Proceedings Paper:
Nan, Y, Zhang, LX and Sun, X (2016) An Efficient Downlink Channel Estimation Approach for TDD Massive MIMO Systems. In: Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd. 23rd Vehicular Technology Conference (VTC), 15-18 May 2016 IEEE. ISBN: 978-1-5090-1698-3.

https://doi.org/10.1109/VTCSpring.2016.7504124

An Efficient Downlink Channel Estimation Approach for TDD Massive MIMO Systems

Yang Nan, Li Zhang, and Xin Sun

Abstract—**In this paper, the channel estimation problem for the downlink massive multi-input multi-output (MIMO) system is considered. Motivated by the observation that the channels in massive MIMO systems always exhibit sparsity and the path delays vary slowly in one uplink-downlink process, we propose a novel channel estimation method under the framework of the weighted compressive sensing. Unlike the conventional methods which do not make use of any *a priori* information or assume the path delays are invariable, we estimate the probabilities that the paths are nonzero in the downlink channel by exploiting the channel impulse response (CIR) estimated from the uplink channel estimation. Based on these probabilities, we propose the Weighted Structured Subspace Pursuit (WSSP) algorithm to efficiently reconstruct the massive MIMO channel. Simulation results show that compared to the conventional methods, the WSSP could achieve a significant reduction in the number of pilots while maintain good channel estimation performance.**

Index Terms—**Massive MIMO, channel estimaion, compressive sensing.**

I. Introduction

As a promising technology for future communication technologies, massive multiple-input multiple-output (MIMO) systems where the base station (BS) is equipped with a large number of antennas have received enormous attention [1]. It is shown that with increase number of BS antennas, the massive MIMO techniques could provide unprecedented spectral efficiency and array gains. However, to implement this technique in practice, there are still many issues that need to be properly addressed. One of the key challenges is the downlink channel estimation, where the required number of the downlink pilots is proportional to the large number of antennas at the BS side, which is unaffordable for the massive MIMO system.

To solve this problem, some efficient downlink channel estimation schemes have been proposed based on the structured compressive sensing (SCS), which could recovery the channel with relatively few pilots by exploiting the sparse nature of the massive MIMO channels [3]-[6]. In [3], the structured subspace pursuit (SSP) algorithm is proposed for downlink massive MIMO system by using the superimposed pilots. In [4], block based orthogonal matching pursuit scheme is proposed for downlink massive multiple-input single-output (MISO) systems. In order to further improve the channel estimation, the authors in [2] exploit the channel reciprocity in time division duplex (TDD) mode whereby the user terminals (UTs) could utilize the channel support estimated from the uplink training to enhance the downlink channel estimation [2]. Based on this idea, they propose the Auxiliary information based Block Subspace Pursuit (ABSP) algorithm. However, the assumption of channel reciprocity sometimes does not hold since the support of CIR changes over time if there is a relative movement between the UTs and the base station. As a result, using the uplink channel support directly in the downlink channel estimation may lead to performance deterioration.

In this paper, we propose a new Weighted Structured Subspace Pursuit (WSSP) algorithm for the downlink channel estimation in TDD massive MIMO systems. This approach is inspired by the weighted CS (WCS) [7] and the observation that in the massive MIMO systems the path delays (active paths) in the downlink channel change slowly over one uplink-downlink process [8][9]. Firstly, we present a method that easily estimates the probabilities of the nonzero paths delays in the downlink channel based on the knowledge of the previous uplink CIRs. After that we combine these probabilities with the SSP by adjusting the weights of SSP according to the probabilities to improve the channel estimation performance. Compared with the conventional SSP method and ABSP method, the proposed method reduces the pilot overhead significantly while maintains an accurate channel estimation performance.

The rest of the paper is organized as follows. We first describe the downlink massive MIMO system model in Section II. Then the WSSP algorithm is proposed in Section III. Section IV presents the simulation results. Finally, section VI 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. $diag\{c\}$ is the diagonal matrix with x at its main diagonal. $\mathcal{P}(x)$, $\|x\|_p$ and $supp_K(x)$ denote the probability, the ℓ_p-norm and the largest K elements in the support of x, respectively.

II. Massive MIMO OFDM System Model

Consider a downlink massive MIMO OFDM system where the BS with M antennas is serving a large number of U autonomous single-antenna UTs ($M > U$). The CIR between the mth BS antenna and one certain UT can be denoted as $\boldsymbol{h}_m = [h_m(0), h_m(1), \cdots, h_m(L-1)]^T$ with $1 \leq m \leq M$, where L is the maximum delay spread of the CIR. Under the

Y. Nan is with the School of Electrical and Information Engineering, Beijing Jiaotong University, Beijing, 100044 China e-mail: ynanbjtu@hotmai.com.

X. Sun is with the School of Electrical and Information Engineering, Beijing Jiaotong University, Beijing, 100044 China e-mail: xsun@bjtu.edu.cn.

L. Zhang is with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT the United Kindom e-mail: L.X.Zhang@leeds.ac.uk.

Manuscript received ; revised .

assumption of channel sparsity, only K elements are nonzero in $\boldsymbol{h}_m$, satisfying $K \ll L$ [3]. Meanwhile, it is reasonable to assume that the CIRs of the uplink channels between the UT and different BS antennas share a common support, considering that the distance between UT and BS is much bigger than the space between the antennas at the BS [3]-[5]. In other words, by defining $S_m = supp\{\boldsymbol{h}_m\} = \{\tau : |h_m(\tau)| > 0\}_{\tau=0}^{L-1}$, we have $S_1 = S_2 = \cdots = S_M$.

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 the mth antenna of the BS to one certain UT as $\boldsymbol{x}_m = [x_{P_{1,m}}, \cdots, x_{P_{j,m}}, \cdots, x_{P_{N_p,m}}]$, where $\boldsymbol{p}_m = [P_{1,m}, \cdots, P_{j,m}, \cdots, P_{N_P,m}]$ is the corresponding subcarriers indices of N_p pilots. To reduce the pilot overhead, we adopt the superimposed pilot pattern that the pilots at different transmit antennas share the same locations, i.e. $\boldsymbol{p}_1 = \boldsymbol{p}_2 = \cdots = \boldsymbol{p}_M$, but each pilot sequence $\boldsymbol{x}_m$ is unique. In this paper, we simply generate the pilots by setting $x_{P_{j,m}} = 1$ or $x_{P_{j,m}} = -1$ with $1 \leq j \leq N_p$, following the identically and independently distributed (i.i.d) random Bernoulli distribution [3].

At the UT side, the received pilot vectors from different BS antennas are distinct due to the distinct $\boldsymbol{x}_m$ and path gains in each downlink channel. The superposition of M pilot vectors can be expressed as

$$\begin{aligned} \boldsymbol{y} &= \Sigma_{m=1}^{M} diag\{\boldsymbol{x}_m\}\boldsymbol{F}\boldsymbol{h}_m + \eta_m \\ &= \Sigma_{m=1}^{M} \boldsymbol{A}_m\boldsymbol{F}\boldsymbol{h}_m + \eta_m \end{aligned} \quad (1)$$

where $\boldsymbol{A}_m = diag\{\boldsymbol{x}_m\}$ is the diagonal matrix with $\boldsymbol{x}_m$ on its main diagonal, $\boldsymbol{F}$ is a $N_p \times L$ submatrix comprising the $\boldsymbol{p}$ rows and the first L columns of the standard $N \times N$ discrete Fourier transform matrix, and η_m is the additive white Gaussian noise. Moreover, let Φ denote the $N_p \times LM$ matrix as

$$\Phi = [\boldsymbol{A}_1\boldsymbol{F}, \boldsymbol{A}_2\boldsymbol{F}, \cdots, \boldsymbol{A}_M\boldsymbol{F}], \quad (2)$$

and $\boldsymbol{h} = [\boldsymbol{h}_1^T, \cdots, \boldsymbol{h}_M^T]^T = [h(0), h(1), \cdots, h(ML-1)]^T$, then we have

$$\boldsymbol{y} = \Phi\boldsymbol{h} + \eta. \quad (3)$$

Since $\boldsymbol{h}$ is a sparse vector, we can formulate the channel estimation problem as a classic ℓ_1 norm minimisation problem,

$$\arg\min \|\boldsymbol{h}\|_1 \ s.t. \ \|\boldsymbol{y} - \Phi\boldsymbol{h}\|_2 \leq \varepsilon, \quad (4)$$

for some suitable $\varepsilon > 0$.

Besides the spatial correlation which ensures the common support of the sparse MIMO channel, we have also noticed the temporal correlation of wireless fading channels, whereby the channels exhibit identical common support in one uplink-downlink process [2]. However, using the uplink channel support directly to aid the downlink channel estimation is sometimes unreliable since the support of CIRs may change over time if there is a relative movement between the UT and BS. Despite of this, we can still find some useful information from the uplink support, owing to the substantial correlation of CIRs between uplink and downlink [8][9].

In this paper we apply a novel channel estimation method called WSSP which is able to produce accurate channel estimation performance from even fewer number of pilots than the SSP by exploiting *a priori* information obtained from the estimated uplink CIRs.

III. WSSP Algorithm for Uplink Massive MIMO Systems

The main idea of WSSP is based on the observation that the support of the downlink channel is closely similar to the support of the uplink channel in one uplink-downlink process. As a result, we could be able to explore some useful information about the support of $\boldsymbol{h}$ from the CIR of the uplink channel. Since the path delays vary slowly, we could calculate the probabilities that the elements in $\boldsymbol{h}$ are nonzero according to $\hat{S}$, where $\hat{S}$ is the channel support estimated from the uplink training, and then use these probabilities to aid the reconstruction of $\boldsymbol{h}$. This approach is divided into the following two steps: 1) Estimation of the probabilities; 2) Downlink channel estimation with the aid of the probabilities.

A. Estimation of the probabilities

Note that two channel taps are not resolvable if the time interval is no more than $\frac{1}{2B}$ [2]. That is, a channel tap with delay τ can be recognized to be $h(j)$ if $\tau_j - \frac{1}{2B} \leq \tau \leq \tau_j + \frac{1}{2B}$, where $\tau_j = \frac{j}{B}$ denotes the delay of $h(j)$. For example, assume $B = 7.56$MHz [10], then we can obtain the minimum resolvable interval of channel taps as $\frac{1}{2B} \approx 0.06\mu s$. Let the maximum delay spread τ_{max} be $20\mu s$ according to the ITU-VB channel with 120km/h receiver velocity and the maximum variation rate of the delays within one uplink-downlink process is $\nu = \pm 0.5\%$, then we can acquire the variation of the delays by $|var| \leq \nu \times \tau_{max} = 0.1\mu s < \frac{1}{B}$ [11]. Thus, we can conclude that the path delay in uplink channel will either be invariant ($|var| \leq \frac{1}{2B}$) or move next to its original location ($\frac{1}{2B} < |var| \leq \frac{1}{B}$) in the downlink.

Let $\boldsymbol{t} = [t_0, \cdots, t_l, \cdots, t_{ML-1}]$ be the probabilities that the lth element in $\boldsymbol{h}$ is nonzero , i.e., $t_l = \mathcal{P}(h(l) \neq 0)$ for $0 \leq l \leq ML-1$. It is obvious that if $t_l = 1$ then the lth element in $\boldsymbol{h}$ should be nonzero (though the value of $h(l)$ is still unknown), while if $t_l = 0$ the lth element in $\boldsymbol{h}_i$ should be definitely zero. Based on the assumption that the path delays only vary slightly in one uplink-downlink process, there is a good chance that the positions of nonzero elements in the uplink channel would be either unchanged or just shift to the nearby locations in the downlink process. In addition, it is also reasonable to predict that zero elements will appear in similar locations with uplink. Thus, suppose $l \in \hat{S}$, then there would be a good chance that this spike remains at this location or move to the locations in the vicinity of l in the downlink, i.e., $h(l \pm \varsigma) \neq 0$ where ς is a small integer. Meanwhile, the probability of $h(l \pm \varsigma) \neq 0$ decreases as ς gets bigger. This behaviour inspires us to express the probability t_l by using the Gaussian function,

$$t_l = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(l-b)^2/2\sigma^2}, \quad (5)$$

where σ is the standard deviation which depends on the variation rate ν. Assume $\hat{S} = \{10\}$ and $\nu = \pm 0.5\%$, we plot the measured probability of this spike in downlink

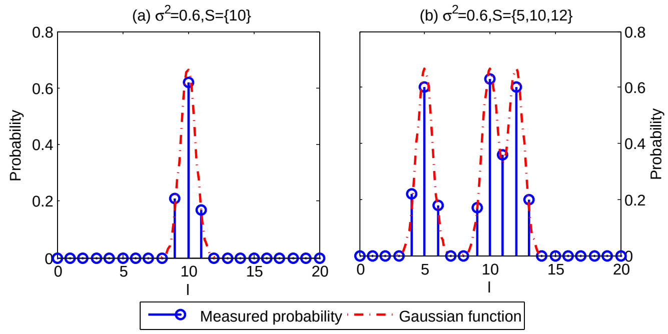

Fig. 1. Comparison between the measured probability and Gaussian function

channel by performing 100 independent trails in Fig.1(a). Also, the Gaussian functions with $\sigma^2 = 0.6$ is draw as comparison. It is obvious that the measured probability and the Gaussian function match very well. Therefore, we can obtain the estimated probability of this spike in downlink channel from (5) as $t_{10} = 0.66, t_9 = t_{11} = 0.16$. Moreover, if there were multiple paths, we could simply acquire the probabilities by summing the probabilities of each nonzero spikes as

$$t_l = \sum_{b \in \hat{S}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(l-b)^2/2\sigma^2} \tag{6}$$

In Fig.1(b), we illustrate an example with three spikes ($\hat{S} = \{5, 10, 12\}$). It is worth noting that t_{11} has higher probability than t_9 since both of its neighbours are in the support $\hat{S}$, which increases the probability that the path gain at $j = 11$ is nonzero.

B. Downlink channel estimation with the aid of the probabilities

In order to incorporate the probability t_l into the CS algorithm, we replace the ℓ_1 norm in (4) with a weighted norm, where the weights are adjusted according to the probability t_l, and then obtain

$$\arg\min \|\boldsymbol{W}\boldsymbol{h}\|_1 \ s.t. \ \|\boldsymbol{y} - \Phi\boldsymbol{h}\|_2 \le \varepsilon, \tag{7}$$

where $\boldsymbol{W} = diag([w_0, \cdots, w_l, \cdots, w_{ML-1}])$ is the matrix with the weights w_l $(0 \le l \le ML-1)$ on its diagonal. Intuitively, smaller weights could encourage elements in $\boldsymbol{h}$ to remain nonzero, while the larger weights urge elements to be zero. Therefore, smaller weights should be set to the elements with higher probabilities of being nonzero. In this paper, we first initialize the weights w_l equal to the reciprocal of t_l as[1]

$$w_l = \begin{cases} \frac{1}{t_l}, & for \ t_l \ne 0 \\ 0, & for \ t_l = 0 \end{cases}, \tag{8}$$

and then normalize w_l by

$$w_l = \frac{w_l}{\max\limits_{0 \le j \le ML-1}(w_j)}, \tag{9}$$

[1]It is worth noting that the weighted ℓ_1 minimization problem in (7) is equivalent to the ℓ_1 minimization problem in (4) when no *a priori* information t_l is available (where all the weights could be 1).

To solve the weighted ℓ_1 minimization problem, based on the classical SSP algorithm [3] which does not take any auxiliary information into account, we propose the WSSP approach to perform channel estimation with the aid of weights. The proposed WSSP algorithm is summarized in Algorithm 1.

Algorithm 1

1: **Input:** Received pilot sequence $\boldsymbol{y}$, sensing matrix Φ, previous support $\hat{S}$, approximated channel sparsity $K = \|\hat{S}\|_0$
2: **Initialization:**
3: The initial residual $\boldsymbol{v}_1 = \boldsymbol{y}_i$, $k = 1$ and $\boldsymbol{W} \leftarrow 0$
4: If $i = 1$ then
5: $\boldsymbol{W} \leftarrow 1$
6: Else
7: Obtain $\boldsymbol{W}$ from (5) and (8)
8: $\Omega \leftarrow supp(\boldsymbol{W})$
9: End
10: **while** $\|\boldsymbol{v}_k\|_2 < \|\boldsymbol{v}_{k-1}\|_2$ **do**
11: $\boldsymbol{z} \leftarrow \Phi^H \boldsymbol{v}_k$
12: $d(l) \leftarrow \sum_{m=0}^{M-1} |z^{(l+mL)}|^2, 0 \le l \le L-1$
13: $\Omega \leftarrow \Omega \cup supp_K(\boldsymbol{d})$
14: $\Gamma \leftarrow \Omega \cup [\Omega + L] \cup, \cdots, \cup[\Omega + L(M-1)]$
15: If $w_{l \in \Gamma} = 0$ $(0 \le l \le ML-1)$ then
16: $w_l = \max\limits_{j \in \Gamma}(w_j)$;
17: Elseif $w_{l \in \Gamma} > 0$
18: $w_l = w_l \times (1 - \alpha)$;
19: Else
20: $w_l = w_l / \max\limits_{j \in \Gamma}(w_j)$;
21: End
22: If $w_l = 1$ then
23: $w_l = 0$;
24: End
25: $\boldsymbol{z} \leftarrow (\Phi \boldsymbol{W}_\Gamma)^\dagger \boldsymbol{y}_i$
26: $d(l) \leftarrow \sum_{m=0}^{M-1} |z^{(l+mL)}|^2, 0 \le l \le L-1$
27: $\Omega \leftarrow supp_K(\boldsymbol{d})$
28: $\Gamma \leftarrow \Omega \cup [\Omega + L] \cup, \cdots, \cup[\Omega + L(M-1)]$
29: $\hat{\boldsymbol{h}}_\Gamma \leftarrow \Phi_\Gamma^\dagger \boldsymbol{y}_i$
30: $k \leftarrow k + 1$
31: $\boldsymbol{v}_k \leftarrow \boldsymbol{y} - \Phi^H \hat{\boldsymbol{h}}$
32: **end while**
33: **Output:** The estimatied CIR vector $\hat{\boldsymbol{h}}$

Note that $\boldsymbol{d} = [d(0), d(1), \cdots, d(L-1)]^T$, $\boldsymbol{z} = [z(0), z(1), \cdots, z(ML-1)]^T$, $\boldsymbol{v}_k$ denotes the residual at the kth iteration. $0 < \alpha < 1$ is a user-selected parameter that controls the decay rates of $\boldsymbol{W}_\Gamma$, where $\boldsymbol{W}_\Gamma$ denotes the submatrix collecting the columns of $\boldsymbol{W}$ according to Γ, and $[\Omega + a]$ means adding a to each element of Ω.

Since we consider not only the original elements in $\hat{S}$, but also the elements in their vicinity, we initialize the set Ω as $\Omega \leftarrow supp(\boldsymbol{W})$ in Line 8, instead of $\Omega \leftarrow \hat{S}$ as the traditional methods did [2]. Moreover, note that $supp(\boldsymbol{W})$ determines the paths that can be processed by the WSSP, while the nonzero values in $\boldsymbol{W}$ affect the probabilities that whether the corresponding paths could be included into the support Γ. Hence, we adjust the weights $\boldsymbol{W}$ iteratively according to the following three situations: 1) The l is one of the K largest magnitude entries of $\boldsymbol{d}$, which means the lth path is selected to be one of the possible nonzero paths by the WSSP, but the current lth element in $\boldsymbol{W}$ is zero, indicating the lth path is regarded to be zero according to the previous steps, then we should assign w_l to a relatively large value as $w_l = \max_{j\in\Gamma}(w_j)$, so that the lth path could be processed in the following WSSP steps but with less influence (Line 16); 2) Otherwise, if the l is one of the K largest magnitude entries of $\boldsymbol{d}$ and $w_l \neq 0$, which indicates that this index has already been considered to be the nonzero path by the previous steps and is collected again in the current step, implying it may belong to the correct support of $\boldsymbol{h}$ with high probability, so we reduce the value of w_l at a rate α to encourage the lth element in $\boldsymbol{h}$ to remain nonzero (Line 18); 3) For the other l' which is not one of the K largest magnitude entries of $\boldsymbol{d}$ but with $w_{l'} > 0$, we divide it by $\max_{j\in\Gamma}(w_j)$ as a punishment, i.e., $w_{l'} = w_{l'} / \max_{j\in\Gamma}(w_j)$ (Line 20). Due to this punishment, the maximum values of $\boldsymbol{W}$ will increase to 1. Then, we set these elements to zeros since the corresponding channel taps are zeros with high probabilities (Line 23).

The main computational complexity of WSSP comes from the matrix inversion operation required to obtain the path delays (Line 25) during each iteration. Although WSSP requires to process more columns of Φ in the first iteration than SSP to ensure the accuracy of the estimated support, it will rule out the redundant columns from Φ quickly by setting $w_l = 0$ according to Line 23[2]. After that, the complexity of WSSP will be exactly the same as SSP. Moreover, owing to the introduction of *a priori* information, the proposed WSSP could converge in fewer iterations than that required by the SSP, which could also decrease the total complexity. Therefore, it is reasonable to say that the computational complexity of WSSP is at the same order of SSP.

IV. Simulation Results

In this section, simulation studies are conducted to investigate the performance of the proposed WSSP algorithm. Consider a massive MIMO system with $M = 8$ BS antennas. For the downlink transmission, $N = 4096$ OFDM subcarriers are used. The maximum delay spread $L = 200$ is considered with only $K = 12$ nonzero elements. The signal bandwidth B is 7.56MHz and the maximum variance rate of the delays within one uplink-downlink process is $\nu = \pm 0.5\%$ [11]. Moreover, the decay rate α is 0.5 while the variance of the Gaussian distribution $\sigma^2 = 0.6$.

In Fig.2 we set the signal-to-noise ratio (SNR) to $25dB$, and compare the success rate of channel recovery between the conventional SSP algorithm, ABSP as well as the proposed WSSP, when a varying number of pilots N_p is employed. 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 is better than 10^{-1}. The number of pilots N_p is varied from 20 to 340, while 100 independent trails are implemented for each N_p. It can be seen from Fig.2

[2]In fact, WSSP could eliminate most of the redundant columns in the first iteration.

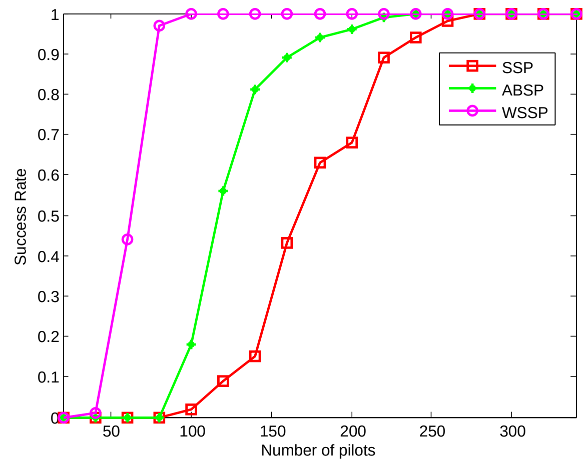

Fig. 2. Success rate comparisons between SSP, ABSP and WSSP.

that WSSP is superior to the other algorithms. Specifically, it requires no more than 70 pilots to achieve the success rate of 50% and 100 pilots for 100%. By contrary, the ABSP and SSP require 240 pilots and 280 pilots respectively to reach a 100% success rate. Moreover, it also can be seen that the performance improvements of the proposed approach is obvious when the same number of pilots is used. For example, when $N_p = 100$, the success rates of WSSP is 100%, while that of the SSP and ABSP are 2% and 18%, respectively. Thus, it can be concluded that the proposed approach could reduce the pilot overhead significantly and thus improve the spectral efficiency.

Next, we set $N_p = 128$ and present the MSE comparison between the proposed WSSP, the conventional SSP and ABSP with $\nu = 0$ and $\nu = \pm 0.5\%$ in Fig.3 respectively. From Fig.3(a) we can see that the MSE performance of ABSP is just the same as WSSP, indicating its good performance in the static channel where the path delays are invariant during one uplink-downlink process. However, it suffers from deterioration when $\nu = \pm 0.5\%$ as shown in Fig.3(b), since the the channel is time-varying. By contrary, WSSP could achieve good performance in both of the static channel and the time-varying channel, since it considers not only the original path delays in the uplink but also their vicinity, which improves its performance in the time-varying channel. Specifically, WSSP outperforms ABSP by about 7dB in Fig.3(b) when the channel estimation MSE $10^{-0.7}$ is considered. Moreover, it is obvious that both of the ABSP and WSSP outperform the conventional SSP, thanks to the use of *a priori* information.

Finally, we set $\nu = \pm 0.5\%$ and $N_p = 300$, and then plot the MSE performance of different approaches in Fig.4. The exact least square (LS) method which perfectly knows the channel support is also included as the performance bound. It is observed that all the approaches could achieve the performance bound when both of the SNR and the number of pilots are sufficient. However, when $SNR < 15$, only WSSP performs as good as the exact LS method, which is superior to the other conventional methods, indicating that it still tracks the correct path delays although the SNR is low.

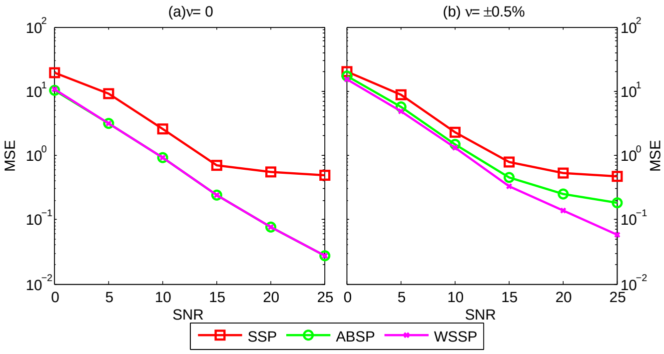

Fig. 3. MSE comparisons between SSP, ABSP and WSSP with different ν.

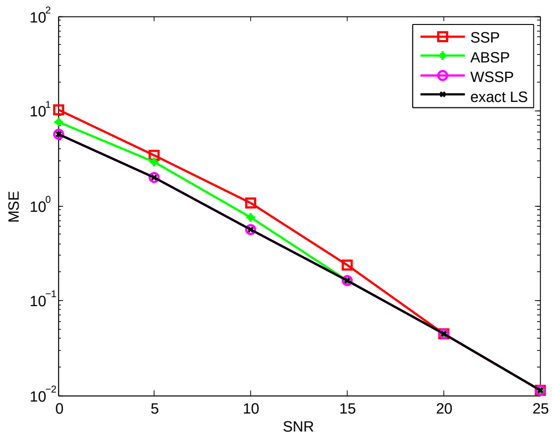

Fig. 4. MSE comparisons between SSP, ABSP and WSSP with $N_p = 300$.

V. CONCLUDING REMARKS

This paper considers the downlink channel estimation for massive MIMO system. By extracting the probability information of the path delays from the uplink channel, the weighted SSP algorithm is proposed to efficiently solve the channel estimation problem with only few pilots. Simulation results have shown that the proposed scheme could achieve higher spectral efficiency as well as more reliable performance over the time-varying channel.

ACKNOWLEDGEMENTS

The authors would like to thank the support by "the Fundamental Research Funds for the Central Universities (2014YJS006)" and "the Grants of National Basic Research Program of China (61201199)".

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited number of base station antennas", *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
[2] Y. Nan, L. Zhang, X. Sun, "Efficient Downlink Channel Estimation Scheme Based on Block-Structured Compressive Sensing for TDD Massive MU-MIMO Systems", *IEEE Wireless Communication Letters*, accepted on Mar. 2015.
[3] Z. Gao, L. Dai, and Z. Wang "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems", *Electornics Letters*, vol.50, no. 12, pp. 896-898, Jun. 2014.
[4] W. Hou, C. Lim, "Structured compressive channel estimation for large-scale MISO-OFDM systems", *IEEE Communications Letters*, vol. 18, no. 5, pp. 765-768, May. 2014.
[5] Y. Barbotin, M. Vetterli, "Estimation of sparse MIMO channels with common support", *IEEE Transations on Communications*, vol. 60, no. 12, pp. 93-100, Dec. 2012.
[6] C. Qi, L. Nan, "Uplink channel estimation for massive MIMO systems exploring joint channel sparsity", *Electronics Letters*, vol. 50, no. 23, pp. 1770-1772, Nov. 2014.
[7] W. Dai, O. Milenkovic "Weighted-CS for reconstruction of highly under-sampled dynamic MRI sequences", *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012 Asia-Pacific, pp. 1-5, Dec. 2012.
[8] L. Dai, Z. Wang, Z. Yang "Spectrally and efficient time-frequency training OFDM for mobile large-scale MIMO systems", *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 251-263, Feb. 2013.
[9] F. Rusek, D. Persson, et al., "Scaling up MIMO: opportunities and challenges with very large arrays", *IEEE Signal Processing Magazine*, vol.30, no. 1, pp. 40-60, Jan. 2013.
[10] Z. Gao, C. Zhang, et al., "Priori-Information Aided Iterative Hard Threshold: A Low-Complexity High-Accuracy Compressive Sensing Based Channel Estimation for TDS-OFDM", *IEEE Transactions on Wireless Communications*, vol.14, no.1, pp. 242-251, Jan. 2015.
[11] *Guideline for Evaluation of Radio Transmission Technology for IMT-2000*, ITU-R M. 1225 Standard, Dec. 1997.
[12] W. Ding, F. Yang, et al., "Compressive sensing based channel estimation for OFDM systems under long delay channels", *IEEE Transactions on Broadcasting*, vol.60, no.2, pp. 313-321, June. 2014.