cdots, P_{N_p}]$ $(1 \leq P_1 < \cdots < P_j < \cdots < P_{N_p} \leq N)$ where P_j denotes the index of the jth pilot. Thus the pilots of the ith OFDM symbol can be expressed as $\bar{\boldsymbol{x}}_i = [x_{i,P_1}, x_{i,P_2}, \cdots, x_{i,P_{N_p}}]$.

For the kth receiving antenna at the BS side, the channel impulse response (CIR) $\boldsymbol{h}_i^k (1 \leq k \leq N_r)$ of the ith OFDM symbol can be represented as

$$\boldsymbol{h}_i^k = [h_{i,0}^k, \ldots, h_{i,l}^k, \ldots, h_{i,L-1}^k]^T$$

where $h_{i,l}^k$ is the path gain of the lth path with the path delay $\tau_{i,l}^k$, L is the maximum channel spread. Therefore, the received pilots $\boldsymbol{y}_i^k = [y_{i,P_1}^k, y_{i,P_2}^k, \ldots, y_{i,P_{N_p}}^k]$ at the kth antenna can be denoted in the frequency domain as [20]

$$\begin{aligned} \boldsymbol{y}_i^k &= diag\{\bar{\boldsymbol{x}}_i\} \boldsymbol{F}_{p,L} \boldsymbol{h}_i^k + \boldsymbol{v}_i^k \\ &= \boldsymbol{P}_i^k \boldsymbol{h}_i^k + \boldsymbol{v}_i^k, \end{aligned} \tag{1}$$

*Note that although we focus on the rectangular array configuration in this paper, there is no limitations for our approach to be applied to any other array configurations.

where $diag\{\bar{\boldsymbol{x}}_i^k\}$ is a $N_p \times N_p$ diagonal matrix with $\bar{\boldsymbol{x}}_i^k$ on its diagonal, $\boldsymbol{F}_{p,L}$ is a partial discrete Fourier transform (DFT) matrix indexed by $\boldsymbol{p} = [P_1, P_2, \cdots, P_{N_p}]$ in row and $[1, 2, \cdots, L]$ in column from a standard $N \times N$ DFT matrix, $\boldsymbol{P}_i^k = diag\{\bar{\boldsymbol{x}}_i^k\}\boldsymbol{F}_{p,L}$, $\boldsymbol{v}_i^k$ denotes the additive white Gaussian noise (AWGN). For the sake of brevity, without loss of generality we hereafter omit the script of i and k unless these are required. Hence (1) becomes

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{v}. \tag{2}$$

2.2. Channel properties

Note that extensive literature has demonstrated that the wireless channels are sparse in nature. This indicates that there are only few entries of the channel matrix containing the significant fraction of the channel information while most of the other entries are ignorable [21], [22] (e.g., the ITU Vehicular B channel with 200 samples has only 6 resolvable paths [23]). As such we can draw a conclusion that the number of resolvable propagation paths (or most significant taps) S is much smaller than the maximum channel spread L ($S \ll L$).

Moreover, it is suggested in [24] that two channel taps are resolvable if the time interval of arrival is larger than $\frac{1}{10B}$ where B is the bandwidth of signal. Therefore, it is reasonable to assume that the CIRs measured at different BS antennas share almost the same locations of the significant taps from a certain transmitter if $\frac{d_{max}}{C} \leq \frac{1}{10B}$, where d_{max} is the maximum distance between two BS antennas and C is the speed of light. In other words, the path delays of nonzero elements in CIRs between different transmit-receive pair are identical while the path gains could be distinct, e.g.,

$$supp(\boldsymbol{h}^m) = supp(\boldsymbol{h}^n),\ m \neq n, \tag{3}$$

where $\boldsymbol{h}^m$ is the CIR at the mth BS antenna, and $supp(\boldsymbol{h}^m)$ denotes the support of $\boldsymbol{h}^m$ defined as

$$supp(\boldsymbol{h}^m) = \begin{cases} 1 & h^m(l) \neq 0 \\ 0 & h^m(l) = 0 \end{cases},\ 0 \leq l \leq L-1. \tag{4}$$

In Table 1 we summarize the system parameters of two classical communication standards in terms of the bandwidth B, the maximum resolvable distance D_{max} and the distance between two adjacent antennas $d = \frac{\lambda}{2}$ where λ is the signal wavelength [19]. It is clear that the antenna array could fall in one of the two possible scenarios:

1) The antenna array has the common support when the maximum distance between two antennas $d_{max} \leq D_{max}$. For example, consider a 16×16 array in the 3GPP LTE standard, we have $d_{max} = 15d < D_{max}$. Thus $\boldsymbol{h}^k$, $k = 1, 2, \cdots, N_r$ have the common support. For convenience, we call such arrays common support arrays (CSA).

2) The antenna array has the approximate support when $\frac{d_{max}}{C} > \frac{1}{10B}$. For the BS equipped with a large antenna array, e.g., 32×32 array where $d_{max} = 31d > D_{max}$ in the 3GPP LTE standard, the channel support varies across the array but with slow rate. We call such arrays the approximate support array (ASA).

Table 1: Parameters of communication systems

Standard	Bandwidth(B)	$D_{max} = \frac{C}{10B}$	$d = \frac{\lambda}{2}$
CDMA2000	1.25 MHz	24 m	0.15m
3GPP LTE	20 MHz	1.5 m	0.058m

While most of the recent literature in massive MIMO channel estimation considers the common support case [25, 26], little attention has been paid to the approximate support scenario. In this paper, we propose the approach that is applicable to both of the CSA and ASA cases with two steps:

1) Weighted Homotopy based channel estimation at each antenna, and

2) Joint channel estimation.

3. Weighted Homotopy based channel estimation

The set of channel estimation approaches that we propose in this paper use a modified version of the Homotopy algorithm proposed by the authors in [12] and

[15]. In this section, we first describe the standard Homotopy briefly, and then propose our improved algorithm. Note that in this section the proposed channel estimation algorithm is utilized at each BS antenna, where no cooperation occurs among the antennas. The joint channel estimation strategy which considers both of the CSA and ASA cases will be issued in the next section.

3.1. Standard Homotopy

To estimate the CIR $\boldsymbol{h}$ in (2), we first introduce the famous convex optimization model known as the Least Absolute Shrinkage and Selection Operator (LASSO) or Basis Pursuit De-Noising (BPDN) as follows [27]:

$$\min_{\boldsymbol{x}} \ \tau\|\boldsymbol{x}\|_1 + \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{5}$$

where $\boldsymbol{y} \in \mathbb{R}^{N_p \times 1}$ is the measurement vector, $\boldsymbol{x} \in \mathbb{R}^{L \times 1}$ is the unknown signal of interest and $\boldsymbol{A} \in \mathbb{R}^{N_p \times L}$ is a known matrix. The ℓ_1 term in (8) limits the sparsity of the solution while the ℓ_2 term keeps the solution close to the measurements. $\tau > 0$ is a user-selected regularization parameter that transfers ℓ_1 minimization problem to ℓ_2 minimization problem. The LASSO or BPDN has been demonstrated to yield good performances in a variety of practical applications [11][14][23]. One of the neoteric approaches for solving (5) is the Homotopy algorithm, which traces a solution path for a range of decreasing values of τ, and terminates when τ converges to some threshold η [28]. The solution path is followed by maintaining the optimality condition of (5) at each point along the path.

Assume $\tilde{\boldsymbol{x}}$ is the solution to (5). The Homotopy algorithm starts at $\tilde{\boldsymbol{x}} = 0$ and τ a large value which shrinks toward a threshold η in a sequence of computationally inexpensive steps. At the same time, $\tilde{\boldsymbol{x}}$ follows a piecewise-linear path and converges to the solution of a noiseless sparse recovery problem, denoted as

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \ s.t. \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{6}$$

In every Homotopy step, $\tilde{\boldsymbol{x}}$ is updated according to two parameters: the step-size and the update direction, both of which are determined by the support and

sign sequence of $\tilde{\boldsymbol{x}}$. The support of $\tilde{\boldsymbol{x}}$ changes only at certain critical values of τ, when either a new nonzero element enters the support or an existing nonzero element shrinks to zero. For every Homotopy step, we jump from one critical value of τ to the next while updating the support of the solution, until τ is reduced to its desired value. Specifically, let $f(\mathbf{x})$ denote the objective function of (5), the smaller critical value of τ can be easily obtained by calculating the following equation [29]:

$$\partial f(\boldsymbol{x}) = \boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + \tau\partial\|\boldsymbol{x}\|_1, \tag{7}$$

where $\partial\|\boldsymbol{x}\|_1$ is the subgradient of $\|\boldsymbol{x}\|_1$ obtained by

$$\partial\|\boldsymbol{x}\|_1 = \left\{\boldsymbol{z} \in \mathbb{R}^L \middle| \begin{array}{ll} z_i = sgn(x_i), & x_i \neq 0 \\ z_i \in [-1, 1], & x_i = 0 \end{array}\right\},$$

where x_i and z_i account for the ith component of $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. It is clear that a necessary condition for $\tilde{\boldsymbol{x}}$ to be the optimal value of (5) is that

$$\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + \tau\partial\|\boldsymbol{x}\|_1 = 0. \tag{8}$$

Thus, let $\Gamma = \{i|\tilde{x_i} \neq 0\}$ be the active support of $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_\Gamma$ be the vector with the components of $\tilde{\boldsymbol{x}}$ on support Γ and $\boldsymbol{z} = sgn(\tilde{\boldsymbol{x}}_\Gamma)$. At any given value of τ, we have

$$\begin{cases} \boldsymbol{A}_\Gamma^*(\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y}) = -\tau\boldsymbol{z} & \text{(9a)} \\ \|\boldsymbol{A}_{\Gamma^c}^*(\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y})\|_\infty \leq \tau & \text{(9b)} \end{cases}$$

where $\boldsymbol{A}_\Gamma$ denotes a sub-matrix of $\boldsymbol{A}$ collecting the column vectors of $\boldsymbol{A}$ in the active support Γ, and $\boldsymbol{A}_{\Gamma^c}$ represents the sub-matrix collecting the column vectors in the inactive support Γ^c where $\Gamma^c = \{i|\tilde{x_i} = 0\}$. When we reduce τ by a small value of δ, the new solution moves in a direction $\partial\boldsymbol{x}$ as

$$\begin{cases} \boldsymbol{A}_\Gamma^*[\boldsymbol{A}(\tilde{\boldsymbol{x}} + \delta\partial\boldsymbol{x}) - \boldsymbol{y}] = -(\tau - \delta)\boldsymbol{z} \\ \|\boldsymbol{A}_{\Gamma^c}^*[\boldsymbol{A}(\tilde{\boldsymbol{x}} + \delta\partial\boldsymbol{x}) - \boldsymbol{y}]\|_\infty \leq (\tau - \delta) \end{cases}, \tag{10}$$

and then we have

$$\partial\boldsymbol{x} = \begin{cases} (\boldsymbol{A}_\Gamma^*\boldsymbol{A}_\Gamma)^{-1}\boldsymbol{z} & on\ \Gamma \\ 0 & otherwise \end{cases}. \tag{11}$$

The solution will continue to move along the direction $\partial \boldsymbol{x}$ until one of the following two cases happens: (i) an element of $\tilde{\boldsymbol{x}}$ shrinks to zero, which violates (9a) and indicates this element should be removed from Γ, or (ii) an element in Γ^c increases in magnitude beyond τ which violates (9b), and thus this element should be added to Γ. Assume $\boldsymbol{p} = \boldsymbol{A}_{\Gamma^c}^*(\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y})$, $\boldsymbol{d} = \boldsymbol{A}_{\Gamma^c}^* \boldsymbol{A}\partial \boldsymbol{x}$, δ^- and δ^+ are the step-sizes in cases (i) and (ii), respectively, then we can obtain the step-sizes according to the following equations:

$$\begin{cases} \tilde{x}_i + \delta^- \partial x_i = 0, \ i \in \Gamma & \text{(12a)} \\ |p_i + \delta^+ d_i| = \tau - \delta^+ & \text{(12b)} \end{cases}$$

where p_i and d_i are the ith element of $\boldsymbol{p}$ and $\boldsymbol{d}$, $\tilde{x}_i$ and ∂x_i are the ith element of $\tilde{\boldsymbol{x}}$ and the direction $\partial \boldsymbol{x}$, respectively. Note that once an estimate $\tilde{x}_i$ decreases to zero, it should be expelled from the active support, which explains why (12a) is used as a constraint condition. Therefore, we have

$$\delta^- = \min_{i \in \Gamma} \left(\frac{-\tilde{x}_i}{\partial x_i} \right)_+ . \tag{13a}$$

Next, by solving (12b), we get the step-size δ^+ as

$$\delta^+ = \min_{i \in \Gamma^c} \left(\frac{\tau - p_i}{1 + d_i}, \frac{-\tau - p_i}{-1 + d_i} \right)_+ , \tag{13b}$$

The smallest step-size is obtained by selecting the smaller step-sizes as

$$\delta = \min(\delta^+, \delta^-). \tag{14}$$

After that we get the new critical value of τ as $\tau = \tau - \delta$ and the new channel estimate $\tilde{\boldsymbol{x}}$ as $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}} + \delta \partial \boldsymbol{x}$. We repeat the above steps until τ decreases to a desired threshold η.

3.2. Weighted Homotopy

To improve the recovery performance, it naturally leads us to substitute the ℓ_1 norm or ℓ_2 norm in (5) for a weighted norm (see more details for the weighted ℓ_2 norm case in [16]-[18]). According to the system model discussed in Section 2, the weighted ℓ_1 norm minimization problem could be described as

$$\min_{\boldsymbol{h}} \ \sum_{i=0}^{L-1} w_i |h_i| + \frac{1}{2} \|\boldsymbol{P}\boldsymbol{h} - \boldsymbol{y}\|_2^2 \tag{15}$$

where $\boldsymbol{y} \in \mathbb{C}^{N_p \times 1}$ is the received signal vector, $\boldsymbol{h} \in \mathbb{C}^{L \times 1}$ is the unknown CIR with h_i the ith element, $\boldsymbol{P} \in \mathbb{C}^{N_p \times L}$ is the partial DFT matrix, and w_i accounts for the weight of h_i.

Note that all the quantities in (15) are complex-valued, except for w_i a positive real number. Thus, we first prove that the optimality conditions of the complex weighted Homotopy could be still satisfied.

Lemma: $\boldsymbol{h}$ is the minimum of (15) if and only if there exists a complex vector $\boldsymbol{z}$ which is the subgradient of $|\boldsymbol{h}|$, satisfying $z_j = \frac{h_j}{|h_j|}$ if $j \in \{j|h_j \neq 0\}$, *and* $|z_j| \leq 1$ *if* $j \in \{j|h_j = 0\}$, such that

$$\boldsymbol{w}\boldsymbol{z} + \boldsymbol{P}^*(\boldsymbol{P}\boldsymbol{h} - \boldsymbol{Y}) = 0, \tag{16}$$

where $\boldsymbol{w} = [w_0, w_1, \cdots, w_{L-1}]$.

Proof: Denoting $f(\boldsymbol{h}) \triangleq \boldsymbol{w}|\boldsymbol{h}| + \frac{1}{2}\|\boldsymbol{P}\boldsymbol{h} - \boldsymbol{Y}\|_2^2$, which is the objective function of (15) to be minimized. First we declare that, for $j \in \{j|h_j \neq 0\}$, (also expressed as $j \in act$), we have

$$|h_j + \delta h_j| \simeq |h_j| + \frac{\Re(h_j^* \delta h_j)}{|h_j|},$$

where δh_j is a small complex number. Also, for $j \in \{j|h_j = 0\}$ ($j \in inact$), we have

$$|h_j + \delta h_j| = |\delta h_j|.$$

Then by defining $\boldsymbol{c} \triangleq \boldsymbol{P}^*(\boldsymbol{P}\boldsymbol{h} - \boldsymbol{Y})$, successively we have

$$\begin{aligned} & f(\boldsymbol{h} + \delta\boldsymbol{h}) \\ & \simeq f(\boldsymbol{h}) + \frac{1}{2}(\delta\boldsymbol{h}^*\boldsymbol{c} + \boldsymbol{c}^*\delta\boldsymbol{h}) + \sum_{j \in act} \frac{w_j \Re(h_j^* \delta h_j)}{|h_j|} + \sum_{j \in inact} w_j |\delta h_j| \\ & \simeq f(\boldsymbol{h}) + \Re(\boldsymbol{c}^*\delta\boldsymbol{h}) + \sum_{j \in act} \frac{w_j \Re(h_j^* \delta h_j)}{|h_j|} + \sum_{j \in inact} w_j |\delta h_j|. \end{aligned} \tag{17}$$

A necessary condition for $\boldsymbol{h}$ to minimize $f(\boldsymbol{h})$ is $\partial f(\boldsymbol{h}) = 0$, where $\partial f(\boldsymbol{h})$ is the derivative of $f(\boldsymbol{h})$. Thus we have $\boldsymbol{c} + \boldsymbol{w}\boldsymbol{z} = 0$. Then we can rewright the term $\Re(\boldsymbol{c}^*\delta\boldsymbol{h})$ in (17) as the sum of two parts

$$\begin{aligned} \Re(\boldsymbol{c}^*\delta\boldsymbol{h}) &= -\Re[(\boldsymbol{w}\boldsymbol{z})^*\delta\boldsymbol{h}] \\ &= -\sum_{j \in act} \frac{w_j \Re(h_j^* \delta h_j)}{|h_j|} - \sum_{j \in inact} w_j \Re(z_j^* \delta h_j). \end{aligned} \tag{18}$$

Substituting (18) into (17), we have

$$
\begin{aligned}
&f(\boldsymbol{h}+\delta\boldsymbol{h}) \\
&\simeq f(\boldsymbol{h}) + \sum_{j\in act} w_j(-\frac{\Re(h_j^*\delta h_j)}{|h_j|} + \frac{\Re(h_j^*\delta h_j)}{|h_j|}) \\
&+ \sum_{j\in inact} w_j(-\Re(z_j^*\delta h_j) + |\delta h_j|) \\
&\simeq f(\boldsymbol{h}) + \sum_{j\in inact} w_j[1-|z_j|\cos(z_j,\delta h_j)]|\delta h_j|.
\end{aligned}
$$

It is obvious that $1-|z_j|\cos(\ldots) \geq 0$ if $|z_j| \leq 1$, then we can draw a conclusion that $f(\boldsymbol{h}+\delta\boldsymbol{h}) \geq f(\boldsymbol{h})$, and vice-versa.

From above discussion, we can see that the optimality conditions in (16) are thus identical to (8) but with independent weight w_i instead of τ, and the subgradient $\partial|\boldsymbol{h}|$ obtained by

$$
\partial|\boldsymbol{h}| = \left\{ \boldsymbol{z}\in\mathbb{C}^L \middle| \begin{array}{lll} z_j = \frac{h_j}{|h_j|} & , & h_j \neq 0 \\ |z_j| \leq 1 & , & h_j = 0 \end{array} \right\}.
$$

Thus we can urge the approximate CIR $\tilde{\boldsymbol{h}}$ to converge towards the optimal solution of (15) by decreasing the weights $\boldsymbol{w}$ separately as long as the optimality condition is satisfied.

Unlike the traditional Homotopy where τ controls the convergence speed of all the estimates $\boldsymbol{h}$ on a same level, in the proposed method the independent weight w_i adjusts each estimate h_i separately. To be specific, we let the weights on the active support shrink at a faster rate, while the weights elsewhere shrink at a slower rate. Thus, the estimates on the active support will be more likely to remain nonzero while the estimates on inactive support are driven to be zero, whereby the proposed method could achieve more accurate estimation and faster covergence rate.

In details, we take large values for all the weights w_i as $w_i = \max_j |\mathcal{P}_j^*\boldsymbol{y}|$, $i = 0, 1, \ldots, L-1$ at the beginning, where $\mathcal{P}_j$ denotes the jth column in $\boldsymbol{P}$. Then we divide these indices w_i into the active set and the inactive set according to the support of solution, e.g. we put w_i into the active set if $i \in \Gamma$, otherwise

we put w_i into the inactive set. Some desired values of w_i are set at each Homotopy step, to guarantee a faster reducing rate of w_i in the active set and a lower reducing rate of w_i in the inactive set. Assume $\tilde{w}_i$ is the desired value of w_i, we set $\tilde{w}_i$ in the active set as $\tilde{w}_i = \frac{w_i}{\kappa|\tilde{h}_i|}$ where $\tilde{h}_i$ is the solution of the last iteration, and $\kappa = N_r \frac{\|\tilde{h}\|_2^2}{\|\tilde{h}\|_1^2}$ is a proportional constant [15]. It is advisable to define κ in this manner because we can adjust the value of w_i according to the solution $\tilde{\boldsymbol{h}}$ where the weight w_i decreases to smaller value when N_r is large or larger value when $\tilde{\boldsymbol{h}}$ is denser. For the $\tilde{w}_i$ in the inactive set, we can select it in a variety of ways, e.g., we can set each of $\tilde{w}_i$ in the inactive set equal to $\max_{j\in\Gamma} \tilde{w}_j$.

Next, by comparing (15) with (5), we can rewrite the constraint conditions in (10) as follows:

$$\begin{cases} \mathcal{P}_i^*(\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y}) = -w_i z_i, \ for\ all\ i \in \Gamma & \text{(19a)} \\ |\mathcal{P}_i^*(\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y})| \le w_i, \quad for\ all\ i \in \Gamma^c & \text{(19b)} \end{cases}$$

Define $p_i = \mathcal{P}_i(\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y})$ and $d_i = \mathcal{P}_i \boldsymbol{P}\partial\boldsymbol{h}$. As we decrease w_i toward the desired weight $\tilde{w}_i$, the solution $\tilde{\boldsymbol{h}}$ moves in a direction $\partial\boldsymbol{h}$ as $\tilde{\boldsymbol{h}} + \delta\partial\boldsymbol{h}$, which must obey the constraint conditions

$$\begin{cases} \boldsymbol{P}_\Gamma^*(\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y}) + \delta^- \boldsymbol{P}_\Gamma^* \boldsymbol{P}\partial\boldsymbol{h} = -\boldsymbol{W}\boldsymbol{z} + \delta^-(\boldsymbol{W} - \tilde{\boldsymbol{W}})\boldsymbol{z} & \text{(20a)} \\ |p_i + \delta^+ d_i| \le w_i - \delta^+(w_i - \tilde{w}_i), \ i \in \Gamma^c & \text{(20b)} \end{cases}$$

where $\boldsymbol{z} = \frac{\tilde{\boldsymbol{h}}_\Gamma}{|\tilde{\boldsymbol{h}}_\Gamma|}$, $\tilde{\boldsymbol{h}}_\Gamma$ is the vector with the components of $\tilde{\boldsymbol{h}}$ on support Γ, $\boldsymbol{W}$ and $\tilde{\boldsymbol{W}}$ denote the $|\Gamma| \times |\Gamma|$ diagonal matrices whose elements of the primary diagonal are the values of w_i and $\tilde{w}_i$ on Γ, respectively.

Note that $\tilde{h}_i$ is a complex value. We can see from (12a) that the breakpoint only occurs when both of the real part and imaginary part of $\tilde{h}_i$ ($i \in \Gamma$) shrink to zero. In other words, if

$$\frac{\Re(\tilde{h}_i)}{\Re(\partial h_i)} = \frac{\Im(\tilde{h}_i)}{\Im(\partial h_i)}, \ i \in \Gamma \tag{21}$$

we have

$$\delta^- = \min_{i\in\Gamma} \left(\frac{-\Re(\tilde{h}_i)}{\Re(\partial h_i)} \right)_+ . \tag{22a}$$

Thus, from (20b) we have

$$\delta^+ = \begin{cases} \min\limits_{i\in\Gamma^c} \left(\frac{w_i^2 - |p_i|^2}{2[\Re(p_i^* d_i) - s_i w_i]} \right), & |d_i| = 1 \\ \min\limits_{i\in\Gamma^c} \left(\frac{s_i w_i - \Re(p_i^* d_i) - v_i}{|d_i|^2 - s_i^2}, \frac{s_i w_i - \Re(p_i^* d_i) + v_i}{|d_i|^2 - s_i^2} \right), & |d_i| \neq 1 \end{cases} \tag{22b}$$

where

$$s_i = \tilde{w}_i - w_i,$$

$$v_i = \sqrt{[s_i w_i - \Re(p_i^* d_i)]^2 - (s_i^2 - |d_i|^2)(w_i^2 - |p_i|^2)}.$$

Combining (19a) and (20a), we can get the update direction $\partial \boldsymbol{h}$ as:

$$\partial \boldsymbol{h} = \begin{cases} (\boldsymbol{P}_\Gamma^* \boldsymbol{P}_\Gamma)^{-1} (\boldsymbol{W} - \tilde{\boldsymbol{W}}) \boldsymbol{z} & on\ \Gamma \\ 0 & on\ \Gamma^c \end{cases}. \tag{23}$$

This process continues until all the weights $\boldsymbol{w}$ drop below a certain threshold η.

Observing (21), we find that the condition in (21) is hard to be satisfied since it only occurs when both of the real part and imaginary part of $\tilde{h}_i$ shrink to zero simultaneously. That is to say, it is less likely to remove an element from Γ than adding a new element into Γ. Meanwhile, the optimality condition in (20b) indicates that as long as the inequality is satisfied, we can change the corresponding weight in the inactive set to an arbitrary value while the solution to (15) is still optimal. Therefore, we alter the selecting rule of δ in (14) as follows. Suppose δ^- makes an element $\tilde{h}_{\gamma^-}$ ($\gamma^- \in \Gamma$) decrease to zero. If $\delta^- < 1$, then we remove γ^- from Γ and set $\delta = \delta^-$. Otherwise, we assume that $\boldsymbol{w}$ has turned to $\tilde{\boldsymbol{w}}$ and set $\delta = \delta^+ = 1$. After that we select an index γ^+ as

$$\gamma^+ = \arg\max\nolimits_{i\in\Gamma^c} |\mathcal{P}_i^*(\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y})|, \tag{24}$$

and then add γ^+ into Γ. In order to satisfy the inequality in (20b), we should also change the choice strategy of the desired weight $\tilde{w}_i$ in the inactive set. By substituting $\delta^+ = 1$ into (20b), we have $\tilde{w}_i \geq |p_i + d_i|$ for $i \in \Gamma^c$. Therefore, we set the $\tilde{w}_i$ for $i \in \Gamma^c$ either equal to $|p_i + d_i|$ or equal to $\max\limits_{j\in\Gamma} \tilde{w}_j$, whichever is larger. For convenience, we name the proposed weighted Homotopy as WH

Algorithm 1

Input: $\boldsymbol{P}$, $\boldsymbol{y}$ and η

Output: $\tilde{\boldsymbol{h}}$

Step 1. Initialization:

$\tilde{\boldsymbol{h}} \leftarrow 0, w_i \leftarrow \max_j |\mathcal{P}_j^* \boldsymbol{y}|$ *for all* i, $\Gamma \leftarrow \arg\max_j |\mathcal{P}_j^* \boldsymbol{y}|$

Step 2.

While $\max_i(w_i) > \eta$ do

1) Determine the desired weights value: $\tilde{w}_i = \frac{w_i}{\kappa |\tilde{h}_i|}$ for $i \in \Gamma$, and $\tilde{w}_i \leftarrow \max(|p_i + d_i|, \max_{j \in \Gamma} \tilde{w}_j)$ for $i \in \Gamma^c$

2) Compute δ^- in (22a)

3) Compute $\partial \boldsymbol{h}$ in (23)

4) Compute $\delta = \min(1, \delta^-)$

5) $\tilde{\boldsymbol{h}} \leftarrow \tilde{\boldsymbol{h}} + \delta \partial \boldsymbol{h}$

6) $w_i \leftarrow w_i + \delta(\tilde{w}_i - w_i)$

7) if $\delta^- < 1$

$\Gamma \leftarrow \Gamma / \gamma^-$ (removing γ^- from Γ)

Else

$\gamma^+ = \arg\max_{j \in \Gamma^c} |\mathcal{P}_j^* (\boldsymbol{P}\tilde{\boldsymbol{h}} - \boldsymbol{y})|$

$\Gamma \leftarrow \Gamma \cup \gamma^+$ (adding γ^+ into Γ)

End if

End while

method and present the pseudocode in Algorithm 1.

Note that the WH method enhances the channel estimation performance with faster convergence rate by adjusting the weights separately according to the channel coefficients. To make further improvement, the common support of MIMO channels will be exploited and incorporated with WH method.

4. Joint channel estimation

In the joint channel estimation method, the BS antennas collaborate with each other to take advantage of the common support (approximate support) of the MIMO channels. Since this collaboration method is realized at the baseband, an additional processor on the baseband card is necessary [19].

The core idea of the collaboration strategy is that the antennas within a "safety zone" share the information of the estimated channel support with each other to improve the channel estimation performance. A safety zone is part of the integrated antenna array where common support property is satisfied. In other words, the distance between any two antennas within the safety zone is no larger than $\frac{C}{10B}$. Therefore, it is safe to assume these antennas share the common support of the CIRs. As a result, antennas in the same safety zone are able to incorporate information from their neighbors to enhance its decision about the support of the CIR, which will ultimately improve the channel estimation. Since we consider the rectangular array configuration in this paper, the "safety zone" is defined as a square containing G antennas. In Fig. 1 we use a 8×8 array with 4 safety zones as an example. Clearly, this strategy has two distinctive features:

1) It is flexible to change the size of safety zone according to the signal bandwidth and antenna interval, which is critical to achieve the ASA.

2) The communication overhead as well as the computational complexity are reduced since there is no necessary to incorporate all the antennas for cooperation.

Based on this information sharing strategy, we propose the joint channel

Figure 1: A 8×8 array with 4 safety zones.

estimation as follows. To estimate the MIMO channels cooperatively, we use the weights of the WH method as the shared information. Note that identical sharing strategy will be adopted within each safety zone, so we only present the approach in a certain safety zone as an example. Assume there are G antennas in the safety zone. We start by estimating the CIR using the WH method on each antenna, respectively. During each WH iteration, we record the weights $\boldsymbol{w}^k$ for $k = 1, 2, \cdots, G$ and calculate the averages $\boldsymbol{a}$ as

$$\boldsymbol{a} = \sum_{k=1}^{G} \boldsymbol{w}^k / G. \tag{25}$$

Let $\boldsymbol{w}_\Gamma^k$ denote the elements of $\boldsymbol{w}^k$ on the active support Γ and $\boldsymbol{w}_{\Gamma^c}^k$ be the elements of $\boldsymbol{w}^k$ on the inactive support. Note that $\boldsymbol{w}_\Gamma^k \leq \boldsymbol{w}_{\Gamma^c}^k$ since $\boldsymbol{w}_\Gamma^k$ are always reduced at a faster rate than $\boldsymbol{w}_{\Gamma^c}^k$, according to the WH method proposed in Section 3.2. We assign Λ as the estimated channel support, collecting the

indices of the S minimum values in $\boldsymbol{a}^*$. After that we define a complementary set $\varpi = \Gamma - \Lambda$ that collects the indices which do not belong to the channel support but are added to the active support accidentally. Then we define a factor ν as

$$\nu = \frac{\max\limits_{i \in \varpi}(\boldsymbol{w}_i^k)}{\max\limits_{i \in \Gamma^c}(\boldsymbol{w}_i^k)}. \tag{26}$$

Thus, we can refine the desired weights as

$$\tilde{w}_i^k = \begin{cases} \tilde{w}_i^k \nu, for\ i \in \Lambda \\ \tilde{w}_i^k / \nu, for\ i \in \varpi \end{cases}. \tag{27}$$

Note that we only adjust the desired weights in Γ and increase the $\tilde{\boldsymbol{w}}_{\varpi}^k$ not larger than the maxmimum of $\tilde{\boldsymbol{w}}_{\Gamma^c}^k$ due to the choice of ν. Therefore this modification will not violate the optimality conditions. Thus, in this fashion, we can further reduce the weights on the set Λ and enlarge the weights on the set ϖ while maintain the solution to be optimal. As a result, the elements of $\tilde{\boldsymbol{h}}$ on Λ will be encouraged to remain nonzero while the elements elsewhere are driven to zero. This estimate will be more accurate since the antennas have shared their information to strengthen their beliefs about the estimated support Λ. Note that this information sharing method can be regarded as an auxiliary algorithm which should be included after Line 5 of Algorithm 1. The details of this information sharing strategy is presented in Algorithm 2.

Algorithm 2

1) At every iteration, record the weights of WH method $\boldsymbol{w}^k$, $k = 1, 2, \cdots, G$.

2) Compute the average weight $\boldsymbol{a}$ by (25).

3) Refine the the desired weights of each antenna according to (26) and (27).

It is worth noting that the proposed information sharing method is independent of the antenna array configuration. Since each antenna only utilizes the

*The expected number of the active taps in the CIR can be detected by the iterative support detection (ISD) method. For details, readers are referred to [30].

knowledge of their neighbors in the same safety zone, this method could improve the channel estimation performance as long as the antennas are located in the right safety zone, no matter what the topology of the antenna array. By exploiting the common support of channel in WH method, the performance is clearly improved.

5. performance analysis

5.1. Boundary analysis

In the CSA case, since the channel support is assumed to be identical cross the array, contribution from as many antennas as possible will always strengthen the belief in the estimated support Λ. Hence, we may choose $G = N_r$ which indicates that all the BS antennas are included in one safety zone and share channel information with each other. However, it might be unnecessary to collect such a large number of antennas in one safety zone. In [19], the authors have proposed a loose lower bound based on the lemma 1 in [31], where a relationship between the number of antennas G, the sparsity of channel S and the number of required pilots N_p is established as

$$S \leq \lceil (N_p + G)/2 - 1 \rceil, \tag{28}$$

where $\lceil \cdot \rceil$ is the ceiling operation. Thus we can obtain the lower bound of G accordingly as

$$G > 2S - N_p. \tag{29}$$

In the ASA case, the contribution of information sharing is based on how fast the channel support change over the array. Clearly, if the support varies fast, large number of G will lead to performance deterioration. Therefore, we should select the number of antennas G that satisfies the CSA according to the former discussion in Section 2.2. In other words, we could still get benefits from information sharing by selecting a suitable G for the ASA case. Note that the safety zone is considered as a square configuration in this paper. Therefore, the distance between the farthest antennas in a safety zone with G antennas is

$(\sqrt{G}-1)d$. Therefore, to ensure the common support property in a safety zone, G should satisfy the following inequation,

$$\begin{aligned}(\sqrt{G}-1)d &\leq \frac{C}{10B},\\ G &\leq \lfloor(\frac{C}{10Bd}+1)^2\rfloor,\end{aligned} \tag{30}$$

where $\lfloor\cdot\rfloor$ is the floor operation.

5.2. Complexity analysis

From [15] we know that the computational cost for one iteration of the standard Homotopy method is about $GL + GQ + 3Q^2 + O(L)$ FLOPs where Q is the size of support Γ. By observing the proposed algorithm, we find the proposed Homotopy algorithm requires roughly the same number of FLOPs as the standard Homotopy algorithm plus additional computations to calculate the update direction $\partial\boldsymbol{h}$ in (23) where the weight matrix $(\boldsymbol{W}-\tilde{\boldsymbol{W}})$ is considered. Since $(\boldsymbol{W}-\tilde{\boldsymbol{W}})$ is a $Q\times Q$ diagonal matrix, the additional cost for computing $\partial\boldsymbol{h}$ involves nearly Q^2 FLOPs*. As such, the computational complexity of one iteration of the proposed Homotopy method is about $GL + GQ + 4Q^2 + O(L)$ FLOPs, which is a little higher than that of the standard Homotopy method. However, due to the use of weights and information sharing method, the proposed algorithm converges much faster than the standard Homotopy. Therefore, the proposed algorithm requires fewer iterations, and thus less complexity (more details will be shown in Section 6). Besides, in terms of the asymptotic complexity, the standard Homotopy is roughly on the same order as OMP with $O(G^2L)$ while the asymptotic complexity of convex optimization algorithm is about $O(L^3)$[13]. Obviously, the complexity of the Homotopy method is less than the convex optimization methods as long as the wireless channel is sparse.

*For the sake of brevity, we neglect small complexity of scalar multiplications and additions of matrices and vectors.

6. Simulation results

In this section, simulation studies are carried out to compare the performance of the proposed method with Homotopy[†] and two other CS algorithms: YALL1 [11] and OMP. YALL1 is a convex optimization algorithm which is proved to be the best compared with two other famous BP algorithms including ℓ_1-LS and SpaRSA in [11], and OMP is widely used due to its fast convergence speed and good estimation accuracy [5]. The threshold η for both of the standard Homotopy and proposed Homotopy is 0.01 [15]. All the simulations in this paper are performed using MATLAB 2012a, running on a standard computer with an Intel Core i3-2100 CPU at 3.10GHz and 4GB of memory. A 16×16 MIMO configuration is considered with the signal bandwidth of 20MHz at the central frequency of 2.6GHz, as specified in the 3GPP-LTE standard. The number of OFDM subcarriers N is 4096, and the length of CP (N_{CP}) is 256 which could combat channels with the maximum multi-path delay spread of $12.8\mu s$. The sparse Rayleigh channels with $S = 6$ significant taps and the maximum delay spread of $10\mu s$ is considered. The IlmProp channel modeling tool [32] is employed for the channel generations of both of the CSA and ASA scenarios. Specifically, we place point-like scatterers and the UT randomly and obstruct the line-of-sight [19]. To make sure the desired sparsity S, we set the number of scatterers accordingly and eliminate the small non-zero components in the resulting CIR while maintain the top S components. Moreover, to generate the CSA and ASA behavior, we set the antenna spacing $d = 0.058m$ for CSA to satisfy the common support assumption while set the antenna spacing $d = \frac{C}{20B} = 0.75m$ for ASA to ensure that the central antenna and its 8-neighbors have approximately common support while the channel support of other antennas may change slowly.

Firstly, we set $G = 9$ and compare the success rate of the WH method with other different channel estimators when a varying number of pilots N_p is used

[†]We have extended the standard Homotopy to the complex field in the similar way as the proposed approach.

in Fig.2. The top row of Fig.2 shows the success rate of channel recovery when the signal-to-noise ratio (SNR) $SNR = 10dB$ while the bottom row shows the success rate comparison when $SNR = 20dB$. The success rate is defined as the ratio of the number of success trails to the number of total trails, where a trail is recognized to be successful as the MSE is better than 10^{-2}. It is obvious that all of the channel estimators require more pilots to achieve the MSE boundary of 10^{-2} when SNR decreases from 20dB to 10dB. However, when the same number of pilots is used, WH method always obtains higher success rate in both of the CSA and ASA scenarios. For example, In Fig.2(a) where $SNR = 10dB$, the WH method uses only 60 pilots to achieve a success rate over 50% and 76 pilots to achieve 100% success rate. On the contrary, the traditional algorithms use at least 68 pilots for the success rate over 50%, and none of them can achieve the 100% success rate due to the limitation of the SNR and the number of pilots. Besides, for the ASA case as shown in Fig.2(b) and Fig.2(d), although the WH performs not as well as it does in the CSA scenario, it still outperforms the traditional methods. Specifically, when $N_p = 60$ is considered in Fig.2(b), the WH achieves a success rate of 48% while the highest success rate the traditional methods could attain is 27%.

Next, we investigate in Fig.3 the MSE performance comparison between the proposed WH scheme and conventional OMP, YALL1 as well as the Homotopy scheme. In this experiment, we use $G = 9$ and $N_p = 32$ pilots to estimate the channel. It can be seen from the figure that for CSA the WH method outperforms other three schemes by more than 2 dB when the target MSE of 10^{-2} is considered. To be specific, when the MSE is 10^{-2}, OMP requires the SNR of 24.8 dB, YALL1 and Homotopy require the SNR of 21.9 dB and 19.7 dB, respectively, while the improved Homotopy needs only a SNR of 17.1 dB. It is also worth noting that the WH method performs equally well for both CSA and ASA, which demonstrates its robustness in the support-variant scenario.

Besides achieving the optimum MSE performance, another strength of the WH method is its fast convergence. Next, we set $G = 1$ and compare the convergence speed of the WH method and Homotopy in Fig.4(a) (for $SNR = 15dB$)

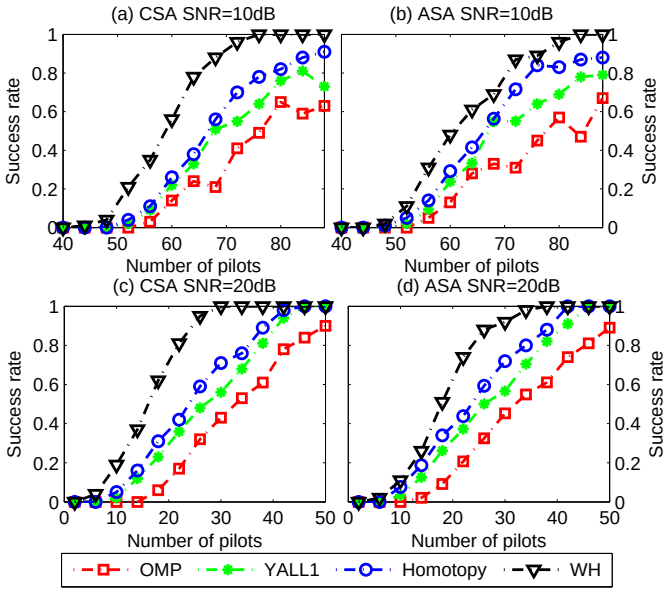

Figure 2: Success rate comparison between different algorithms when a varying number of pilots is used.

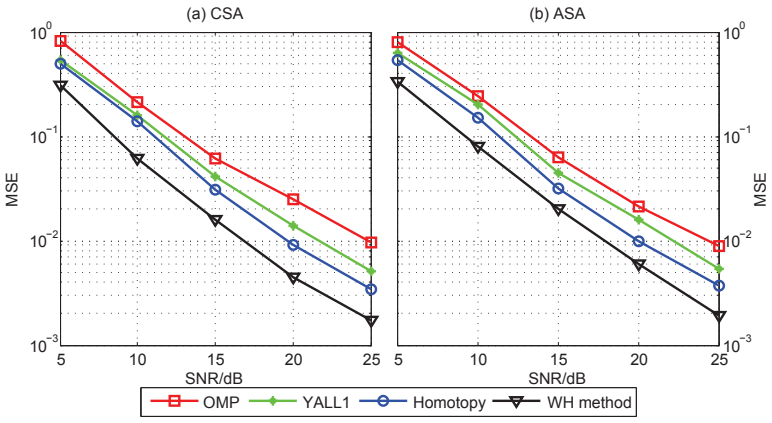

Figure 3: MSE comparison between the WH method and conventional schemes.

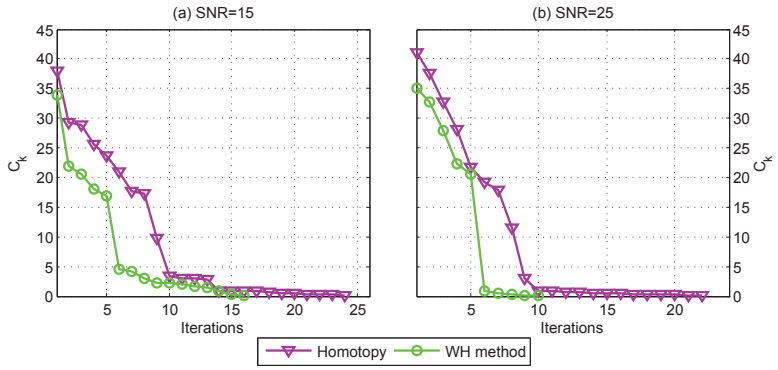

Figure 4: Convergence speed comparison of the Homotopy and the WH method for different SNR values.

Table 2: Running times of different algorithms

Algorithms	OMP	YALL1	Homotopy	WH method
CPU time (in s)	0.1873	1.4751	0.3177	0.2246

and Fig.4(b) (for $SNR = 25dB$), where $C_k = \tau$ for the Homotopy and $C_k = \max_i w_i$ for the proposed method are used as ordinate while the horizontal axis is denoted by the number of iterations. The threshold of C_k in this simulation is set to 0.01. We can see from both of Fig.4(a) and Fig.4(b) that comparing with Homotopy, the proposed method converges to the threshold more quickly. For example, in Fig.4(b) the C_k in proposed method decreases significantly faster than the standard Homotopy method and achieves the threshold 0.01 after 10 iterations, while the number of iterations required by Homotopy is 22. This can be explained by the cooperation of information sharing strategy and the weighted ℓ_1 term which democratically penalize the nonzero coefficients, and in turn reduce the total number of iteration steps. Moreover, an important observation is that the overall complexity of the WH method is hence less than the standard Homotopy as depicted in Table 2 which compares the runtime for all the algorithms.

In order to evaluate the effect of the number of collaborating antennas on

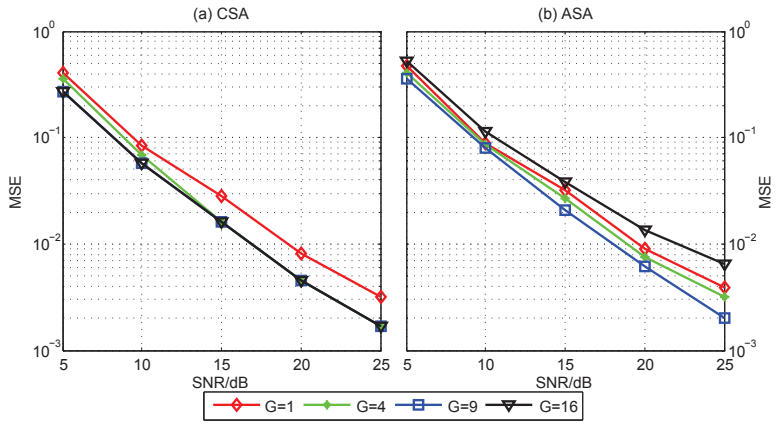

Figure 5: MSE comparison of the WH method for different G.

channel estimation performance for both of the CSA and ASA, we plot in Fig.5 the MSE comparison of the WH method for different G. It is evident from Fig.5(a) for CSA that the MSE decreases when we use multiple antennas to estimate the channel cooperatively, indicating that the information sharing strategy improves the channel estimation performance. Meanwhile, we observe that no further improvement can be attained when $G > 9$, implying that we can achieve a balance between channel estimation accuracy and computation complexity by choosing $G = 9$ in this case. In fact, there are several facts might have effects on the performance, e.g., the number of pilots or the sparsity of channel [19]. For example, with fewer pilots, we might also observe improvement by using $G > 9$. For the ASA case from Fig.2(b), we can see that even though the antenna cooperation with $G \leq 9$ leads to some performance improvement, it degrades the performance when $G = 16$. This is because that antennas with spacing over $\frac{C}{10B}$ would not share the same common support, the cooperation with them will degrade the channel estimation performance. Therefore, it is recommended that a small G is better in the ASA, which is compliant to the boundary in (30).

7. Conclusion

In this paper, we proposed an improved Homotopy method, which adaptively reweights the ℓ_1 norm inside a Homotopy iteration, to estimate the massive MIMO OFDM channel matrix. Since the proposed algorithm emphasizes more on adding elements into the active set and less on removing elements from the active set, it is appropriate to consider the approach as a compromise between standard Homotopy and LARS which is similar with standard Homotopy but omits the step that removes variables from the active set [28]. In addition, to further reduce the required pilots while enhance the estimation accuracy, we proposed a Homotopy-based joint channel estimation method where the antennas collaborate with their neighbors within the same safety zone. In the simulations, we investigated how the estimation performance is affected by the number of pilots and cooperative antennas, and proved that the proposed approach can yield high quality channel reconstructions in various scenarios while using small number of pilots.

References

[1] S.Nguyen, A.Ghrayeb, Compressive sensing-based channel estimation for massive multiuser mimo systems, IEEE Wireless Communications and Networking Conference (WCNC) (2013) 2890–2895.

[2] N.Shariati, E.Bjornson, M.Bengtsson, M.Debbah, Low-complexity polynomial channel estimation in large-scale mimo with arbitrary statistics, IEEE Journal of Selected Topics in Signal Processing 8 (5) (2014) 815–830.

[3] H.Taoka, K.Higuchi, Experiments on peak spectral efficiency of 50 bps/hz with 12-by-12 mimo multiplexing for future broadband packet radio access, International Symposium on Communications, Control and Signal Processing (ISCCSP) (2010) 1–6.

[4] G.Breit, et al., Channel modeling, Doc. IEEE 802.11-09/0088r0.

[5] H.Ngo, E.Larsson, Evd-based channel estimation in multicell multiuser mimo systems with very large antenna arrays, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2012) 3249–3252.

[6] S.Kay, Fundamentals of statistical signal processing: Estimation theory, Fundamentals of Statistical Signal Processing.

[7] J.Kotecha, A.Sayeed, Transmit signal design for optimal estimation of correlated mimo channels, IEEE Transactions on Signal Processing 52 (2) (2004) 546–557.

[8] D.Donoho, I.Johnstone, A.Maleki, A.Montanari, Compressed sensing over ℓ_p-balls: Minimax mean square error, IEEE International Symposium on Information Theory Proceedings (ISIT) (2011) 129–133.

[9] C.Berger, Z.Wang, J.Huang, S.Zhou, Application of compressive sensing to sparse channel estimation, IEEE Communications Magazine 48 (11) (2010) 164–174.

[10] S.Nguyen, A.Ghrayeb, M.Hasna, Iterative compressive estimation and decoding for network-channel-coded two-way relay sparse isi channels, IEEE Communications Letters 16 (12) (2012) 1992–1995.

[11] J.Huang, C.Berger, S.Zhou, J.Huang, Comparison of basis pursuit algorithms for sparse channel estimation in underwater acoustic ofdm, IEEE OCEANS 2010 (2010) 1–6.

[12] C.Qi, L.Wu, X.Wang, Underwater acoustic channel estimation via complex homotopy, IEEE International Conference on Communications (ICC) (2012) 3821–3825.

[13] D.Donoho, A.Maleki, A.Montanari, Message passing algorithms for compressed sensing: I. motivation and construction, IEEE Information Theory Workshop (ITW) (2010) 1–5.

[14] W.Chen, M.Rodrigues, I.Wassell, Penalized ℓ_1 minimization for reconstruction of time-varying sparse signals, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2011) 3988–3991.

[15] M.Asif, J.Romberg, Fast and accurate algorithms for re-weighted ℓ_1-norm minimization, IEEE Transactions on Signal Processing 61 (23) (2013) 5905–5916.

[16] P.Holland, R.Welsch, Robust regression using iteratively reweighted least-squares, Communications in Statistics Theory and Methods 6 (9) (1977) 813–827.

[17] I.Gorodnitsky, B.Rao, Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm, IEEE Transactions on Signal Processing 45 (3) (1997) 600–616.

[18] I.Daubechies, R.DeVore, M.Fornasier, C.Gunturk, Iteratively reweighted least squares minimization for sparse recovery, Communications on Pure and Applied Mathematics 63 (1) (2010) 1–38.

[19] M.Masood, L.Afify, T.AlNaffouri, Efficient coordinated recovery of sparse channels in massive mimo, IEEE Transactions on Signal Processing 63 (1) (2015) 104–118.

[20] L.Dai, Z.Wang, Z.Yang, Spectrally efficient time-frequency training ofdm for mobile large-scale mimo systems, IEEE Journal on Selected Areas in Communications 31 (2) (2013) 251–263.

[21] E.Bonek, Experimental validation of analytical mimo channel models, e&i Elektrotechnik und Informationstechnik 122 (6) (2005) 196–205.

[22] H.Ozcelik, N.Czink, E.Bonek, What makes a good mimo channel model?, IEEE Vehicular Technology Conference (2005) 156–160.

[23] Guideline for evaluation of radio transmission technology for imt-2000, Tech. rep., Recommendation ITU-R M.1225.

[24] Y.Barbotin, A.Hormati, S.Rangan, M.Vetterli, Estimation of sparse mimo channels with common support, IEEE Transactions on Communications 60 (12) (2012) 3705–3716.

[25] Z.Gao, L.Dai, Z.Wang, Structured compressive sensing based superimposed pilot design in downlink large-scale mimo systems, Electronics Letters 50 (12) (2014) 896–898.

[26] C.Qi, L.Wu, Uplink channel estimation for massive mimo systems exploring joint channel sparsity, Electronics Letters 50 (23) (2014) 1770–1772.

[27] A.Maleki, L.Anitori, Z.Yang, R.Baraniuk, Asymptotic analysis of complex lasso via complex approximate message passing (camp), IEEE Transactions on Information Theory 59 (7) (2013) 4290–4308.

[28] D.Donoho, Y.Tsaig, Fast solution of ℓ_1-norm minimization problems when the solution may be sparse, IEEE Transactions on Information Theory 54 (11) (2008) 4789–4812.

[29] J.Fuchs, On sparse representations in arbitrary redundant bases, IEEE Transactions on Information Theory 50 (6) (2004) 1341–1344.

[30] Y.Wang, W.Yin, Sparse signal reconstruction via iterative support detection, SIAM J. Imaging Sci. 49 (6) (2010) 2543–2563.

[31] B.Rao, K.Engan, S.Cotter, Diversity measure minimization based method for computing sparse solutions to linear inverse problems with multiple measurement vectors, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04) (2004) 17–21.

[32] Ilmprop, [Online]. Availabel: http://www.tu-ilmenau.de/nt/en/ilmprop [Online; accessed Jan. 10, 2014].