Article:

Nan, Y, Zhang, L and Sun, X (2017) Weighted Compressive Sensing Based Uplink Channel Estimation for TDD Massive MIMO Sytems. IET Communications, 11 (3). pp. 355-361. ISSN 1751-8628

https://doi.org/10.1049/iet-com.2016.0625

Weighted Compressive Sensing Based Uplink Channel Estimation for TDD Massive MIMO Systems

Yang Nan[1,*], Li Zhang[2], Xin Sun[1]

[1]School of Electrical and Information Engineering, Beijing Jiaotong University, Beijing, 100044
[2]School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT
[*]ynanbjtu@hotmai.com

Abstract: In this paper, the channel estimation problem for the uplink massive multi-input multi-output (MIMO) system is considered. Motivated by the observations that the channels in massive MIMO systems may exhibit sparsity and the channel support changes slowly over time, we propose one efficient channel estimation method under the framework of compressive sensing. By exploiting the channel impulse response (CIR) estimated from the previous OFDM symbol, we firstly estimate the probabilities that the elements in the current CIR are nonzero. Then, we propose the probability-weighted subspace pursuit (PWSP) algorithm exploiting these probability information to efficiently reconstruct the uplink massive MIMO channel. Moreover, noting that the massive MIMO systems also share a common support within one channel matrix due to the shared local scatterers in the physical propagation environment, an antenna collaborating method is exploited for the proposed method to further enhance the channel estimation performance. Simulation results show that compared to the existing compressive sensing methods, the proposed methods could achieve higher spectral efficiency as well as more reliable performance over time-varying channel.

1. Introduction

As a promising technology for future communication networks, massive multiple-input multiple-output (MIMO) systems where the base station (BS) is equipped with a large number of antennas have gained much interest in the research community [1]. In order to benefit from the advantages of massive MIMO systems, there are still many issues that need to be properly addressed. For example, the channel state information (CSI) is crucial for both of the uplink and downlink. But in typical massive MIMO systems, reliable CSI acquisition is a challenging problem, especially for the downlink where each user has to estimate the channels from large number of BS antennas. To solve this problem, lots of works today have considered massive MIMO of time division duplex (TDD) systems, where CSI can be acquired only at the uplink, and then utilised at both transmission directions based on the assumption of channel reciprocity [2][3]. The uplink channel estimation is therefore critical for TDD massive MIMO since it affects the signal detection at both transmission directions.

Recently, there has been a growing interest in compressive sensing (CS) based channel estimation algorithms [4][5]. By exploiting the inherent sparsity of the MIMO channels, sparse channel estimation can give better estimation performance than conventional schemes such as least square (LS) and minimum mean square error (MMSE) [6]. In [4], a structured subspace pursuit (SSP) algorithm is proposed to estimate the massive MIMO channel through superimposed pilots. In

[5], the block based orthogonal matching pursuit scheme is proposed for massive multiple-input single-output (MISO) systems. Noting that consecutive frames tend to share some common multipaths even in time-varying propagation environment, some recent schemes exploit the previously estimated channel support to enhance current channel estimation [7]-[11]. [7]-[9] proposed the modified basis pursuit (MBP) utilizing the *a priori* signal support. In [10], the Auxiliary information based Block Subspace Pursuit (ABSP) is proposed, which utilizes the previous channel support to initialize the estimated support for subspace pursuit algorithm. However, using the *a priori* support information directly may lead to even worse performance when the channel is fast varying, since the path delays may change from their previous locations. In [11], the authors have further considered the incorrect indices in the previous support, and exclude them adaptively. However, this method does not take into account delay variance and the relationships between the channel supports in adjacent frames.

In this paper, we propose an efficient channel estimation scheme in uplink massive MIMO system. Specifically, inspired by the observation that in massive MIMO systems, the path delays change slowly although the path gains may change quickly [15][16], we first explore the temporal correlation of channel support in uplink TDD massive MIMO systems. And then we propose a method that easily estimates the probabilities of the nonzero path delays in current channel impulse response (CIR) based on the knowledge of the previous CIR. After that we utilize the probabilities as *a priori* information in the subspace pursuit (SP) algorithm [17][18], and propose the probability-weighted SP (PWSP) algorithm to improve the uplink massive MIMO channel estimation. Moreover, noting that the massive MIMO system may also share common support within one channel matrix [19]. Based on this assumption, an antenna collaborating method is exploited for the proposed channel estimation approach, so that the antennas could share information with their neighbouring antennas to strengthen their beliefs about the locations of the path delays, and thus to further improve the channel estimation performance. Compared with the conventional SP method, the proposed method reduces the pilot overhead significantly while improving the channel estimation performance.

The rest of the paper is organized as follows. We first describe the massive MIMO OFDM system model in Section II. Then the PWSP algorithm is proposed in Section III. Section IV discusses the simulation results while Section V concludes the paper.

Notations: Throughout this paper, boldface lower and upper case symbols represent vectors and matrices, respectively. Operators ^{T}, ^{H} and † represent transpose, Hermite and Moore-Penrose matrix inversion, respectively. $diag\{x\}$ is the diagonal matrix with x at its main diagonal. $\mathcal{P}(x)$, $|x|$, $\|x\|_p$ $supp(x)$ and $supp_K(x)$ denote the probability, cardinality, ℓ_p-norm, support and largest K elements in the support of x, respectively. $x\rangle_k$ denotes the k-sparse vector of x which retains the k largest elements of x while setting the rest of the elements to zero.

2. MASSIVE MIMO SYSTEM MODEL

Consider the uplink TDD massive MIMO OFDM system where the BS with M antennas is serving a large number U autonomous single-antenna user terminals (UTs) ($M > U$). During the ith OFDM symbol, the CIR between the mth BS antenna and one certain UT can be denoted as $\boldsymbol{h}_{i,m} = [h_{i,m}(0), h_{i,m}(1), \cdots, h_{i,m}(L-1)]^T$ with $1 \leq m \leq M$, where L is the maximum delay spread of the CIR. Under the assumption of channel sparsity, only K elements are nonzero in $\boldsymbol{h}_{i,m}$, satisfying $K \ll L$ [4].

Suppose the total number of OFDM subcarriers is N, among which N_p subcarriers are randomly employed to transmit pilot symbols. Thus we can denote the pilot sequence transmitted from one certain UT to the BS as $\boldsymbol{x} = [x_{P_1}, \cdots, x_{P_j}, \cdots, x_{P_{N_p}}]$, where $\boldsymbol{p} = [P_1, \cdots, P_j, \cdots, P_{N_P}]$ is the corresponding subcarrier indices of N_p pilots. At the BS side, the pilot vectors received by different BS antennas are distinct due to the different path gains in each uplink channel. We denote the $N_p \times 1$ received pilot vector as

$$\boldsymbol{y}_{i,m} = diag\{\boldsymbol{x}\}\boldsymbol{F}\boldsymbol{h}_{i,m} + \boldsymbol{n}_{i,m}, \tag{1}$$

where $\boldsymbol{F}$ is a $N_p \times L$ submatrix comprising the N_p rows and the first L columns of the standard $N \times N$ discrete Fourier transform matrix, and $\boldsymbol{n}_{i,m}$ is the additive white Gaussian noise with zero mean and unit variance. Moreover, let $\boldsymbol{Y}_i = [\boldsymbol{y}_{i,1}, \boldsymbol{y}_{i,2}, \cdots, \boldsymbol{y}_{i,M}]$ represent the $N_p \times M$ received pilots for all M BS antennas, $\boldsymbol{H}_i = [\boldsymbol{h}_{i,1}, \boldsymbol{h}_{i,2}, \cdots, \boldsymbol{h}_{i,M}]$ be the $L \times M$ CIR matrix and $\boldsymbol{N}_i = [\boldsymbol{n}_{i,1}, \boldsymbol{n}_{i,2}, \cdots, \boldsymbol{n}_{i,M}]$ be the $N_p \times M$ channel's AWGN, the signal model (1) can be equivalently written as

$$\boldsymbol{Y}_i = \Phi\boldsymbol{H}_i + \boldsymbol{N}_i, \tag{2}$$

where $\Phi \triangleq diag\{\boldsymbol{x}\}\boldsymbol{F}$.

Based on many practical measurements of the channel matrix of massive MIMO systems [11]-[14], we have the following two important observations:

Observation I (Temporal Correlation Between Channel Matrices) In [4], the authors exploit the temporal correlation of wireless fading channels, whereby the channels exhibit identical common support in R adjacent OFDM symbols, indicating that $S_{1,m} = \cdots = S_{i,m} = \cdots = S_{R,m}$, where $S_{i,m} = supp\{\boldsymbol{h}_{i,m}\} = \{l : |h_{i,m}(l)| > 0\}_{l=0}^{L-1}$. However, this method suffers from performance deterioration even when there is only slight changes in the support. In the Fig.3 of [18], the authors illustrate the CIRs of four adjacent OFDM symbols over the International Telecommunications Union Vehicular B (ITU-VB) channel with 120km/h receiver velocity and signal bandwidth B=7.56 MHz [15][16]. As shown in these figures, the path delays are not completely the same, although they are nearly invariant, since the support of CIR changes over time when there is a relative movement between the UT and the BS. Nevertheless, note that the path delays just vary within their vicinity, we can still extract some useful information from the CIR of the previous OFDM symbol (i.e., $S_{i-1,m}$) to assist current channel estimation.

Observation II (Common Support within Channel Matrix) Some existing researches [5]-[11] indicate that the massive MIMO channels are usually correlated at the BS side. That is to say, the CIRs of different uplink channels share a common support since the distance between the BS and UT is far larger than the antenna spacing at the BS side. This can also be proved by the fact that two channel taps are not resolvable if the time interval of arrival is smaller than $\frac{1}{10B}$ [19], i.e. $\frac{d_{max}}{c} < \frac{1}{10B}$ where d_{max} is the maximum resolvable distance, and c is the speed of light. Consider a 16×16 massive MIMO system where $B = 20$MHz and the signal wavelength is $\lambda = 0.116m$ [22]. The distance between two adjacent antennas is $d = \frac{\lambda}{2} = 0.058$m. Hence, the maximum distance between two BS antennas is $D = 15d = 0.87$m (D is a function of the antenna spacing d and the number of antennas in the array), while $d_{max} = \frac{c}{10B} = 3.96$m which is much larger than D. As a result, the BS uplink channels share the common support S_i, i.e., $S_i = S_{i,1} = \cdots = S_{i,m} = \cdots = S_{i,M}$ [10].

Note that Observation I refers to the temporal shared support between different channel matrices, while Observation II refers to the spatial common support within one channel matrix. In the next section, we shall propose an efficient channel estimation method based on these observations.

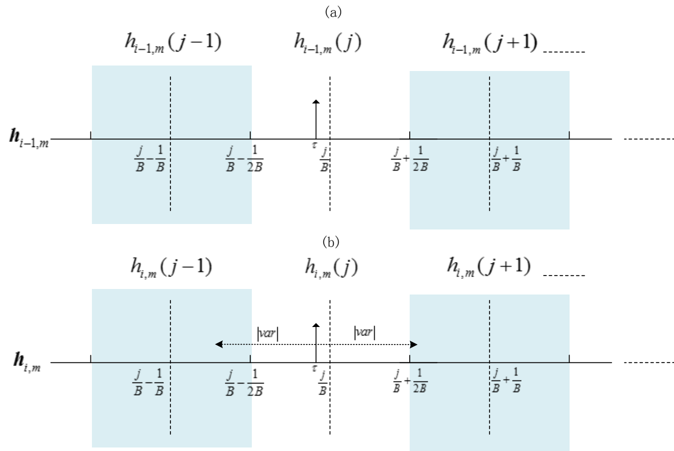

Fig. 1. *The relationship between the signal bandwidth B and channel taps in two successive OFDM symbols.*

3. *A PRIORI* INFORMATION AIDED CHANNEL ESTIMATION ALGORITHMS

3.1. Estimation of probabilities

Note that the spatial resolution of channel taps is determined by the inverse bandwidth $\frac{1}{B}$ [19], in Fig.1 we demonstrate the relationship between B and the channel taps in two successive OFDM symbols. From Fig.1(a) we can see that the channel tap with delay τ can be recognized as $h_{i-1,m}(j)$ (i.e., $j \in S_{i-1}$) if $\frac{j}{B} - \frac{1}{2B} \leq \tau \leq \frac{j}{B} + \frac{1}{2B}$. Assume $B = 7.56$MHz [18], then we can obtain the minimum resolvable interval of the channel taps as $\frac{1}{2B} \approx 0.06\mu s$. Let the maximum delay spread $\tau_{max} = 20\mu s$ according to the ITU-VB channel with 120km/h receiver velocity, and the maximum variation rate of the delays between two adjacent OFDM symbols is $\nu = \pm 0.5\%$[§], then we can acquire the variation of the delays by $|var| \leq |\nu \times \tau_{max}| = 0.1\mu s$ [20]. Noting that $|var| < \frac{1}{B}$, from Fig.1(b) we can see that in the following OFDM symbol, this path tap will (i) be invariant (i.e., $j \in S_i$) if $\frac{j}{B} - \frac{1}{2B} \leq \tau + var \leq \frac{j}{B} + \frac{1}{2B}$; (ii) move to $j-1$ (i.e., $j-1 \in S_i$) if $\tau + var < \frac{j}{B} - \frac{1}{2B}$; or (iii) move to $j+1$ (i.e., $j+1 \in S_i$) if $\tau + var > \frac{j}{B} + \frac{1}{2B}$.

In addition, by assuming that the value of var occurs uniformly within $|var| \leq 0.1\mu s$, we have $\mathcal{P}(|var| \leq \frac{1}{2B}) = \frac{1/2B}{\nu\tau_{max}} = \frac{3}{5}$ and $\mathcal{P}(|var| > \frac{1}{2B}) = \frac{\nu\tau_{max} - 1/2B}{\nu\tau_{max}} = \frac{2}{5}$. Therefore, the conditional probabilities of the nonzero channel taps can be expressed as

$$\begin{cases} \mathcal{P}(h_{i,m}(l) \neq 0 | l \in S_{i-1}) = \frac{3}{5} \\ \mathcal{P}(h_{i,m}(l \pm 1) \neq 0 | l \in S_{i-1}) = \frac{2}{5} \end{cases}, \tag{3}$$

Assume ς is a small integer, indicating the offset of delays from the $(i-1)$th symbol to the ith symbol. Based on the previous analysis, we can easily derive the relationship between the maximum ς (denoted as ς_{max}), the signal bandwidth B, the maximum delay spread τ_{max} and the variation ν as

$$\frac{\varsigma_{max}}{B} - \frac{1}{2B} < \nu\tau_{max} < \frac{\varsigma_{max}}{B} + \frac{1}{2B}. \tag{4}$$

[§]Since the channel support depends on the large scale properties of the scattering environment and changes slowly over a very long timescale in practice, we can obtain the knowledge of ν easily based on the *a priori* knowledge of the propagation environment [11].

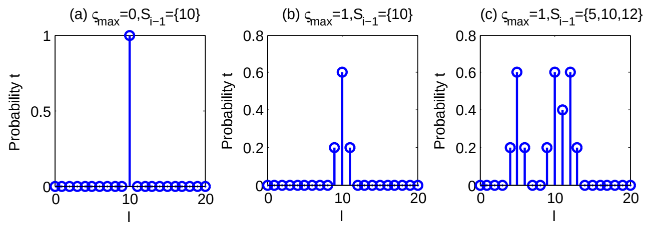

Fig. 2. *The probability t_l based on the previous channel support S_{i-1}.*

Thus, we have

$$\varsigma_{max} = \lceil \frac{2B\nu\tau_{max} - 1}{2} \rceil_+, \tag{5}$$

where $\lceil \cdot \rceil_+$ denotes the ceil function. Moreover, we can further exploit the relationship between B, τ_{max}, ν and the probability $\mathcal{P}(h_{i,m}(l+\varsigma) \neq 0 | l \in S_{i-1})$. It is obvious that if $\varsigma_{max} = 0$, the path delays are invariant, that is

$$\begin{cases} \mathcal{P}(h_{i,m}(l) \neq 0 | l \in S_{i-1}) = 1 \\ \mathcal{P}(h_{i,m}(l) \neq 0 | l \notin S_{i-1}) = 0 \end{cases}. \tag{6}$$

When $\varsigma_{max} > 0$, we have

$$\begin{cases} \mathcal{P}(h_{i,m}(l) \neq 0 | l \in S_{i-1}) = \mathcal{P}(|var| \leq \frac{1}{2B}) \\ \mathcal{P}(h_{i,m}(l \pm \varsigma) \neq 0 | l \in S_{i-1}) = \mathcal{P}(\frac{2\varsigma - 1}{2B} < |var| \leq \frac{2\varsigma+1}{2B}) \\ \qquad\qquad for\ 0 < \varsigma < \varsigma_{max} \\ \mathcal{P}(h_{i,m}(l \pm \varsigma_{max}) \neq 0 | l \in S_{i-1}) = \mathcal{P}(\frac{2\varsigma_{max} - 1}{2B} < |var|) \end{cases}. \tag{7}$$

We can then obtain the probabilities that the lth delay in $(i-1)$th symbol is shifted to the position $l \pm \varsigma$ in the ith symbol as,

$$\mathcal{P}(h_{i,m}(l \pm \varsigma) \neq 0 | l \in S_{i-1}) = \begin{cases} \frac{1}{2B\nu\tau_{max}}, & \varsigma = 0 \\ \frac{1}{B\nu\tau_{max}}, & 0 < \varsigma < \varsigma_{max} \\ 1 - \frac{2\varsigma_{max}-1}{2B\nu\tau_{max}}, & \varsigma = \varsigma_{max} \\ 0, & else \end{cases}. \tag{8}$$

Next, let $\boldsymbol{T} = diag([t_0, \cdots, t_l, \cdots, t_{L-1}])$ where t_l denotes the probability that the lth element in $\boldsymbol{h}_{i,m}$ is nonzero, i.e. $t_l = \mathcal{P}(h_{i,m}(l) \neq 0)$. Therefore, we have

$$t_l = \sum_{b=0}^{L-1} \mathcal{P}(h_{i,m}(l) \neq 0 | b \in S_{i-1}) \mathcal{P}(b \in S_{i-1}), \tag{9}$$

where $\mathcal{P}(b \in S_{i-1})$ equals 1 or 0 is known *a priori* according to the previous CIR.

In Fig.2, we illustrate three examples of the probabilities t_l with (a) $\varsigma_{max} = 0, S_{i-1} = \{10\}$, (b) $\varsigma_{max} = 1, S_{i-1} = \{10\}$ and (c) $\varsigma_{max} = 1, S_{i-1} = \{5, 10, 12\}$, respectively. Clearly, in Fig.2(a) the channel is static where the path delays are invariable during the successive symbols. On the other hand, in Fig.2(b) and Fig.2(c) the channels are time-varying, where the path delay may change but only within its neighbourhood due to $\varsigma_{max} = 1$. It is also worth noting that t_{11} is larger than t_9 in Fig.2(c) since both of its neighbours are in the support of $\boldsymbol{h}_{i-1}$, which increases the probability that the path delay at $l = 11$ is nonzero.

3.2. Uplink channel estimation with the aid of the probabilities

In order to incorporate the probability t_l into the CS algorithm, we assign a diagonal matrix $\boldsymbol{W} = diag([w_0, \cdots, w_l, \cdots, w_{L-1}])$ based on the probability $\boldsymbol{T}$. In this paper, we first initial the weights by

$$w_l = 1 + \alpha t_l, 0 \leq l \leq L-1. \tag{10}$$

where $\alpha \geq 0$ is a user-selected parameter.

Note that the weights $\boldsymbol{W}$ are exploited based on not only the probabilities $\boldsymbol{T}$ but also the parameter α, where the value of α determines the degree that the probabilities $\boldsymbol{T}$ may impact the algorithm. Obviously, in the slow time-varying channel where the variance rate ν is small, a larger number of indices in $supp(\boldsymbol{T})$ is correct, which indicates a more reliable $\boldsymbol{T}$. Therefore, by using a larger α, we encourage the PWSP algorithm to select indices from $supp(\boldsymbol{T})$ to form the support set Ω. On the contrary, in the fast time-varying channel where ν is large, less indices in $supp(\boldsymbol{T})$ are correct. In such cases, we reduce the impact of the *a priori* information on the support detection by using a relatively smaller α. In this way, the *a priori* support information $supp(\boldsymbol{T})$ is utilized adaptively based on the variance rate ν, and hence better recovery performance can be achieved[§]. Since α is inversely proportional to ν, we assume $\nu = \beta\%$ and employ $\alpha = \frac{1}{|\beta|}$.

Then we propose the PWSP approach designed by exploiting the *a priori* information and the antenna collaboration based on the conventional subspace pursuit [4]. The details of PWSP is presented in Algorithm 1.

Algorithm 1

Input: Received pilot matrix $\boldsymbol{Y}_i$, sensing matrix Φ, previous support S_{i-1}, approximated channel sparsity $K = \|S_{i-1}\|_0$.
Initialization:
  The initial residual $\boldsymbol{V}_0 = \boldsymbol{Y}_i$, $\boldsymbol{W} = \boldsymbol{1} + \boldsymbol{T}/\nu$, $\alpha = \frac{1}{100|\nu|}$, $\Gamma = S_{i-1}$ and $k = 1$;
while $\|\boldsymbol{V}_k\|_F < \|\boldsymbol{V}_{k-1}\|_F$ **do**
  $\Omega = \arg\max\limits_{|\Omega'|=K} \|(\boldsymbol{W}\Phi^H\boldsymbol{V}_k)^{\Omega'}\|_F$;
  $\Gamma \leftarrow \Gamma \cup \Omega$;
  $\Gamma \leftarrow \arg\max\limits_{|\Omega'|=K} \|(\boldsymbol{W}_\Gamma\Phi_\Gamma^\dagger\boldsymbol{Y}_i)^{\Omega'}\|_F$;
  Set $\hat{\boldsymbol{H}}_i^\Gamma = \Phi_\Gamma^\dagger\boldsymbol{Y}_i$ and $\hat{\boldsymbol{H}}_i^{[\{0,\cdots,L-1\}\backslash\Gamma]} = \boldsymbol{0}$;
  $k \leftarrow k+1$;
  $\boldsymbol{V}_k \leftarrow \boldsymbol{Y}_i - \Phi^H\hat{\boldsymbol{H}}_i$;
end while
Output: The estimated CIR vector $\hat{\boldsymbol{H}}_i$.

In Algorithm 1, Γ is the estimated support set, $\boldsymbol{V}_k$ is the residual at the kth iteration. Φ_Γ and Φ^Γ denote the sub-matrix formed by collecting the column vectors and row vectors of Φ respectively, whose indices belong to Γ.

Note that the PWSP is developed based on the classical SP algorithm but with two major changes:

1) Antenna Collaborating. In the SP algorithm, the K significant path delays are captured by maximizing the correlations between the column vectors of Φ and the residual measurement, i.e.

[§]Note that when $M = 1$ and $\nu \to \infty$, Algorithm 1 will reduce to conventional SP [17].

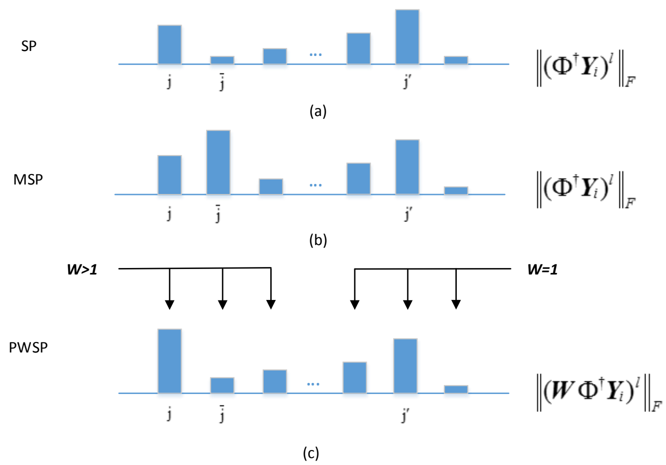

Fig. 3. *The impact of the weights on support detection.*

$\Gamma \leftarrow \arg \max_{|\Omega'|=K} \|(\Phi_\Gamma^\dagger \boldsymbol{y}_{i,m})^{\Omega'}\|$, where $\boldsymbol{y}_{i,m}$ is the ith received pilot sequence on the mth antenna. As a result, the estimated support may be inaccurate due to the imperfect Φ with only approximate orthogonal columns [17]. In the proposed PWSP method, the receive antennas collaborate with each other to take advantage of Observation II and estimate the path delays jointly, i.e. $\Gamma \leftarrow \arg \max_{|\Omega'|=K} \|(\Phi_\Gamma^\dagger \boldsymbol{Y}_i)^{\Omega'}\|_F$. Therefore, according to the law of large numbers, it is more likely to capture the correct path delays since the antennas share information with each other to reach a decision on the most probable support (see more details in our previous work [10]).

2) Weights Aided. Suppose that $\bar{j} \in S_{i-1}$, $j = \bar{j}-1 \in S_i$ and $j' \notin S_{i-1} \cup S_i$. For the classical SP algorithm with antenna collaborating, consider an ideal case that all the columns of Φ are orthogonal, the K significant path delays can be easily captured by calculating $arg \max_{|\Omega'|=K} \|(\Phi^\dagger \boldsymbol{Y}_i)^{\Omega'}\|_F$, since $\|(\Phi^\dagger \boldsymbol{Y}_i)^j\|_F \gg \|(\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F$. However, in the practical case, the columns of Φ are only locally near-orthogonal. As a result, it may happen that the value of $\|(\Phi^\dagger \boldsymbol{Y}_i)^j\|_F$ is smaller than the value of $\|(\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F$, and this leads to incorrect support detection. On the other hand, in the modified SP (MSP) algorithm [7], the previous channel support (i.e., the $\bar{j}$th path) is used to enhance the channel estimation directly, which is equivalent to adding a large weight on the $\bar{j}$th path. Consequently, MSP could improve the channel estimation performance if the channel is invariant (i.e., $\bar{j} \in S_{i-1} \bigcap S_i$). However, in the time-varying channel where $\bar{j} \notin S_i$, this scheme may lead to bad performance.

Compared with these conventional algorithms, the proposed PWSP exploits more information from the previous channel support. Specifically, according to the analyses in Section 3.1, we have $w_j > w_{j'}$. Hence, it is more likely that $\|(\boldsymbol{W}\Phi^\dagger \boldsymbol{Y}_i)^j\|_F > \|(\boldsymbol{W}\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F$ although $\|(\Phi^\dagger \boldsymbol{Y}_i)^j\|_F \leq \|(\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F$. Consequently, PWSP is more probable to obtain the correct support. To illustrate this point, we give a simple example in Fig. 3. Assume $\bar{j} \in S_{i-1}$, $j = \bar{j} - 1 \in S_i$ and $j' \notin S_{i-1} \cup S_i$, it can be seen from Fig.3(a) that $\|(\Phi^\dagger \boldsymbol{Y}_i)^{\bar{j}}\|_F < \|(\Phi^\dagger \boldsymbol{Y}_i)^j\|_F < \|(\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F$. As a result, the conventional SP algorithm will capture the incorrect path delay j', which leads to performance deterioration. For the MSP scheme in Fig. 3(b), we can see that $\|(\Phi^\dagger \boldsymbol{Y}_i)^j\|_F < \|(\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F < \|(\Phi^\dagger \boldsymbol{Y}_i)^{\bar{j}}\|_F$. Hence, the MSP will capture the incorrect index $\bar{j}$, On the contrary, with the PWSP scheme in Fig. 3(c), we have $\|(\boldsymbol{W}\Phi^\dagger \boldsymbol{Y}_i)^{\bar{j}}\|_F < \|(\boldsymbol{W}\Phi^\dagger \boldsymbol{Y}_i)^{j'}\|_F < \|(\boldsymbol{W}\Phi^\dagger \boldsymbol{Y}_i)^j\|_F$. Therefore, the proposed PWSP will still capture the correct index j into the support set.

4. SIMULATION RESULTS

In this section, we conduct simulation studies to evaluate the performance of the proposed channel estimation algorithms. The signal bandwidth is 7.56 MHz locating at a central frequency of 760 MHz. For the uplink transmission, the system parameters are set as: OFDM subcarriers $N = 2048$, number of BS antennas $M = 64$, sparsity of channel $K = 6$, maximum variation $\nu = \pm 0.5\%$ and maximum delay spread $L = 151$ [20]. We assume the path delays of the previous CIR is known *a priori**. In order to verify the effectiveness of the proposed scheme, we compare the performance of PWSP with the conventional SP [17], modified SP (MSP), Basis Pursuit (BP) [17], modified Basis Pursuit (MBP) [7], and exact Least Square (exact LS) with perfectly known support of the sparse channel as the performance benchmark [4]. The MSP and MBP are developed based on the conventional SP and BP respectively, with exploitation of the previous support information.

4.1. Channel Estimation Performance Versus Overhead N_p

Firstly, in Fig.4 we compare the mean square error (MSE) of the estimated channel versus the number of pilots, with signal-to-noise ratio (SNR) $SNR = 25dB$. From the figure, we observe that the channel estimation improves with the increase of N_p, and the proposed PWSP outperforms other conventional methods. Specifically, PWSP achieves a substantial performance gain over the MSP. This is because the *a priori* channel support $S_{i-1,m}$ is not always correct (i.e., $S_{i-1,m}$ is not identical to $S_{i,m}$). In the MSP, $S_{i-1,m}$ is used directly to aid current channel estimation. As a result, the incorrect *a priori* information may lead to performance deterioration. On the other hand, PWSP also exploits how reliable the *a priori* channel support is as illustrated in Section 3.1. Therefore, it could provide more stable performance in time varying channel. Moreover, we observe that PWSP performs closely to the exact LS when $N_p > 23$, which is caused by the fact that Φ has imperfect but approximate orthogonal columns.

In order to evaluate the stability, we compare the success rate of channel recovery versus the number of pilots N_p in Fig.5. 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 when the MSE of channel estimation is better than 10^{-1} [22]. The number of pilots N_p is varied from 1 to 35, while 100 independent trails are implemented for each N_p. It can be seen from the figure that PWSP evidently outperforms other conventional methods. Specifically, when 15 pilots are adopted, PWSP can achieve a success rate of 66% while the MSP, the best among the conventional methods, can only reach the success rate of 26%. In addition, PWSP requires 23 pilots to achieve a success rate of 100%, while MSP requires at least 27 pilots, which indicates that the proposed PWSP could lead to a significant improvement (about 17%) in spectral efficiency.

4.2. Channel Estimation Performance Versus Transmit SNRs

In Fig.6 we present the MSE performance comparison of the estimated channel versus the transmit SNR with $N_p = 23$. From the figure, we can see that the proposed PWSP can achieve enormous MSE gains over other conventional methods, and approach the performance bound (i.e., exact LS) in higher SNR regions.

*This assumption is reasonable since we could always employ more pilots in the first OFDM symbol to obtain accurate CIR, and then reduce the number of pilots by using PWSP.

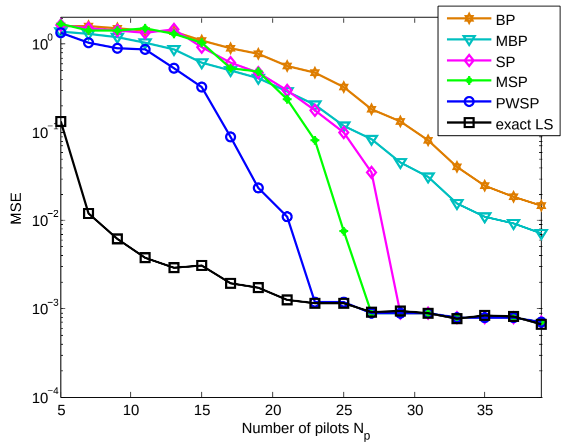

Fig. 4. *MSE comparison of estimated channel versus N_p.*

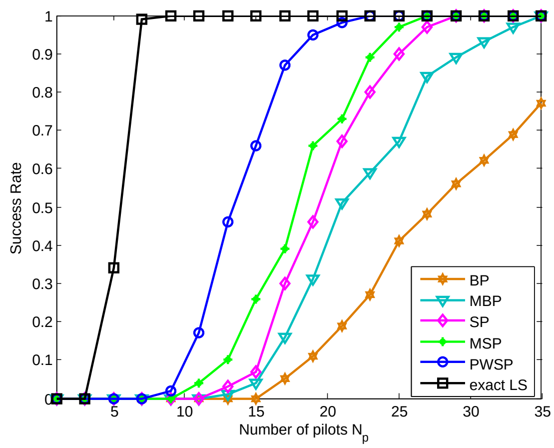

Fig. 5. *Success rate comparison of estimated channel versus N_p.*

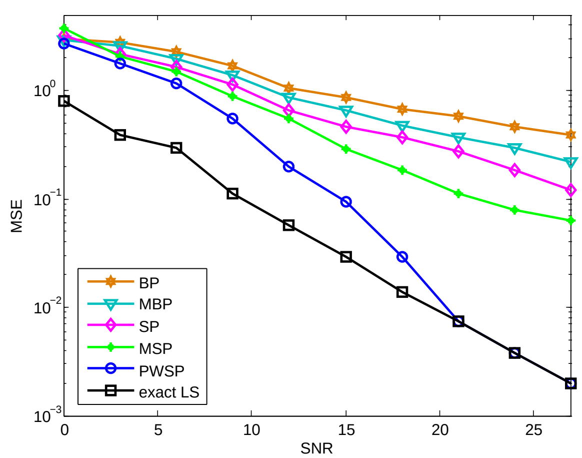

Fig. 6. *MSE of estimated channel versus transmit SNR with $N_p = 23$.*

4.3. Channel Estimation Performance With Different Number of BS Antennas M

In Fig.7, we set $K = 12$ and compare the MSE performance of the PWSP scheme with different number of BS antennas for both $N_p = 16$ and $N_p = 32$. It is evident from both Fig. 7(a) and Fig.

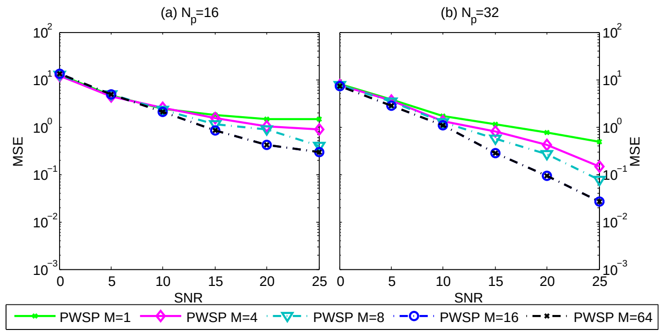

Fig. 7. *MSE performance comparisons of PWSP with different number of BS antennas M.*

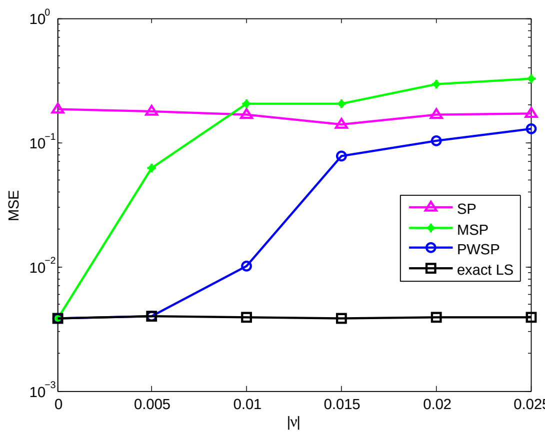

Fig. 8. *MSE performance comparisons of PWSP versus the variation rate $|\nu|$.*

7(b) that the channel estimation performance improves as M increases. This is because that the BS antennas could coordinate with each other and reach a decision on the most probable channel support for channel estimation (i.e., Line 5 and Line 7 of Algorithm 1) [10]. Thus, according to the law of large numbers, more coordinated BS antennas could provide more accurate support detection, as long as Observation II is satisfied.

4.4. Channel Estimation Performance Versus the Variation Rate $|\boldsymbol{\nu}|$

In order to evaluate the impact of ν on the proposed PWSP, we set $N_p = 23$, $M = 64$ and illustrate the MSE comparisons with different $|\nu|$ in Fig.8. It is obvious that both of MSP and PWSP achieve a substantial performance gain over the SP when the channel is invariant ($|\nu| = 0$). However, the performance of the MSP degrades severely when $|\nu| > 0$. On the contrary, the PWSP is relatively stable and still provides performance gains over the SP and MSP, demonstrating the robustness of the proposed PWSP scheme in fast time-varying channel[§].

5. CONCLUDING REMARKS

This paper considers the uplink channel estimation for massive MIMO system. By extracting the *a priori* information from the previous CIRs and exploiting the spatial correlation of massive MIMO

[§]Note that PWSP is reduced to the conventional SP when $|\nu| = \infty$.

channels, the PWSP algorithm is proposed to efficiently perform the channel estimation with only few pilots. Simulation results have shown that the proposed scheme could achieve higher spectral efficiency as well as more reliable performance. It is also demonstrated that PWSP could still achieve performance gain in fast time-varying channel.

6. Acknowledgements

The authors would like to thank the support by "National Natural Science Foundation of China (61201199)".

7. References

[1] Marzetta T.: 'Noncooperative cellular wireless with unlimited number of base station antenns', IEEE Trans. Wirel. Commun., 2010, 9, (11), pp. 3590-3600

[2] Yansha, D., Lifeng, W., Kaikit, Wong.: 'Safeguarding massive MIMO aided hetnets using physical layer security', 2015 Int. Conf. Wirel. Commun. & Signal Process. (WCSP), Nanjing, Oct. 2015, pp. 1-5

[3] Yansha, D., Lifeng, W., Maged, E., *et al.*: 'K-tier heterogeneous cellular networks with wireless power transfer', 2016 IEEE Int. Conf. Commun. (ICC), Kuala Lumpur, Jul. 2016, pp. 1-6

[4] Gao Z., Dai L., Wang Z.: 'Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems', Electron. Lett., 2014, 50, (23), pp. 896-898

[5] Hou W., Lim C.: 'Structured compressive channel estimation for large-scale MISO-OFDM systems', IEEE Commun. Lett., 2014, 18, (5), pp. 765-768

[6] Qi C., Nan L.: 'Uplink channel estimation for massive MIMO systems exploring joint channel sparsity', Electron. Lett., 2014, 50, (23), pp. 1770-1772

[7] Vaswani N., Lu W.: 'Modified-CS: Modifying compressive sensing for problems with partially known support', IEEE Trans. Signal Process., 2010, 58, (9), pp. 4595-4607

[8] Jacques L.: 'A short note on compressed sensing with partially known signal support', Signal Process., 2010, 90, (12), pp. 3308-3312

[9] Amel R., Feuer A.: 'Adaptive identification and recovery of jointly sparse vectors, IEEE Trans. Signal Process., 2014, 62, (2), pp. 354-362

[10] Nan Y., Zhang L., Sun X.: 'Efficient Downlink Channel Estimation Scheme Based on Block-Structured Compressive Sensing for TDD Massive MU-MIMO Systems', IEEE Wirel. Commun. Lett., 2015, 4, (4), pp. 345-348

[11] Rao X., Lau V.: 'Compressive Sensing With Prior Support Quality Information and Application to Massive MIMO Channel Estimation With Temporal Correlation', IEEE Trans. Signal Process. 2015, 63, (18), pp. 4914-4924

[12] Kyritsi P., Cox D., Valenzuela R., *et al.*: 'Correlation analysis based on MIMO channel measurements in an indoor environment', IEEE J. Sel. Areas Commun., 2003, 21, (5), pp. 713-720

[13] Kaltenberger F., Gesbert D., Knopp R., *et al.*: 'Correlation and capacity of measured multi-user MIMO channels', in Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC), Cannes, Sep. 2008, pp. 1-5

[14] Hoydis J., Hoek C., Wild T., *et al.*: 'Channel measurements for large antenna arrays', in Proc. IEEE Int. Symp., Pairs, Aug. 2012, pp. 811-815

[15] Dai L., Wang Z., Yang Z.: 'Spectrally and efficient time-frequency training OFDM for mobile large-scale MIMO systems', IEEE J. Sel. Areas Commun., 2013, 31, (2), pp. 251-263

[16] Rusek F., Persson D.: 'Scaling up MIMO: opportunities and challenges with very large arrays', IEEE Signal Process. Mag., 2013, 30, (1), pp. 40-60

[17] Wei D., Milenkovic O.: 'Subspace Pursuit for Compressive Sensing Signal Reconstruction', IEEE Trans. Inform. Theory, 2009, 55, (5), pp. 2230-2249

[18] Gao Z., Zhang C.: 'Priori-Information Aided Iterative Hard Threshold: A Low-Complexity High-Accuracy Compressive Sensing Based Channel Estimation for TDS-OFDM', IEEE Trans. Wirel. Commun., 2015, 14, (1) pp. 242-251

[19] Barbotin Y., Vetterli M.: 'Estimation of sparse MIMO channels with common support', IEEE Trans. Commun., 2012, 60, (12), pp. 93-100

[20] M.1225: 'Guideline for Evaluation of Radio Transmission Technology for IMT-2000', 1997

[21] Ding W., Yang F.: 'Compressive sensing based channel estimation for OFDM systems under long delay channels', IEEE Trans. Broadcasting, 2014, 60, (2), pp. 313-321

[22] Masood M., Afify L., AlNaffouri T.: 'Efficient Coordinated Recovery of Sparse Channels in Massive MIMO', IEEE Trans. Signal Process., 2015, 63, (1), pp. 104-118