Proceedings Paper:

Nan, Y, Sun, X and Zhang, L (2015) Near-optimal pilot allocation in sparse channel estimation for massive MIMO OFDM systems. In: Proceedings IEEE International Symposium on Wireless Communication Systems (ISWCS 2015). Twelfth International Symposium on Wireless Communication Systems, 25-28 Aug 2015, Brussels, Belgium. IEEE . ISBN 9781467365406

https://doi.org/10.1109/ISWCS.2015.7454372

Near-optimal pilot allocation in sparse channel estimation for massive MIMO OFDM systems

Yang Nan
Beijing Jiaotong University, 100044
Email: ynanbjtu@hotmail.com

Xin Sun
Beijing Jiaotong University,100044
Email: xsun@bjtu.edu.cn

Li Zhang
University of Leeds, LS2 9JT
Email: l.x.zhang@Leeds.ac.uk

***Abstract*—Inspired by the success in sparse signal recovery, compressive sensing has already been applied for the pilot-based channel estimation in massive multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) systems. However, little attention has been paid to the pilot design in the massive MIMO system. To obtain the near-optimal pilot placement, two efficient schemes based on the block coherence (BC) of the measurement matrix are introduced. The first scheme searches the pilot pattern with the minimum BC value through the simultaneous perturbation stochastic approximation (SPSA) method. The second scheme combines the BC with probability model and then utilizes the cross-entropy optimization (CEO) method to solve the pilot allocation problem. Simulation results show that both of the methods outperform the equispaced search method, exhausted search method and random search method in terms of mean square error (MSE) of the channel estimate. Moreover, it is demonstrated that SPSA converges much faster than the other methods thus are more efficient, while CEO could provide more accurate channel estimation performance.**

***Keywords*—Massive MIMO, optimal pilot allocation, compressive sensing.**

I. Introduction

The combination of massive (or large-scale) multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) is expected to be the key technology for future communications [1]. However, this technique suffers from high complexity load in channel estimation due to the high dimensional channel matrix to be estimated. To solve this problem, more and more attentions have been focused on the sparse recovery algorithms in compressive sensing (CS), which could estimate the sparse channel by relatively less pilots [2]. One of the challenges in sparse channel estimation is pilot pattern design, which significantly affects the channel estimation performance. To optimize the pilot allocation, a promising idea is to utilize the restricted isometry property (RIP) [3], which demonstrates that random sampling may guarantee the sparse recovery with a decent probability, indicating that an optimal way to design the pilot pattern is random search. However, this method encounters difficulties in implementation due to its large storage and low efficiency [4].

In [5] and [6], two near-optimal pilot allocation schemes based on the mutual incoherence property (MIP) are proposed. In [7] a pilot allocation method based on the mean square error (MSE) is proposed. However, all of these schemes are based on the single-input-single-output (SISO) OFDM. To the best of the authors' knowledge, few works have been done on the near optimal pilot allocation for the massive MIMO OFDM systems.

In this paper, we investigate the block coherence (BC) of the measurement matrix of the massive MIMO systems [8] and then propose two efficient pilot allocation methods for the sparse channel estimation. The first method obtains the near-optimal pilot pattern by minimizing the BC through the simultaneous perturbation stochastic approximation (SPSA) method [9]. The second method combines BC with a probability model and then employs the cross-entropy optimization (CEO) [10] to solve the minimization problem of BC. The proposed schemes are compared with the equispaced search method, exhaustive search method and random search method. Simulation results show that the SPSA has advantage in convergence speed while the CEO is superior to the others in terms of the BC performance and the mean square error (MSE) performance.

The rest of the paper is organized as follows. We first describe the massive MIMO OFDM system model and define the problem in Section II. Then the proposed pilot design methods are addressed in Section III. Numerical experiments are presented in Section IV. Finally, section V concludes the paper.

The notations used in this paper are concluded as follows. Matrices and vectors are denoted by symbols in bold letters. Operators $(\cdot)^T$, $\lceil\cdot\rceil$, $|\cdot|$ and $\|\cdot\|$ represent transpose, ceiling function, complex modulus and norm function, respectively.

II. System Model and Problem Definition

Considering a downlink massive MIMO OFDM system where the receiver and transmitter are equipped with N_r and N_t antennas, respectively. The ith MIMO OFDM symbol for the kth transmitting antenna ($1 \le k \le N_t$) is composed of N subcarriers, among which N_p subcarriers are used to transmit the pilot signals. The transmitting pilots of the kth antenna can be expressed as $\boldsymbol{x}_i^{(k)} = [x_i^{(k)}(P_1), x_i^{(k)}(P_2), \cdots, x_i^{(k)}(P_{N_p})]$ where $\boldsymbol{p} = [P_1, P_2, \cdots, P_{N_p}]$ ($1 \le P_1 < P_2 < \cdots < P_{N_p} \le N$) contains the corresponding indices of the pilots. To distinguish the channels corresponding to different transmit antennas at the receiver, the pilot sequence $\boldsymbol{x}_i^{(k)}$ ($1 \le k \le N_t$) is generated respectively, based on the identically and independently distributed (i.i.d) random Bernoulli distribution (± 1) [11]. Therefore, we have $\boldsymbol{x}_i^m \ne \boldsymbol{x}_i^n$ if $m \ne n$.

In wireless communications, the channel impulse response (CIR) $\boldsymbol{h}_i^{(k)}$ of the ith OFDM symbol between the kth transmitting antenna and a certain receiving antenna can be represented as

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

978-1-4673-6540-6/15/$31.00 ©2015 IEEE

where L is the maximum channel spread. Note that identical channel estimation processing will be adopted on every receiving antenna, so we have omitted the index of receiving antennas in this paper. On the receiver side, we have

$$\begin{aligned} \boldsymbol{y}_i &= \sum_{k=1}^{N_t} diag\{\boldsymbol{x}_i^{(k)}\}\boldsymbol{F}_{p,L}\boldsymbol{h}_i^{(k)} + \boldsymbol{w}_i^{(k)} \\ &= \sum_{k=1}^{N_t} \boldsymbol{D}_i^{(k)}\boldsymbol{F}_{p,L}\boldsymbol{h}_i^{(k)} + \boldsymbol{w}_i^{(k)}, \end{aligned} \tag{1}$$

where $\boldsymbol{y}_i$ is the ith received pilot sequence coming from N_t different transmit antennas, $\boldsymbol{D}_i^{(k)} = diag\{\boldsymbol{x}_i^{(k)}\}$ is a diagonal matrix with $\boldsymbol{x}_i^{(k)}$ 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, $\boldsymbol{w}_i^{(k)}$ denotes the additive white Gaussian noise (AWGN). For the sake of brevity, we hereafter omit part of the scripts in (1) and rewrite (1) as [11]

$$\boldsymbol{y} = \Psi\tilde{\boldsymbol{h}} + \boldsymbol{w}, \tag{2}$$

where $\Psi = [\boldsymbol{D}_i^{(1)}\boldsymbol{F}_{p,L}, \boldsymbol{D}_i^{(2)}\boldsymbol{F}_{p,L}, \cdots, \boldsymbol{D}_i^{(N_t)}\boldsymbol{F}_{p,L}]$ of size $N_p \times N_t L$ and $\tilde{\boldsymbol{h}} = [(\boldsymbol{h}_i^{(1)})^T, (\boldsymbol{h}_i^{(2)})^T, \cdots, (\boldsymbol{h}_i^{(N_t)})^T]^T$ is the CIR vector of size $N_T L \times 1$. Moreover, we rearrange $\tilde{\boldsymbol{h}}$ as $\boldsymbol{b} = [\boldsymbol{b}_0^T, \cdots, \boldsymbol{b}_l^T, \cdots, \boldsymbol{b}_{L-1}^T]^T$ with $\boldsymbol{b}_l = [h_{l,i}^{(1)}, \cdots, h_{l,i}^{(k)}, \cdots, h_{l,i}^{(N_t)}]$. Then the system can be modeled by

$$\boldsymbol{y} = \sum_{l=0}^{L-1} \boldsymbol{A}_l \boldsymbol{b}_l + \tilde{\boldsymbol{w}}_l = \boldsymbol{A}\boldsymbol{b} + \tilde{\boldsymbol{w}}, \tag{3}$$

where $\boldsymbol{A}_l = [\Psi_l, \Psi_{L+l}, \cdots, \Psi_{(N_t-1)L+l}]$ is the lth $N_p \times N_t$ block (submatrix) of Ψ with Ψ_l the lth column of Ψ, $\boldsymbol{A} = [\boldsymbol{A}_0, , \cdots, \boldsymbol{A}_l, \cdots, \boldsymbol{A}_{L-1}]$. In the same way, $\tilde{\boldsymbol{w}}_l$ is the lth block of the reordered AWGN $\tilde{\boldsymbol{w}}$.

It has been proved that the wireless channel of massive MIMO system is sparse in nature, where the significant paths (non-zero taps) of the channel, denoted as S, is much smaller than the maximum channel spread L, e.g., $S \ll L$. Moreover, it is shown recently in [11] that the CIR of different transmit antennas shares a common support in downlink massive MIMO systems since the antenna spacing at the BS side is negligible compared to the long distance between BS and the receivers. In other words, although the paths amplitudes and phases may be distinct, the path delays of different transmit antennas are identical, e.g.,

$$supp(\boldsymbol{h}_i^{(m)}) = supp(\boldsymbol{h}_i^{(n)}), m \neq n. \tag{4}$$

As a result, we can jointly consider the channel estimation problem of massive MIMO systems by exploiting their joint sparsity.

According to the recent progress in CS, a necessary condition for successfully reconstructing $\boldsymbol{b}$ from $\boldsymbol{y}$ is that measurement matrix $\boldsymbol{A}$ satisfies the RIP [8]. However, it is computationally complex to validate whether a given matrix could satisfy the RIP. To solve this problem, former works have utilized the MIP of $\boldsymbol{A}$ as an alternative, which obtains the near-optimal pilot allocation by minimizing the mutual coherence of $\boldsymbol{A}$, denoted as [12]

$$\boldsymbol{p}_{opt} = \arg\min_{\boldsymbol{p}} \mu(\boldsymbol{p}), \tag{5}$$

where $\boldsymbol{p}_{opt}$ denotes the near-optimal pilot pattern, and

$$\mu(\boldsymbol{p}) = \max_{1\le m<n\le L} |\langle \boldsymbol{a}_m, \boldsymbol{a}_n \rangle|, \tag{6}$$

where $\boldsymbol{a}_m$ is the mth column of $\boldsymbol{A}$, and $\langle\cdot\rangle$ is the correlation function.

In traditional SISO or MIMO systems, orthogonal pilot pattern is always adopted, which could be optimized according to the MIP. However, the superimposed pilot pattern based on the block CS [11] becomes more popular in massive MIMO systems since the overhead of orthogonal pilot pattern increases with the number of transmit antennas, which reduces the spectral efficiency significantly in the massive MIMO system. Note that MIP is no longer applicable for the superimposed pilot pattern, we exploit the block coherence as an alternative in this paper.

III. Pilot design methods

A. Analysis of block coherence

We first define the block coherence of $\boldsymbol{A}$ as [8]

$$\mu_B(\boldsymbol{p}) = \frac{1}{N_t} \max_{0\le m<n\le L-1} \rho(\boldsymbol{A}_m^H \boldsymbol{A}_n), \tag{7}$$

where $\rho(\boldsymbol{X})$ is the spectrum norm of a given matrix $\boldsymbol{X}$ defined as $\rho(\boldsymbol{X}) = \lambda_{max}^{1/2}(\boldsymbol{X}^H\boldsymbol{X})$, with $\lambda_{max}^{1/2}(\boldsymbol{X}^H\boldsymbol{X})$ denoting the largest eigenvalue of the positive semi-definite matrix $\boldsymbol{X}^H\boldsymbol{X}$. It is obvious that computing the spectrum norm of $\boldsymbol{A}_m^H\boldsymbol{A}_n$ is computationally expensive. To solve this problem, we need to simplify the expression in (7). Note that $|x_i^{(k)}(P_j)| = 1$ $(1 \le j \le N_p)$ and

$$\boldsymbol{F}_{p,L} = \begin{pmatrix} 1 & \varphi_{P_1} & \cdots & \varphi_{P_1\times(L-1)} \\ 1 & \varphi_{P_2} & \cdots & \varphi_{P_2\times(L-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \varphi_{P_{N_p}} & \cdots & \varphi_{P_{N_p}\times(L-1)} \end{pmatrix},$$

where $\varphi_\kappa = e^{-j2\kappa\pi/N}$, we have

$$\begin{aligned} \mu_B(\boldsymbol{p}) &= \frac{1}{N_t} \max_{0\le m<n\le L-1} \rho(\boldsymbol{A}_m^H \boldsymbol{A}_n) \\ &\le \frac{1}{N_t} \max_{0\le m<n\le L-1} |\Psi_m^H \Psi_n| \qquad (8) \\ &= \frac{1}{N_t} \max_{0\le m<n\le L-1} \sum_{j=1}^{N_p} |\varphi_{P_j\times(n-m)}| \\ &= \frac{1}{N_t} \max_{1\le d\le L-1} \sum_{j=1}^{N_p} |\varphi_{P_j\times d}|, \qquad (9) \end{aligned}$$

where $d = n - m$. Note that (8) is derived from $Ger\check{s}gorin's$ $disc\ theorem$ in [14, Corollary 6.1.5]. Therefore, we could obtain the near-optimal pilot allocation by solving the following optimization problem

$$\boldsymbol{p}_{opt} = \arg\min_{\boldsymbol{p}} \mu_B(\boldsymbol{p}). \tag{10}$$

B. Pilot allocation method based on the block coherence

By observing (9) and (10), it is obvious that the most intuitive approach to optimize the superimposed pilot pattern is to perform an exhaustive search over all the possible pilot patterns. However, the effort for computing $\mu_B(\boldsymbol{p})$ with different pilot patterns is prohibitive due to the huge number of ways to allocate N_p pilots in N subcarriers. For example, we have 5.82×10^{40} possible pilot allocations when $N_p = 32$ and $N = 256$ are considered. Therefore, we propose two efficient methods to search the near-optimal pilot pattern.

1) Simultaneous Perturbation Stochastic Approximation (SP-SA)

The simultaneous perturbation stochastic approximation (SPSA) algorithm has been proved to be efficient for solving the multivariate optimization problems [9]. By using a relatively small number of measurements of the objective function, SPSA could provide a desirable solution without direct reference to the first derivative of the objective function. As an alternate method for calculating the first derivative, it uses two performance function observations whose variables are simultaneously varied randomly to approximate the gradient during each iteration, regardless of the dimension of target signals. Due to this essential feature of SPSA, it could become a powerful and efficient tool for the pilot pattern design in massive MIMO systems.

Assume $\varrho^t = [\varrho_1^t, \cdots, \varrho_j^t, \cdots, \varrho_N^t]$ where ϱ_j^t denotes the probability of the jth subcarrier to be the pilot subcarrier (e.g., $j \in \boldsymbol{p}$) at the tth iteration. Starting from $\varrho_j^{t=0} = 0.5$ for $1 \leq j \leq N$, at the tth iteration we perturb all the elements in ϱ^{t-1} simultaneously according to a vector δ^t whose elements are generated following Bernoulli distribution. After that, we update ϱ^t as

$$\varrho^t = \varrho^{t-1} + \alpha^t \frac{\mu_B(\varrho^{t-1} + \beta^t\delta^t) - \mu_B(\varrho^{t-1} - \beta^t\delta^t)}{2\beta^t}(\delta^t)^{-1}, \tag{11}$$

where all the parameters are set empirically as follows [14]: $\alpha^t \triangleq \alpha/(B+t+1)^\gamma$ with $\alpha = 1$, $B = 1000$ and $\gamma = 0.602$; $\beta^t \triangleq \beta/(t+1)^\lambda$ is a positive time-varying constant where $\beta = 0.01$ and $\lambda = 0.101$; the inverse of $(\delta^t)^{-1}$ is equal to δ^t since δ^t follows Bernoulli distribution. The inverse function here is defined to be element-wise inverse. Moreover, since the probability is finite, we set 0 and 1 as the lower bound and upper bound of ϱ^t, respectively. The SPSA terminates when the maximum number of iterations t_{max} is reached. After that, we can obtain the near-optimal pilot allocation $\boldsymbol{p}_{opt}$ by collecting the N_p indices with the maximum probability in ϱ. The detailed algorithm is summarized in Algorithm 1.

Algorithm 1

Initialization:
$\varrho^0 = 0.5$, and $t = 1$.
while $t \leq t_{max}$ **do**
1. Generate δ^t following Bernoulli distribution,
2. Set $\alpha^t = \alpha/(B+t+1)^\gamma$ and $\beta^t = \beta/(t+1)^\lambda$,
3. Update ϱ^t according to (11).

end while
Obtain the near-optimal pilot allocation $\boldsymbol{p}_{opt}$ by collecting the N_p indices with the maximum probability in ϱ.
Output: $\boldsymbol{p}_{opt}$.

2) Cross Entropy Optimization (CEO)

Consider equation (10) as an minimization problem,

$$\boldsymbol{p}_{opt} = \min_{\boldsymbol{p}\in\Lambda} \mu_B(\boldsymbol{p}), \tag{12}$$

where $\boldsymbol{p}_{opt}$ is the near optimal pilot pattern and Λ is the set of pilot patterns with cardinality N_p. Firstly we define the probability distribution function (PDF) of $\boldsymbol{p}$ based on the Bernoulli distribution as

$$f(\boldsymbol{p};\boldsymbol{v}) = \prod_{n=1}^{N} (v_n)^{p_n}(1-v_n)^{1-p_n}, \tag{13}$$

where $\boldsymbol{v}$ is a probability vector contains N elements $\{v_n\}_{n=1}^N$, p_n equals to 0 or 1 with probability $\mathcal{P}(p_n = 0) = 1 - v_n$ or $\mathcal{P}(p_n = 1) = v_n$ respectively. In each iteration, we only concern the probability that $\mu_B(\boldsymbol{p})$ is lower than or equal to some real number Γ, which can be denoted as

$$\ell = \mathbb{P}_u(\mu_B(\boldsymbol{p}) \leq \Gamma) = \mathbb{E}_u\Big\{I\{\mu_B(\boldsymbol{p}) \leq \Gamma\}\Big\}, \tag{14}$$

where $\mathbb{P}_u$ and $\mathbb{E}_u$ represent the probability operator and expectation operator, respectively, and $I\{\cdot\}$ is a indicator function where $I\{\mu_B(\boldsymbol{p}) \leq \Gamma\} = 1$ if $\mu_B(\boldsymbol{p}) \leq \Gamma$ and $I\{\mu_B(\boldsymbol{p}) \leq \Gamma\} = 0$ if $\mu_B(\boldsymbol{p}) > \Gamma$. To solve (14) efficiently, we adopt the CEO method which makes adaptive changes to the probability vector $\boldsymbol{v}$ towards the direction of the theoretically optimal PDF $f(\boldsymbol{p};\boldsymbol{v}^*)$ where $\boldsymbol{v}^*$ is the optimal probability vector of $\boldsymbol{p}$ [10]. In details, we choose a sample quantile factor ρ, e.g., $\rho = 0.1$, and then proceed the algorithm as follows:

1. Updating Γ_t. For some $\boldsymbol{v}_{t-1}$ at the tth iteration, let Γ_t be the $(1-\rho)$ quantile of $\mu_B(\boldsymbol{p})$ under PDF $f(\boldsymbol{p};\boldsymbol{v}_{t-1})$. Thus we have

$$\mathbb{P}_{\boldsymbol{v}_{t-1}}(\mu_B(\boldsymbol{p}) \leq \Gamma_t) \geq 1-\rho,$$
$$\mathbb{P}_{\boldsymbol{v}_{t-1}}(\mu_B(\boldsymbol{p}) \geq \Gamma_t) \geq \rho.$$

To update Γ_t, we first draw a random sample $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_z$ from $f(\boldsymbol{p};\boldsymbol{v}_{t-1})$. After that we calculate their BC respectively and then sort them in ascending order. Thus we can get the quantile $\hat{\Gamma}_t$ easily from

$$\hat{\Gamma}_t = \mu_{\lceil(1-\rho)z\rceil}, \tag{15}$$

where $\{\mu_j\}_{j=1}^z$ is the BC in new order.

2. Updating $\boldsymbol{v}_t$. With $\hat{\Gamma}_t$ and $\boldsymbol{v}_{t-1}$, we can obtain the optimal $\boldsymbol{v}_t$ by solving the the following program [10]

$$\frac{1}{z}\sum_{j=1}^{z} I\{\mu_B(\boldsymbol{p}_j) \leq \hat{\Gamma}_t\}\nabla \ln f(\boldsymbol{p}_j;\boldsymbol{v}_t) = 0. \tag{16}$$

Note that as $\boldsymbol{p}$ is generated based on the Bernoulli distribution [7], we could obtain $\boldsymbol{v}_t$ by substituting (13) into (16) as

$$v_{t,n} = \frac{\sum_{n=1}^{N} I\{\mu_B(\boldsymbol{p}_n) \leq \hat{\Gamma}_t\} I\{p_n = 1\}}{\sum_{n=1}^{N} I\{\mu_B(\boldsymbol{p}_n) \leq \hat{\Gamma}_t\}}, \tag{17}$$

where $v_{t,n}$ is the nth element of $\boldsymbol{v}_t$ at the tth iteration.

Consequently, we can update the $v_{t,n}$ iteratively to raise the possibility of $\mu_B(\boldsymbol{p}) \leq \Gamma$. However, once a $v_{t,n}$ turns to 0 or 1, it will be fixed so forever, which is undesirable. One of the efficient ways to solve this problem is to update the $v_{t,n}$ from $v_{t-1,n}$ smoothly, e.g., $\boldsymbol{v}_t = \iota \times \boldsymbol{v}_t + (1-\iota) \times \boldsymbol{v}_{t-1}$, where ι is a small positive constant. This method is summarized in Algorithm 2.

Algorithm 2

Input: sample size z, sample quantile factor ρ, smoothing rate ι, and the maximum iterations t_{max}.

Initialization:

$v_{0,n} = N_p/N$ for $1 \leq n \leq N$, and $t = 1$.

while $t \leq t_{max}$ **do**

1. Generate z random samples $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_z$ based on PDF $f(\boldsymbol{p}; \boldsymbol{v}_{t-1})$,
2. Calculate the BC $\mu_B(\boldsymbol{p}_j)$ for each $1 \leq j \leq z$, and then sort them in ascending order,
3. Let $\hat{\Gamma}_t$ be the $(1-\rho)$ sample quantile as $\hat{\Gamma}_t = \mu_{\lceil (1-\rho)z \rceil}$,
4. Calculate $v_{t,n}$ according to (17), for $n = 1, 2, \ldots, N$,
5. Smooth out $\boldsymbol{v}_t$ by

$$\boldsymbol{v}_t = \iota \times \boldsymbol{v}_t + (1-\iota) \times \boldsymbol{v}_{t-1},$$

6. $t = t + 1$.

end while

Obtain the near-optimal pilot allocation $\boldsymbol{p}_{opt}$ by collecting the N_p indices with the maximum probability in $\boldsymbol{v}_t$.

Output: $\boldsymbol{p}_{opt}$.

IV. Simulation Results

In this section, simulation studies are conducted to investigate the performance of the proposed pilot design schemes. Consider a 16×16 MIMO configuration with $N = 256$ OFDM subcarriers, among which $G = 32$ subcarriers are used as pilot subcarriers. Suppose the maximum channel spread $L = 60$, the sparsity $S = 6$ while $\rho = 0.1$, $\iota = 0.8$ and $t_{max} = 10000$. Note that the BC values have been normalized for brevity. Moreover, all the simulations are performed using MATLAB 2012a, running on a standard computer with an Intel Core i3-2100 CPU at 3.10GHz and 4GB of memory.

In Fig. 1 we show the BC performance of the CEO method for different values of the sample size z. It is obvious that z has a significant influence on the BC performance. In details, the CEO with $z = 50$ has the poorest performance due to its limited sample size. Meanwhile we observe that $z = 150$ has a slight edge over $z = 100$, indicating CEO converges to its best performance with $z = 150$ when the number of iterations reaches 21. Therefore, we choose $z = 150$ for the CEO method in the following simulations.

Next, we evaluate the updating process of BC as a function of running time for both of the proposed methods in Fig. 2. Meanwhile, the conventional equispaced search method, exhaustive search method as well as random search method are included for comparison. It is observed in Fig. 2(a) that the SPSA converges much faster than the other methods. In details, it exceeds the best BC performance of random search could achieve in 1 second by using no more than 0.03 second. On the other hand, though CEO takes more time for convergence, it outperforms the other method in BC performances within 0.2 second. From Fig. 2(b) where the run time is set to 1000 seconds, we can see that similar performances are achieved for all the evaluated methods. Specifically, both of the proposed methods maintain their advantages over the conventional methods, which confirms the reliable performance of the proposed methods. Moreover, we observe that although the exhaustive search can obtain the optimal pilot pattern in theory, it is unable to find a desirable solution within the given time.

In Fig. 3, we compare the channel estimation performances using different pilot patterns. All the pilot patterns are obtained from the simulation in Fig. 2(b) where 1000 seconds are carried out by the corresponding pilot design methods. The structured subspace pursuit (SSP) is employed for the channel

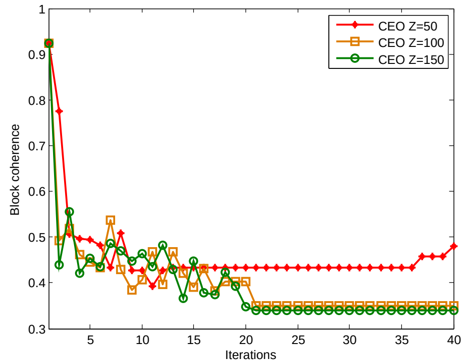
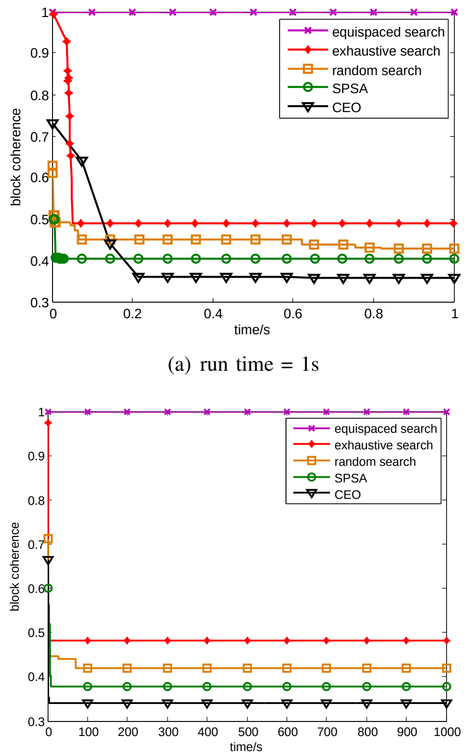


Fig. 1. Convergence of the CEO method for different values of z.



(a) run time = 1s

(b) run time = 1000s

Fig. 2. BC performance comparisons of different pilot design methods.

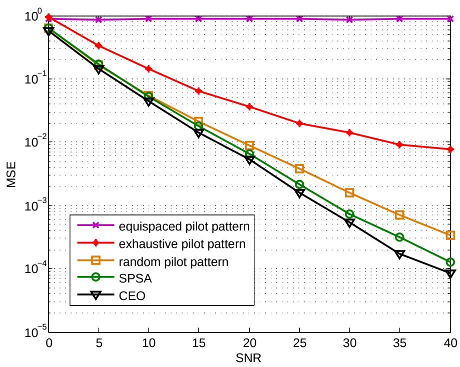


Fig. 3. MSE performance comparisons of channel estimation for different pilot design schemes.

estimation [11]. We observe from Fig. 3 that the pilot pattern obtained from CEO performs slightly better than that of SPSA, while both of the proposed schemes outperform the other conventional schemes. For example, at the MSE of $= 10^{-3}$, the SPSA outperforms the conventional schemes by more than 5.1 dB signal noise ratio (SNR) gain while CEO performs at least 6.4 dB better than those schemes. Moreover, it is noticed that there is no benefit using the equispaced pilot pattern for sparse channel estimation although it is proved to be the best choice in traditional channel estimation. This result is also consistent with its bad BC performance in Fig. 2.

V. Conclusion

In this paper, we have investigated the pilot allocation for sparse channel estimation in massive MIMO OFDM systems. Based on the BC, we have proposed two pilot design methods for searching the near-optimal pilot pattern off-line. Simulation results have shown that both of the proposed methods are superior to the conventional methods in terms of MSE performance. Moreover, it has demonstrated that SPSA converges much faster than the other methods while CEO provides better BC performance and thus MSE performance.

Acknowledgment

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

References

[1] G. Stuber, J. Barry, S. Mclaughlin, Y. Li, M. Ingram, and T. Pratt, "Broadband MIMO-OFDM wireless communications", *Proceeding of the IEEE*, vol. 92, no. 2, pp. 271-294, Feb. 2004.

[2] W. Ding, F. Yang, C. Pan, L. Dai, J. Song, "Compressive Sensing Based Channel Estimation for OFDM Systems Under Long Delay Channels", *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 313-321, Jun. 2014.

[3] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] C. Qi, G. Yue, L. Wu, Y. Huang and N. Arumugam, "Pilot Design Schemes for Sparse Channel Estimation in OFDM Systems", *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1493-1505.

[5] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems", *IEEE Signal Processing Letters*, vol. 18, no. 12, pp. 749-752, Dec. 2011.

[6] C. Qi and L. Wu, "A Study of Deterministic Pilot Allocation for Sparse Channel Estimation in OFDM Systems", *Communications Letters, IEEE* vol. 16, no. 5, pp. 742-744, May. 2012.

[7] J.C. Chen, C.K. Wen, and P. Ting, "An efficient pilot design scheme for sparse channel estimation in OFDM systems", *Commmunications Letters, IEEE* vol. 17, no. 7, pp. 1352-1355, Jul. 2013.

[8] Y. Eldar, P. Kuppinger and H. Bolcskei, "Block-space signals:uncertainty relations and efficient recovery", *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042-3054, Nov. 2010.

[9] J. Spall, "Stochastic optimization and the simultaneous perturbation method", *Simulation Conference Proceedings*, vol. 1, pp. 101-109, 1999.

[10] P. T. Boer, D. P. Kroese, S. Mannor and R. Y. Rubinstein, "A Tutorial on the Cross-Entropy Method", *Annals of Operations Research*, vol. 134, no. 1, pp. 19-67, Feb. 2005.

[11] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems", *Electronics Letters*, vol. 50, no. 12, pp. 896-898, Jun. 2014.

[12] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation", *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[13] R. Horn, C. Johnson, "Matrix Analysis", New York, NY:Cambridge Press, 1985.

[14] D. Zonoobi, A. Kassim, Y. Venkatesh, "Gini Index as Sparsity Measure for Signal Reconstruction from Compressive Samples", *IEEE Journal of Selected Topics in Signal Processing*, vol.5, no.5, pp.927-932 Sep. 2011