Article:

Nan, Y, Sun, X and Zhang, L (2017) Decision Aided Uplink Compressive Channel Estimation for Massive MIMO Systems. Wireless Personal Communications, 96 (1). pp. 153-162. ISSN 0929-6212

https://doi.org/10.1007/s11277-017-4159-6

Noname manuscript No.
(will be inserted by the editor)

Decision Aided Uplink Compressive Channel Estimation for Massive MIMO Systems

Yang Nan · Xin Sun · Li Zhang

Received: date / Accepted: date

Abstract Thank to the observation that in massive multi-input multi-output (MIMO) systems, the channels associated with different base station (BS) antennas may share common sparse support, the significant path delays can be accurately captured by only few pilots, leading to a reduction of pilot overhead. However, when the number of pilots is small, the path gains can not be accurately estimated and this limits the system performance. To solve this problem, in this paper we propose a decision aided compressive sensing based channel estimation scheme, which utilizes the decoded data to refine the channel estimation. This scheme can effectively improve the channel estimation without increasing the length of pilot sequence, which is confirmed by both analyses and simulation results.

Keywords Massive MIMO · channel estimation · compressive sensing

1 Introduction

As a promising technology for the future fifth-generation wireless communication, massive multiple-input multiple-output (MIMO) systems with a large

Yang Nan
School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China
Tel.: 0086-010-51685034
E-mail: ynanbjtu@hotmail.com

Xin Sun
School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China

Li Zhang
School of Electronic and Electrical Engineering, University of Leeds, LEEDs, the United Kingdom

number of antennas at the base station (BS) have received enormous attention [1]. It is shown that with the increase number of BS antennas, the massive MIMO techniques could reduce transmit power while provide unprecedented spectral efficiency and array gains. However, there are also some issues beginning to resurface. One of the key challenges for massive MIMO system is the channel estimation, where a large number of pilots is required to distinguish the channel impulse response (CIR) of each receiving antenna. Things are getting worse at the user terminal (UT) side where each user has to accurately estimate channels from the large number of BS antennas. For this reason, the time division duplex (TDD) mode is more desirable for massive MIMO systems, where the acquired uplink CIR at the BS can be directly fed back to the users thanks to the channel reciprocity in slow time-varying channels [2] or used as basis to obtain the downlink CIR in fast time-varying channels [3]. In this way, the heavy overhead of the pilots in the downlink transmission as well as the battery power consumption at the user terminal for channel estimation can be saved. Thus, accurate CSI acquisition for uplink is critical since it affects the signal detection at both transmission directions directly.

Recently, compressive sensing (CS) based channel estimation becomes popular in massive MIMO systems, since it could accurately reconstruct the channels by exploiting the sparse nature of MIMO channels [4]. Moreover, thank to the observation that the channels of different BS antennas may share common support, the path delays (i.e., the locations of the significant taps) can be estimated cooperatively among the BS antennas [2]-[5]. As a result, BS can acquire the path delays accurately with only few pilots. However, from previous works we found that the limited number of pilots is insufficient for the conventional CS recovery algorithms to accurately estimate the path gains, although the path delays can be well captured. To solve this problem, in this paper we propose a decision aided channel estimation scheme for the uplink massive MIMO systems. To estimate both of the path delays and path gains accurately with only few pilots, we acquire the path delays by cooperating the BS antennas while estimate the path gains by jointly using the pilots and the detected data [8][9]. Theoretical analyses and simulation results show that the proposed method could improve the channel estimation performance without increasing the pilot overhead.

The rest of the paper is organized as follows. We first describe the uplink massive MIMO OFDM system model in Section II. Then the Decision aided Subspace Pursuit (DSP) algorithm is proposed in Section III. Section IV presents the performance analyses of the proposed scheme while Section V discusses the simulation results. Finally, Section VI concludes the paper.

2 MASSIVE MIMO SYSTEM MODEL

Consider the uplink massive MIMO OFDM system with M antennas at the BS and U autonomous single-antenna user terminals (UTs) ($M > U$) using the orthogonal frequency division multiplexing (OFDM). 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 $1 \le m \le M$, where L is the maximum delay spread of the CIR. Normally, the channel is sparse with the number of nonzero elements K in $\boldsymbol{h}_m$, satisfying $K \ll L$ [2].

Suppose the number of OFDM subcarriers is N, among which N_p subcarriers are randomly employed to transmit pilot symbols. Thus we can denote the signal transmitted from one certain UT to the BS as $\boldsymbol{x} = [x_1, \cdots, x_j, \cdots, x_N]$, where $x_j = \pm 1$. Moreover, the corresponding subcarrier indices of N_p pilots are expressed as $\Omega = [P_1, \cdots, P_j, \cdots, P_{N_p}]$. Thus, at the mth antenna of the BS, the received signal vector is $\boldsymbol{y}_m = [y_{m,1}, y_{m,2}, \cdots, y_{m,N}]^T$ while the received pilot vector is

$$\begin{aligned} \boldsymbol{y}_m^{\Omega} &= diag\{\boldsymbol{x}^{\Omega}\}\boldsymbol{F}_L^{\Omega}\boldsymbol{h}_m + \eta_m \\ &= \Phi\boldsymbol{h}_m + \eta_m, \end{aligned} \tag{1}$$

where $\boldsymbol{y}_m^{\Omega} = [y_{m,P_1}, \cdots, y_{m,P_j}, \cdots, y_{m,P_{N_p}}]^T$, $\boldsymbol{x}^{\Omega} = [x_{P_1}, x_{P_2}, \cdots, x_{P_{N_p}}]^T$ and $diag\{\boldsymbol{x}^{\Omega}\}$ is the diagonal matrix with $\boldsymbol{x}^{\Omega}$ at its main diagonal. $\boldsymbol{F}_L^{\Omega}$ is a $N_p \times L$ submatrix formed by collecting the N_p rows whose indices belong to Ω, and the first L columns of the standard $N \times N$ discrete Fourier transform matrix $\boldsymbol{F}$, $\Phi \triangleq diag\{\boldsymbol{x}^{\Omega}\}\boldsymbol{F}_L^{\Omega}$ where $tr(\Phi\Phi^H) = N_p K$, and η_m is the additive white Gaussian noise with mean zero and variance ν [5].

3 DECISION AIDED CHANNEL ESTIMATION ALGORITHMS

It has been proved that the CIRs of different BS antennas may have similar path arrival times since the antenna spacing at the BS is far less than the distance between the UT and BS.[4]. Therefore, it is safe to say that the CIRs share a common support, indicating that $S_1 = S_2 = \cdots = S_M$, where $S_m = supp\{\boldsymbol{h}_m\} = \{l : |h_m(l)| > 0\}_{l=0}^{L-1}$ denotes the largest K elements in the support of $\{\boldsymbol{h}_m\}$. To take advantage of the common support, in [6] a coordinated method between the BS antennas has been proposed to reach a decision on the most probable channel support for all the antennas. In this way, the path delays can be acquired accurately with only few pilots. However, these pilots are not enough to accurately estimate the path gains, which fundamentally limits the performance of channel estimation. To solve this problem, we propose the Decision aided Subspace Pursuit (DSP) (as elaborated in Algorithm 1), which uses the decoded data to refine the estimation of the path gains and hence achieve more accurate channel estimation performance with few pilots.

Note that in Algorithm 1, $\hat{S}$ is the estimated support set, $\boldsymbol{v}_{k,m}$ and $\boldsymbol{c}_{t,m}$ are the residuals at the kth iteration and tth iteration, respectively. Operators ^{H} and † represent Hermite and Moore-Penrose matrix inversion, respectively. There are two main differences between the proposed algorithm and the standard subspace pursuit (SP) algorithm. Firstly, the proposed algorithm uses a simple antenna cooperation method we previously proposed in [6]. In order to

Algorithm 1

1: **Input:** Received pilot sequence $\boldsymbol{y}_m^{\Omega}$, sensing matrix Φ.
2: **Initialization:**
3: $\boldsymbol{v}_{1,m} = \boldsymbol{y}_m^{\Omega}$, $\boldsymbol{c}_{1,m} = \boldsymbol{y}_m$, $\hat{S}_m \leftarrow \emptyset$, $k = 1$, $t = 1$;
4: **while** $\|\boldsymbol{v}_{k,m}\|_2 < \|\boldsymbol{v}_{k-1,m}\|_2$ **do**
5: $\boldsymbol{z}_m \leftarrow \Phi^H \boldsymbol{v}_{k,m}$;
6: $\hat{S}_m \leftarrow \hat{S}_m \cup supp_K(\boldsymbol{z}_m)$;
7: $\boldsymbol{z}_m \leftarrow \Phi^{\dagger}_{\hat{S}_m} \boldsymbol{y}_m^{\Omega}$;
8: $\hat{S}_m \leftarrow supp_K(\boldsymbol{z}_m)$;
9: Obtain $\hat{S}$ according to (2) - (4);
10: $\hat{\boldsymbol{h}}_m \leftarrow \Phi^{\dagger}_{\hat{S}} \boldsymbol{y}_m^{\Omega}$;
11: $k \leftarrow k + 1$;
12: $\boldsymbol{v}_{k,m} \leftarrow \boldsymbol{y}_m^{\Omega} - \Phi^H \hat{\boldsymbol{h}}_m$;
13: **end while**
14: **while** $\|\boldsymbol{c}_{t,m}\|_2 < \|\boldsymbol{c}_{t-1,m}\|_2$ **do**
15: $\hat{\boldsymbol{\Phi}}_m = \boldsymbol{y}_m \hat{\boldsymbol{h}}_m^H$;
16: Obtain $\hat{\boldsymbol{x}}$ according to (5);
17: Randomly select N_g data as the new pilots and add
18: their indices into Ω;
19: $\hat{\boldsymbol{h}}_m \leftarrow \Phi^{\dagger}_{\hat{S}} \boldsymbol{y}_m^{\Omega}$;
20: $t = t + 1$;
21: $\boldsymbol{c}_{t,m} \leftarrow \boldsymbol{y}_m^{\Omega} - \Phi^H \hat{\boldsymbol{h}}_m$;
22: **end while**
23: **Output:** The estimated CIR vector $\hat{\boldsymbol{h}}_m$.

capture the common support efficiently, the BS antennas share their information about the estimated support $\hat{S}_m$ with each other, to reach a decision on the most probable channel support for all the antennas. In details, we define an integer vector $\boldsymbol{o}_m = [o_{0,m}, o_{1,m}, \cdots, o_{L-1,m}]$ as the score vector to record the path delays of the mth antenna and then update the $\boldsymbol{o}_m$ according to $\hat{S}_m$ which is obtained from Line 8 of Algorithm 1 at each iteration as

$$o_{l,m} = \begin{cases} 1, & l \in \hat{S}_m \\ 0, & l \notin \hat{S}_m \end{cases}. \tag{2}$$

Once all the BS antennas have updated their vector scores, an aggregate score of each channel tap is obtained by

$$\bar{o}_j = \sum_{m=1}^{M} o_{j,m},\ 0 \le j \le L-1. \tag{3}$$

According to the law of large numbers, the $\bar{o}_j$ corresponding to the true path delays should get higher scores since $S_1 = S_2 = \cdots = S_M$. Therefore, we sort $\bar{\boldsymbol{o}} = [\bar{o}_0, \bar{o}_1, \cdots, \bar{o}_{L-1}]$ and then obtain the estimated support $\hat{S}$ by selecting the K indices with the largest magnitude in $\bar{\boldsymbol{o}}$ as

$$\hat{S} = argmax_{|\hat{S}|=K} \|\bar{\boldsymbol{o}}_{\hat{S}}\| \tag{4}$$

The other difference is that the proposed algorithm employs detected data to refine the channel estimation iteratively. In details, after detecting the received sensing matrix $\hat{\boldsymbol{\Phi}}_m$ at the mth antenna as shown in Line 15, we easily obtain the decoded data as $\hat{\boldsymbol{x}}_m = \hat{\boldsymbol{\Phi}}_m \boldsymbol{F_L}^H$. Then we combine the data detection on each receive antenna to obtain the estimates of the transmitted data $\hat{\boldsymbol{x}}$, which can be expressed as

$$\hat{x}_j = \begin{cases} 1, & \sum\limits_{m=1}^{M} \hat{x}_{m,j} \geq 0 \\ -1, & \sum\limits_{m=1}^{M} \hat{x}_{m,j} < 0 \end{cases} \tag{5}$$

where $\hat{x}_j$ and $\hat{x}_{m,j}$ are the jth element of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_m$ respectively. Obviously, $\hat{\boldsymbol{x}}$ should be more accurate than $\hat{\boldsymbol{x}}_m$ since the antennas have shared their information within the antenna array to strengthen the beliefs about the detected signals. After that, we randomly select N_g data from $\hat{\boldsymbol{x}}$ and add their indices into Ω. Finally, the matched filter (MF) detector is exploited to obtain the path gains according to "new pilot sequence" (i.e., the original N_p pilots plus N_g decoded data) as shown in Line 19 of Algorithm 1. These steps in the loop are performed until the convergence is achieved.

4 PERFORMANCE ANALYSES

In this section, the performance of the proposed DSP algorithm is analyzed. Specifically, we are interested in both the normalized mean absolute error (NMAE) of the channel matrix [10] and the computational complexity of the proposed scheme. From the closed-form results, we can obtain simple insights into how the decoded data can be exploited to enhance the channel estimation performance.

4.1 NMAE analysis of the channel estimation

By utilizing the method of matrix and probability, we obtain the following theorem on the NMAE of the channel matrix.

Theorem 1: The normalized mean absolute error (NMAE) of $\boldsymbol{h}_m$ satisfies

$$\mathbb{E}\left(\frac{\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_F}{\|\boldsymbol{H}\|_F}\right) \leq \frac{1}{1-\delta_{2K}} + \sqrt{\frac{MK\nu}{N_p(1-\delta_{2k})}} \frac{\Gamma(M-\frac{1}{2})}{\Gamma(M)}, \tag{6}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_M]$, δ_{2k} is the restricted isometry property (RIP) constant [11].

Proof: See Appendix A.

From Theorem 1 we can see that the NMAE of the channel matrix decreases with the increase of the number of pilots N_p. That is to say, we can improve the channel estimation performance by increasing N_p. However, increasing N_p

will consequently reduce the spectral efficiency. Hence, in our proposed scheme, we euphemistically increase the number of pilots by employing the detected data. In this way, the proposed scheme could improve the channel estimation performance without sacrificing spectral efficiency. It is also worth noting that due to the noise and inaccurate channel estimation, it is possible that $\hat{\boldsymbol{x}}$ is not accurate to $\boldsymbol{x}$. However, we still have the belief that our algorithm can achieve better performance for the following reasons: 1) the detection error could be corrected when strong channel coding scheme (e.g., automatic repeat request) is applied in actual applications [8]; 2) the detected data are further refined by the antenna cooperation strategy, which improves the accuracy of the data; 3) the locations of N_g decoded data are selected randomly so that the threats of burst errors can be avoid. Therefore, we believe that even the data is not perfectly detected, the use of these detected data can still improve the channel estimation performance.

4.2 Complexity analysis

The complexity of the traditional channel estimator based on the SP scheme is dominated by the complexity of SP algorithm which is roughly $\mathcal{O}(N_pK^3 + 8K^4)$, where N_p is the number of pilots, K is the channel sparsity and L is the maximum spread of the channel [11]. For the DSP scheme, the complexity of the SP part (from Line 4 to Line 13 of Algorithm 1) is the same as the traditional SP scheme. However, the MF combining (Line 15) with complexity of $\mathcal{O}(N_p + N_g)K^2 + K^3)$ [12] is required in DSP. As a result, its complexity is about $\mathcal{O}(N_pK^2(K+1) + K^3(8K+1) + N_gK^2)$[1]. Suppose G is the total number of iterations, the overall complexity of DSP is about $\mathcal{O}(N_pK^2(K + G) + K^3(8K+G) + GN_gK^2)$. From the simulation results of DSP we can see that G is a small number and N_p can be reduced by roughly 3 times. Therefore, it is reasonable to say that the computational complexity of the proposed DSP is roughly on the same order of the standard SP.

5 SIMULATION RESULTS

We consider an $M = 32$ massive MIMO system with $N = 2048$ subcarriers in one OFDM symbol. The typical 6-tap multipath ITU-VB channel model with the maximum delay spread $L = 151$ is used [6].

In the first experiment, we set $N_g = 32$ and compare the mean square error (MSE) performance of the proposed DSP scheme in 5 iterations. For Fig.1(a), the transmit SNR is set to 0 dB while in Fig.1(b) SNR equals to 10 dB. From these figures we observe that the channel estimation performance improves as t increases under higher value of SNR (e.g., $SNR = 10$dB). However, when the value of SNR is low (e.g., $SNR = 0$dB), t has little effect on the performance

[1] The complexity of antenna cooperation is omitted since it is negligible, and the extra complexity introduced by the MF combining is far less than the complexity of SP.

of DSP. This is because that the data detection is not accurate at low SNR. As a result, using the inaccurate decoded data iteratively can not refine the channel estimation performance.

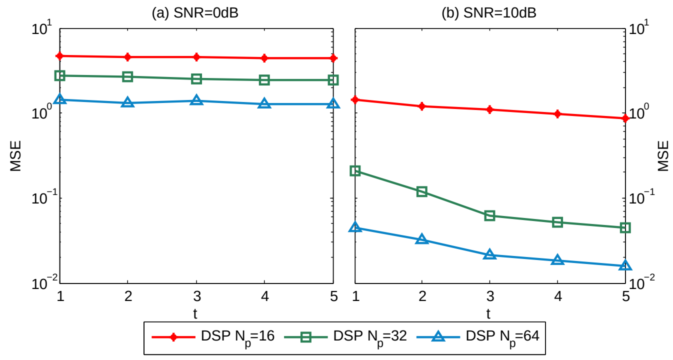

Fig. 1 MSE comparisons of DSP under a varying number of iterations t.

Next, we set $N_p = 32$ and present the MSE comparison of the proposed DSP scheme under a varying number of selected data N_g in Fig.2. Moreover, the structured SP (SSP) algorithm is applied for comparison [2]. It is obvious that the proposed method outperforms the SSP although only few data are used (i.e., $N_g \leq 16$) to assist the channel estimation. In addition, we observe that the improvement of the proposed method increases with N_g and tends to flatten out when $N_g > 32$. Since the computational complexity of DSP also increases with N_g (please refer to Section 4.2 for the details), we choose $N_g = 32$ for the following simulation, considering the tradeoff between the channel estimation performance and computational complexity.

In Fig.3 we present the MSE comparison of DSP under a varying number of pilots N_p. From this figure, we observe that the proposed DSP algorithm has substantial performance gain over the SSP. Specifically, when $N_p = 32$ and $MSE = 10^{-2}$ are considered, DSP is superior to SSP by over 5 dB. Meanwhile, the curve of DSP with $N_p = 32$ is very close to the curve of SSP with $N_p = 64$ when $SNR \geq 10$dB, implying a drastic reduction of pilot overhead at roughly 50% that DSP could offer. Moreover, for the proposed DSP method, $N_p = 32$ pilots only occupy 1.7% subcarriers of the total $N = 2048$ subcarriers. For comparison, with similar estimation accuracy, the pilot occupancy of joint subspace pursuit (JSP) method is 4.6% [7], where data decisions are not exploited in channel estimation.

6 CONCLUDING REMARKS

This paper considers the challenging problem of the uplink channel estimation for massive MIMO system. Decision aided subspace pursuit recovery algorithm

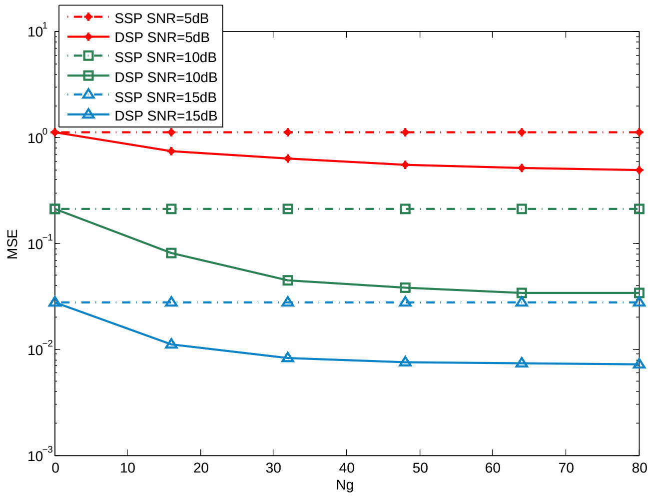

Fig. 2 MSE comparisons of SP and DSP under $N_p = 32$ and a varying number of selected data N_g.

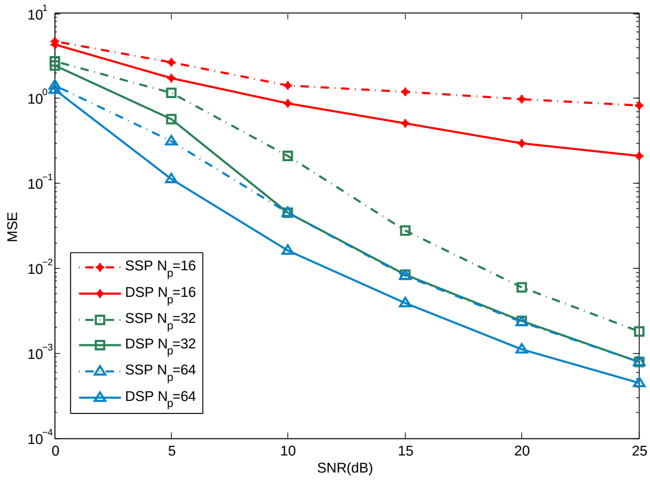

Fig. 3 MSE comparisons of SP and DSP under a varying number of pilots N_p.

is designed to iteratively exploit the detected data to enhance the channel estimation performance. We analyze the normalized mean absolute error and the computational complexity with the proposed scheme. Simulation results show that the proposed DSP could improve the channel estimation performance by over 5 dB compared with the SSP algorithm, without sacrificing the spectral efficiency.

Acknowledgements This work has been supported by “the National Natural Science Foundation of China (61201199)” and “the Fundamental Research Funds for the Central Universities (2014YJS006)”.

Appendix: Proof of Theorem 1

First, equation (1) can be rewritten as

$$\boldsymbol{Y} = \frac{1}{\sqrt{N_p}} \Phi \boldsymbol{H} + \boldsymbol{N} = \Theta \boldsymbol{H} + \boldsymbol{N}, \tag{7}$$

where $\boldsymbol{Y} = [\frac{1}{\sqrt{N_p}}\boldsymbol{y}_1^{\Omega}, \cdots, \frac{1}{\sqrt{N_p}}\boldsymbol{y}_M^{\Omega}]$, $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_M]$ and $\boldsymbol{N} = [\frac{1}{\sqrt{N_p}}\eta_1, \cdots, \frac{1}{\sqrt{N_p}}\eta_M]$. The term $\frac{1}{\sqrt{N_p}}$ is to normalize the measurement matrix Θ to satisfy $tr(\Theta^H \Theta) = K$, so as to fit into the analytical framework of compressive sensing [10].

Note that $\hat{\boldsymbol{H}} = \Theta_{\hat{S}}^{\dagger} \boldsymbol{Y}$ according to Line 19 in Algorithm 1, and $\boldsymbol{Y} = \Theta \boldsymbol{H} + \boldsymbol{N} = \Theta_{\hat{S}} \boldsymbol{H}_{\hat{S}} + \Theta_{S-\hat{S}} \boldsymbol{H}_{S-\hat{S}} + \boldsymbol{N}$, where S and $\hat{S}$ are the support and the estimated support of $\boldsymbol{H}$ respectively [11], we have

$$\begin{aligned} \mathbb{E}\Big(\frac{\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_F}{\|\boldsymbol{H}\|_F}\Big) \le\, & \underbrace{\mathbb{E}\Big(\frac{\|\boldsymbol{H}_{S-\hat{S}}\|_F}{\|\boldsymbol{H}\|_F}\Big)}_{I_1} \\ & + \underbrace{\mathbb{E}\Big(\frac{\|\Theta_{\hat{S}}^{\dagger}\Theta_{S-\hat{S}}\boldsymbol{H}_{S-\hat{S}}\|_F}{\|\boldsymbol{H}\|_F}\Big)}_{I_2} + \underbrace{\mathbb{E}\Big(\frac{\|\Theta_{\hat{S}}^{\dagger}\boldsymbol{N}\|_F}{\|\boldsymbol{H}\|_F}\Big)}_{I_3}, \end{aligned} \tag{8}$$

where $\|\cdot\|_F$ is the F-norm operation.

Firstly, for the term I_1 inside the summation in (8), we can easily obtain $\frac{\|\boldsymbol{H}_{S-\hat{S}}\|_F}{\|\boldsymbol{H}\|_F} \le 1$ since $\|\boldsymbol{H}_{S-\hat{S}}\|_F$ is always smaller than $\|\boldsymbol{H}\|_F$.

Next, according to the restricted isometry property (RIP) in [10], we have $\|\Theta_{\hat{S}}^{\dagger}\Theta_{S-\hat{S}}\| \le \frac{\delta_{2K}}{1-\delta_K}$, where δ_K and δ_{2K} are RIP constants that only relate to the sparsity K. Hence,

$$I_2 \le \mathbb{E}\Big(\frac{\|\boldsymbol{H}_{S-\hat{S}}\|_F}{\|\boldsymbol{H}\|_F}\frac{\delta_{2K}}{1-\delta_K}\Big) \le \frac{\delta_{2K}}{1-\delta_K} \tag{9}$$

Finally, note that $\Theta_{\hat{S}}^{\dagger}$ is a partial discrete Fourier transform matrix and has at most K non-zero singular values with upper bound of $\frac{1}{\sqrt{1-\delta_k}}$ according to the RIP property [10]. Moreover, since $\boldsymbol{N}$ and $\boldsymbol{H}$ are i.i.d complex Gaussian distributed with variance ν and 1 respectively, we have $\|\boldsymbol{N}\|_F \sim \nu/2 \cdot \chi^2(2K)$ and $\|\boldsymbol{H}\|_F \sim 1/2 \cdot \chi^2(2N_p)$ where $\chi^2(2K)$ and $\chi^2(2N_p)$ denotes the chi-square distribution with $2K$ and $2N_p$ degrees of freedom, respectively. Thus [10],

$$\begin{aligned} I_3 = \mathbb{E}\Big(\frac{\|\Theta_{\hat{S}}^{\dagger}\boldsymbol{N}\|_F}{\|\boldsymbol{H}\|_F}\Big) &\le \sqrt{\frac{2MK\nu}{N_p(1-\delta_K)}}\mathbb{E}\Big((\chi_{2M})^{-\frac{1}{2}}\Big) \\ &\le \sqrt{\frac{MK\nu}{N_p(1-\delta_K)}}\frac{\Gamma(M-\frac{1}{2})}{\Gamma(M)} \end{aligned}. \tag{10}$$

Applying (8), (9) and (10) to (7), we have

$$\begin{aligned}\mathbb{E}\left(\frac{\|\boldsymbol{H}-\hat{\boldsymbol{H}}\|_F}{\|\boldsymbol{H}\|_F}\right) &\leq 1+\frac{\delta_{2K}}{1-\delta_K}+\sqrt{\frac{MK\nu}{N_p(1-\delta_K)}}\frac{\Gamma(M-\frac{1}{2})}{\Gamma(M)} \\ &\overset{(a)}{\leq} \frac{1}{1-\delta_{2K}}+\sqrt{\frac{MK\nu}{N_p(1-\delta_{2K})}}\frac{\Gamma(M-\frac{1}{2})}{\Gamma(M)}\end{aligned}, \tag{11}$$

where (a) is a consequence of the fact that $\delta_K \leq \delta_{2K} < 1$ [11].

This completes the proof of the Theorem 1.

References

1. T. L. Marzetta. (2010). Noncooperative cellular wireless with unlimited number of base station antenns. *IEEE Transactions on Wireless Communications*, 9(11), 3590-3600.
2. Z. Gao, L. Dai, and Z. Wang. (2014). Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems. *Electornics Letters*, 50(12), 896-898.
3. Y. Nan, L. Zhang, X. Sun. (2016). An Efficient Downlink Channel Estimation Approach for TDD Massive MIMO Systems [C]. *IEEE Vehicular Technology Conference (VTC)*, 1-5.
4. W. Hou, C. Lim. (2014). Structured compressive channel estimation for large-scale MISO-OFDM systems. *IEEE Communications Letters*, 18(5), 765-768.
5. Y. Nan, L. Zhang, X. Sun. (2015). Efficient Downlink Channel Estimation Scheme Based on Block-Structured Compressive Sensing for TDD Massive MU-MIMO Systems. *IEEE Wireless Communication Letters*, 4(4), 1-4.
6. Y. Nan, L. Zhang, X. Sun. (2016). Weighted Compressive Sensing Based Uplink Channel Estimation for TDD Massive MIMO Systems. *IET Communications.* [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-com.2016.0625.
7. B. Lv, Z. Yang, Y. Feng. (2015). Temporally and spatially correlated uplink channel estimation for massive MIMO systems. *Wireless Communications and Signal Processing (WCSP), 2015 International Conference on*, Nanjing, 1-5.
8. R. Wang, Y. Chen, H. Tan. (2014). Data-Assisted Channel Estimation for Uplink Massive MIMO Systems. *Global Communications Conference (GLOBECOM)*, 3766-3771.
9. M. Masood, L. Afify, T. Al-Naffouri. (2015). Efficient Coordinated Recovery of Sparse Channels in Massive MIMO. *IEEE Transactions on Signal Processing*, 63(1), 104-118.
10. X. Rao, V. Lau. (2014). Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems. *IEEE Transactions on Signal Processing*, 62, 3261-3271.
11. W. Dai, O. Milenkovic. (2009). Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Transactions on Information Theory*, 55(5), 2230-2249.
12. J. Ma, P. Li. (2014). Data-aided channel estimation in large antenna systems. *IEEE Transactions on Signal Processing*, 4626-4631.