Channel Estimation via Oblique Matching Pursuit for FDD Massive MIMO Downlink

Minhyun Kim*, Junho Lee*, Gye-Tae Gil†, and Yong H. Lee*
*Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
†KAIST Institute for Information Technology Convergence (KI ITC), Daejeon, Korea
Email: {mhkim, jhlee}@stein.kaist.ac.kr, {gategil, yohlee}@kaist.ac.kr

***Abstract*—We consider channel estimation for massive multiple-input multiple-output (MIMO) systems operating in frequency division duplexing (FDD) mode. By exploiting the sparsity of significant propagation paths in massive MIMO channels, we develop a compressed sensing (CS) based channel estimator that can reduce the pilot overhead as compared with the conventional least squares (LS) and minimum mean square error (MMSE) estimators. The proposed scheme is based on the oblique matching pursuit (ObMP), an extension of the orthogonal matching pursuit (OMP), that can exploit prior information about the sparse signal vector. Given the channel covariance matrix, we obtain the incidence probability that each quantized angle coincides with the angle-of-departure (AoD) and use the incidence probability for deriving the oblique operator of the proposed scheme. The pilot sequence is designed to minimize the MSE of the oracle estimator. The simulation results demonstrate the advantage of the proposed scheme over various existing methods including the LS, MMSE and OMP estimators.**

I. Introduction

Multiple-input multiple-output (MIMO) systems with very large antenna arrays, called massive MIMO systems, have received considerable attention as a key technology for future mobile communication. In a massive MIMO cellular system, the number of base station (BS) antennas M greatly exceeds the number K of single antenna user terminals (UTs) ($M \gg K$). Compared to the conventional MIMO employing a small number of antennas with roughly equal M and K ($M \simeq K$), massive MIMO can provide huge improvements in network capacity and energy efficiency. Further, massive MIMO can achieve these improvements with simple precoders and combiners using the conjugate-transpose of the channel estimates [1]-[3].

One difficulty encountered in massive MIMO systems is caused by the need for estimating high dimensional channel matrix within a coherence time interval. To relieve this difficulty, most massive MIMO systems assume time division duplexing (TDD) based on channel reciprocity and estimate the channel only at BSs. The pilot training overhead for such systems is proportional to the number of UTs K (i.e., $O(K)$) which is much less than the number of BS antennas M. On the other hand, frequency division duplexing (FDD) dominates current cellular systems and exploring efficient channel estimation techniques for FDD massive MIMO is an important issue. (For the advantages and technical issues associated with FDD and TDD systems, refer [4].)

In FDD systems the downlink channel is estimated at every UT and the obtained channel state information (CSI) is fed back to the BS for precoding. If the conventional least-squares (LS) estimate is employed for the downlink channel estimation, the pilot overhead is $O(M)$ which can be prohibitively large in massive MIMO. Similarly, the CSI feedback becomes challenging due to large M. Recently, to overcome these difficulties in FDD massive MIMO, various downlink channel estimation and feedback techniques have been proposed. In [5], [6], pilot beams that can reduce the overhead of downlink training are designed based on both spatial and temporal channel correlations: the proposed techniques include the open-/closed-loop pilot training techniques with memory [5] and an optimal algorithm for sequentially designing pilot beams based on Kalman filtering [6]. In [7] a compressed sensing (CS) technique is developed for jointly estimating multiuser massive MIMO channels by exploiting the hidden joint sparsity of massive MIMO channels in angular domain where the channel is represented by a small number of significant paths. The pilot overhead for the CS-based technique can be $O(L \log M)$, where L is the number of significant paths, called the sparsity level [8]. For massive MIMO channels with $L \ll M$, this represents a substantial reduction of the overhead compared with the conventional LS method. The techniques for reducing the CSI feedback overhead include the antenna grouping-based method that maps multiple correlated antenna elements to a single representative value [9] and the CS-based methods using the Karhunen-Loève transform (KLT) [10]-[12].

In this paper, we develop a CS-based massive MIMO downlink channel estimator based on the oblique matching pursuit (ObMP) [13], [14], an extension of the orthogonal matching pursuit (OMP), that can exploit prior information about the sparse signal vector. The ObMP algorithm is identical to the OMP algorithm with the exception that an oblique operator is employed in the ObMP instead of the equivalent sensing matrix of the OMP. The proposed ObMP-based estimator is an extension of the grid-based OMP channel estimator employing a redundant dictionary consisting of array response vectors in the CS formulation [15] and uses the statistical covariance matrix of the channel at each UT to enhance its performance.

This work was supported by the ICT R&D program of MSIP/IITP [14-000-04-001, Development of 5G mobile communication technologies for hyper-connected smart services].

In this method, the oblique operator is designed using the probability that each angular grid coincides with the angle-of-departure (AoD). This probability, called the *incidence* probability, is evaluated from the channel covariance matrix. The pilot vectors are designed so that the mean square error (MSE) of the oracle estimator is minimized under the transmit power constraint [16]. In the simulation, we compare the proposed grid-based ObMP with the conventional minimum MSE (MMSE) method, the grid-based OMP, and the KLT-based OMP that has been employed for CSI feedback in massive MIMO [11]. The results show that the proposed ObMP outperforms the other methods in estimating massive MIMO downlink channels. They also show that both the grid-based ObMP and OMP outperform the KLT-based OMP, which suggests employing the grid-based CS methods instead of the KLT-based CS methods in the CSI feedback.

The organization of this paper is as follows. Section II presents the system and channel models. In Section III, we formulate the CS-based channel estimation problem and develop the proposed ObMP-based channel estimation. The process for designing the pilot sequence is also presented in Section III. Simulation results are presented in Section IV, and the conclusion is described in Section V.

Notations: Matrices are denoted by bold-faced uppercase letters and column vectors are denoted by bold-faced lower-case letters. Superscript $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the transpose, the conjugate transpose, and the inverse, respectively. $\mathrm{diag}(\cdot)$ is a diagonal matrix with diagonal elements inside the parenthesis, $\mathbb{E}[\cdot]$ is the expectation operator, and $\Pr(E)$ denotes the probability that an event E occurs. $\|\mathbf{A}\|_F$ is the Frobenius norm and $(\mathbf{A})_{i,j}$ denotes the (i,j)-th entry of the matrix $\mathbf{A}$. $\|\mathbf{a}\|_0$ and $\|\mathbf{a}\|$ are the ℓ_0 and ℓ_2 norms, respectively, and $\mathbf{a}(i)$ denotes the i-th entry of the vector $\mathbf{a}$. $\mathbf{A}_{\mathcal{I}}$ is the sub-matrix formed by collecting the columns of $\mathbf{A}$ whose indices are in set $\mathcal{I}$. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$.

II. System Model

We consider a massive MIMO cellular system operating in FDD mode where the BS is equipped with M antennas and there are K single antenna UTs in the cell. The BS broadcasts a sequence of T pilot vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_T\}$, for $\mathbf{s}_t \in \mathbb{C}^M$, $\|\mathbf{s}_t\|^2 = P$, and $1 \leq t \leq T$ where P is the transmission power. Each UT collects the corresponding received signals and estimates its downlink channel which is assumed to be a flat block-fading channel. The received signal vector $\mathbf{y} = [y_1, \ldots, y_T] \in \mathbb{C}^T$ at a UT is written as

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{n} \tag{1}$$

where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_T]^H \in \mathbb{C}^{T \times M}$, $\mathbf{h} \in \mathbb{C}^M$ is the channel vector, and $\mathbf{n} \in \mathbb{C}^T$ is a noise vector with independent identically distributed (i.i.d.) entries having zero mean and variance σ^2. $\|\mathbf{S}\|_F^2 = PT$ is the total transmit power during the training period T. (Here we drop the index of UT for notational simplicity.)

We adopt the parametric channel model [3], [15]. Assuming the uniform linear array (ULA), the channel vector is represented as

$$\mathbf{h} = \sqrt{\frac{M}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}(\phi_l) \tag{2}$$

where L is the number of paths and α_l is the channel gain of the l-th path which is assumed to be complex Gaussian random variables with zero mean and variance σ_α^2. The angle ϕ_l is the azimuth AoD of the l-th path. The mean angle associated with the cluster of scatterers is uniformly distributed over $[-\pi/2, \pi/2)$, and the distribution of the difference between an AoD and its mean is Laplacian with angular standard deviation σ_{AS}. The vector $\mathbf{a}(\phi_l) \in \mathbb{C}^M$ is the array response vector defined as

$$\mathbf{a}(\phi_l) = \frac{1}{\sqrt{M}} \left[1, e^{-j\frac{2\pi}{\lambda} d \sin(\phi_l)}, \ldots, e^{-j\frac{2\pi}{\lambda}(N-1) d \sin(\phi_l)}\right]^T \tag{3}$$

where d is the antenna spacing and λ is the signal wavelength.

III. CS-based Channel Estimation

In this section, we first formulate the CS problem for estimating both the AoDs and the corresponding complex gains of the L paths in (2). Then the ObMP algorithm that can solve the CS problem is derived using the channel covariance matrix which is assumed to be known.

A. CS Problem Formulation

We first quantize the azimuth angles and define a set $\mathcal{G}$ of quantized angles, called the grid, $\mathcal{G} = \{\theta_1, \ldots, \theta_G\}$. The number of grids G is much larger than the number of paths L ($G \gg L$). Using $\mathcal{G}$, the channel $\mathbf{h}$ in (2) can be approximated as

$$\mathbf{h} \approx \mathbf{A}\mathbf{h}_{\mathrm{a}} \tag{4}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_G)] \in \mathbb{C}^{M \times G}$ is the matrix consisting of G array response vectors corresponding to the grids and $\mathbf{h}_{\mathrm{a}} \in \mathbb{C}^G$ is an L-sparse vector having L non-zero elements which represent the path gains of the quantized AoDs. Ignoring the quantization error, (1) can be rewritten as

$$\mathbf{y} = \mathbf{S}\mathbf{A}\mathbf{h}_{\mathrm{a}} + \mathbf{n} = \mathbf{X}\mathbf{h}_{\mathrm{a}} + \mathbf{n} \tag{5}$$

where $\mathbf{X} \triangleq \mathbf{S}\mathbf{A} \in \mathbb{C}^{T \times G}$. Estimating $\mathbf{h}_{\mathrm{a}}$ in (5) is a CS problem with the dictionary $\mathbf{A}$ and the sensing matrix $\mathbf{S}$ ($\mathbf{A}$ becomes redundant for $G > M$). Their product $\mathbf{X}$ is called the equivalent sensing matrix. The sparse vector $\mathbf{h}_{\mathrm{a}}$ can be estimated by the following optimization:

$$\hat{\mathbf{h}}_{\mathrm{a}} = \arg\min_{\mathbf{h}_{\mathrm{a}}} \|\mathbf{y} - \mathbf{X}\mathbf{h}_{\mathrm{a}}\|^2 \ \text{ s.t. } \ \|\mathbf{h}_{\mathrm{a}}\|_0 = L \tag{6}$$

where $\hat{\mathbf{h}}_{\mathrm{a}}$ denotes the estimate of $\mathbf{h}_{\mathrm{a}}$. Once $\hat{\mathbf{h}}_{\mathrm{a}}$ is obtained, the corresponding channel estimate is given by $\hat{\mathbf{h}} = \mathbf{A}\hat{\mathbf{h}}_{\mathrm{a}}$.

B. ObMP for Channel Estimation

The ObMP algorithm is summarized in Algorithm 1. This algorithm is identical to the conventional OMP with the exception that an operator $\mathbf{Q} \in \mathbb{C}^{T \times G}$, called the *oblique* operator, is employed in place of the equivalent sensing matrix $\mathbf{X}$. At each iteration, the column index of the oblique operator $\mathbf{Q}$ that is most strongly correlated with the residual $\mathbf{r}_{t-1}$ is chosen, where $\mathbf{r}_0 = \mathbf{y}$ (steps 3 and 4). The selected index is appended to $\mathcal{S}_{t-1}$ which is the set of estimated AoDs obtained for iterations $\{1, \ldots, t-1\}$ (step 5). The channel gains corresponding to the estimated AoDs are obtained by the LS approach (step 6) and the residual $\mathbf{r}_t$ is updated by subtracting the contributions of the chosen columns (step 7). This process is repeated until the value of $\|\mathbf{r}_{t-1} - \mathbf{r}_{t-2}\|^2$ falls below the predetermined threshold γ.

The oblique operator can be designed to satisfy a generalized restricted isometry property (RIP) [13]. Alternatively, in [14] the oblique operator is found via optimization of coherence properties of $\mathbf{Q}^H\mathbf{X}$, which allows incorporation of the probability that a non-zero entry appears in each position of the sparse vector. We denote this probability, called the incidence probability, by $w_j = \Pr\left(\mathbf{h}_\mathrm{a}(j) \neq 0\right)$, $j \in \{1, \ldots, G\}$. Then the optimization problem is stated as follows:

$$\min_{\mathbf{Q}} \sum_{i=1}^{G} \sum_{j=1}^{G} w_j \left((\mathbf{Q}^H \mathbf{X})_{ij} - \delta_{ij} \right) \tag{7}$$

where δ_{ij} is the unit-impulse function which is equal to 1 when $i = j$ and zero otherwise. This problem can be decomposed and solved for each column. The closed-form expression of the i-th column of the optimal $\mathbf{Q}$, denoted as $\mathbf{q}_i^*$, can be written as

$$\mathbf{q}_i^* = \frac{\left(\mathbf{X}\mathbf{W}\mathbf{X}^H\right)^{-1} \mathbf{x}_i}{\left\| \left(\mathbf{X}\mathbf{W}\mathbf{X}^H\right)^{-1} \mathbf{x}_i \right\|}, \quad i = 1, \ldots, G \tag{8}$$

where $\mathbf{W} = \mathrm{diag}\left(w_1, \ldots, w_G\right)$ and $\mathbf{x}_i$ is the i-th column of $\mathbf{X}$. Next we show that the incidence probabilities $\{w_j\}$ can be obtained from the channel covariance matrix $\mathbf{R}$. Note that

$$\mathbf{R} = \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] = \mathbf{A}\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right]\mathbf{A}^H \tag{9}$$

where the second equality follows from (4) ignoring the angle quantization error. In (9) the matrix $\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right] \in \mathbb{E}^{G \times G}$ is a diagonal matrix with $\{w_j\}$ on its diagonal as shown below.

Lemma 1. $\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right] = \bar{\sigma}_\alpha^2 \mathrm{diag}\left(w_1, \ldots, w_G\right) = \bar{\sigma}_\alpha^2 \mathbf{W}$ *where* $\bar{\sigma}_\alpha^2 = \frac{M}{L}\sigma_\alpha^2$ *and* σ_α^2 *is the variance of the path gain in (2).*

Proof. Let Ω_{ij} be the event that both θ_i and θ_j are the AoDs of the channel and Ω_{ij}^c be its complementary event. Then $\mathbb{E}\left[\mathbf{h}_\mathrm{a}(i)\mathbf{h}_\mathrm{a}(j)^*\right] = \mathbb{E}\left[\mathbf{h}_\mathrm{a}(i)\mathbf{h}_\mathrm{a}(j)^* | \Omega_{ij}\right] \Pr\left(\Omega_{ij}\right) + \mathbb{E}\left[\mathbf{h}_\mathrm{a}(i)\mathbf{h}_\mathrm{a}(j)^* | \Omega_{ij}^c\right] \Pr\left(\Omega_{ij}^c\right) = \mathbb{E}\left[\bar{\alpha}_i \bar{\alpha}_j^*\right] \Pr\left(\Omega_{ij}\right)$. The second equality holds because $\mathbf{h}_\mathrm{a}(i) = \bar{\alpha}_i = \sqrt{\frac{M}{L}}\alpha_i$ if θ_i is the AoD and zero, otherwise. Since the i-th and j-th paths are statistically independent, $\mathbb{E}\left[\bar{\alpha}_i \bar{\alpha}_j^*\right] \Pr\left(\Omega_{ij}\right) = \bar{\sigma}_\alpha^2\ w_i$ if $i = j$ and $\mathbb{E}\left[\bar{\alpha}_i \bar{\alpha}_j^*\right] \Pr\left(\Omega_{ij}\right) = 0$, if $i \neq j$. This completes the proof. ■

Since $\sum w_j = 1$, this lemma indicates that $\{w_j\}$ can be obtained by normalizing the diagonal entries of $\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right]$. Using this lemma and (9) in (8), the proposed oblique operator can be represented as

$$\mathbf{q}_i^* = \frac{\left(\mathbf{S}\mathbf{R}\mathbf{S}^H\right)^{-1} \mathbf{x}_i}{\left\| \left(\mathbf{S}\mathbf{R}\mathbf{S}^H\right)^{-1} \mathbf{x}_i \right\|}, \quad i = 1, \ldots, G. \tag{10}$$

We now obtain $\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right]$ under the assumption that $M = G$ and $\mathbf{A}$ is unitary, where the latter assumption can be justified by the fact that $\mathbf{A}$ approaches the discrete Fourier transform (DFT) matrix as M goes to infinity [6], [17]. Assuming that $\mathbf{A}$ is unitary for large M,

$$\mathbb{E}\left[\mathbf{h}_\mathrm{a}\mathbf{h}_\mathrm{a}^H\right] = \mathbf{A}^H\mathbf{R}\mathbf{A} = \mathbf{A}^H\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{A} \tag{11}$$

where $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ represents the eigenvalue decomposition. From Lemma 1 and (11), the probabilities $\{w_j\}$ are given by

$$\begin{aligned} w_j &= \frac{1}{\bar{\sigma}_\alpha^2} \mathbf{a}(\theta_j)^H \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \mathbf{a}(\theta_j) \\ &= \frac{1}{\bar{\sigma}_\alpha^2} \sum_{i=1}^{r} |\mathbf{a}(\theta_j)^H \mathbf{u}_i|^2 \lambda_i, \quad i = 1, \ldots, G \end{aligned} \tag{12}$$

where $\mathbf{u}_i$ and λ_i are the i-th eigenvector and eigenvalue, respectively, and r is the rank of $\mathbf{R}$. Using (12) in (10), we obtain the proposed oblique operator. Although the oblique operator is derived for $M = G$, it will be shown in the simulation that this operator enhances the estimation performance of massive MIMO systems with $M < G$.

Algorithm 1 ObMP based Channel Estimation

Require: $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{A}$
1: **Initialize:** $\mathcal{S}_0$: Empty set, $\mathbf{r}_0 = \mathbf{y}$ and $\mathbf{r}_{-1} = \mathbf{0}$, $t = 1$
2: **while** $\|\mathbf{r}_{t-1} - \mathbf{r}_{t-2}\|^2 > \gamma$ **do**
3: $\quad \mathbf{f} = \mathbf{Q}^H \mathbf{r}_{t-1}$
4: $\quad j = \arg\max_i |\mathbf{f}_i|$ where $\mathbf{f}_i$ is the ith entry of $\mathbf{f}$
5: $\quad \mathcal{S}_t = \mathcal{S}_{t-1} \cup \{j\}$
6: $\quad \mathbf{g}_t = \left(\mathbf{X}_{\mathcal{S}_t}^H \mathbf{X}_{\mathcal{S}_t}\right)^{-1} \mathbf{X}_{\mathcal{S}_t}^H \mathbf{y}$
7: $\quad \mathbf{r}_t = \mathbf{y} - \mathbf{X}_{\mathcal{S}_t}\mathbf{g}_t$
8: $\quad t = t + 1$
9: **end while**
10: $\hat{\mathbf{h}}_\mathrm{a}(\mathcal{S}_{t-1}) = \mathbf{g}_{t-1}$ and $\hat{\mathbf{h}}_\mathrm{a}(i) = 0$ for $i \notin \mathcal{S}_{t-1}$
11: **return** $\hat{\mathbf{h}} = \mathbf{A}\hat{\mathbf{h}}_\mathrm{a}$

C. Pilot Sequence Design

In our CS formulation, designing the pilot vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_T\}$ is equivalent to designing the sensing matrix $\mathbf{S}$ in (5). A class of sensing matrices that are popular in the CS area is the set of random matrices that satisfy the RIP with large probability [18]. Such random matrices have been employed in the CS-based channel estimation [7], [8] and the CSI feedback methods in [10]-[12]. On the other hand, some recent results in [16], [19]-[21] indicate that carefully designed sensing matrices can improve the CS performance. In this section, we design $\mathbf{S}$ under the transmission power constraint following the approach in [16] that can minimize the MSE of the oracle estimator. Since the BS broadcasts its pilot sequences to all UTs, it is difficult to consider the channel covariance of each UT. Thus, in our design the knowledge on the channel covariance is not exploited.

The objective function for optimizing the sensing matrix is given by

$$\min_{\mathbf{S}} \|\mathbf{B} - \mathbf{SA}\|_F^2 + \beta\|\mathbf{S}\|_F^2 \tag{13}$$

where $\mathbf{B} \in \mathbb{C}^{T\times G}$ is a semi-unitary matrix satisfying $\mathbf{BB}^H = \mathbf{I}_T$ and $\beta > 0$. In (13) the equivalent sensing matrix $\mathbf{SA} = \mathbf{X}$ is designed as close as possible to the design target $\mathbf{B}$, while minimizing the second term which is essentially the pilot transmission power. On the other hand, the transmission power constraint is given by

$$\|\mathbf{S}\|_F^2 = PT \tag{14}$$

where P is the transmit power of each pilot sequence. Due to the second term of (13), it is not possible to jointly consider the objective function (13) and the power constraint (14). (This is true because jointly considering (13) and (14) reduces the optimization to: $\min_{\mathbf{S}}\|\mathbf{B} - \mathbf{SA}\|_F^2$ subject to $\|\mathbf{S}\|_F^2 = PT$.) To avoid this difficulty, the objective function and the constraint are considered separately: the solution to (13) is given by

$$\mathbf{S} = \mathbf{BA}^H \left(\mathbf{AA}^H + \beta\mathbf{I}_M\right)^{-1}, \tag{15}$$

and each column of $\mathbf{S}$ is normalized to meet $\|\mathbf{s}_t\|^2 = P$ for all $t = 1, \ldots, T$.

IV. Simulation Results

In this section, we compare the proposed grid-based ObMP (G-ObMP) with the conventional LS/MMSE estimation, the grid-based OMP (G-OMP) and the KLT-based OMP (KLT-OMP) through computer simulation. In the LS and MMSE estimation, the channel is estimated by evaluating $\mathbf{Fy}$ where $\mathbf{F} \in \mathbb{C}^{M\times T}$ is given by $\mathbf{F} = \mathbf{S}^H\left(\mathbf{SS}^H\right)^{-1}$ for the LS and $\mathbf{F} = \mathbf{RS}^H\left(\mathbf{SRS}^H + \sigma^2\mathbf{I}_T\right)^{-1}$ for the MMSE, where σ^2 is the noise variance in (1). For the LS and MMSE estimators, we use orthogonal pilot vectors with equal pilot power: $\mathbf{S} \in \{\mathbf{V}\colon \mathbf{V} \in \mathbb{C}^{T\times M},\ \mathbf{VV}^H = P\mathbf{I}_T\}$. The G-OMP is identical to the G-ObMP in Algorithm 1 with the exception that the equivalent sensing matrix $\mathbf{X}$ is used instead of the oblique operator $\mathbf{Q}$. The KLT-OMP employs the unitary dictionary $\mathbf{A}$ consisting of the eigenvectors of the correlation matrix $\mathbf{R}$ (in this case, $M = G$). To design the sensing matrices for the ObMP and OMP methods, we use (15)[1] with $\beta = 1$ (the grid-based ObMP and OMP employ an identical equivalent sensing matrix which is different from that for the KLT-OMP). The remaining parameters for the simulation are as follows. The number of antennas $M = 64$ and the number of grids $G = 180$. The channel has $L = 15$ paths with $\sigma_\alpha^2 = 1$ and the angular standard deviation $\sigma_{AS} = 10$. We set $\gamma = 0.1\sigma^2$ in step 2 of Algorithm 1. The signal-to-noise ratio (SNR) is defined as P/σ^2.

[1]The target matrix $\mathbf{B}$ is designed through the QR decomposition of a Gaussian random matrix.

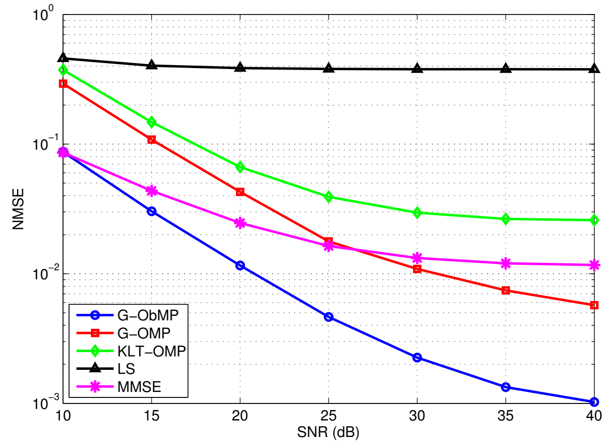

Fig. 1. NMSE performance against SNR when the training overhead $T = 40$.

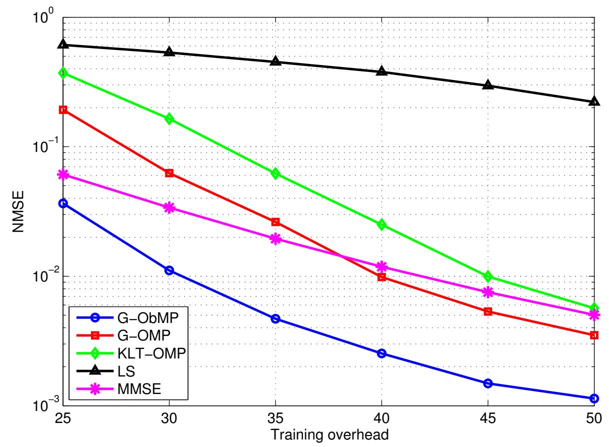

Fig. 2. NMSE performance against training overhead when SNR = 30dB.

To evaluate the performance of the estimators, we use the normalized mean square error (NMSE) defined as $10\log_{10}(\mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2/\|\mathbf{h}\|^2])$. Fig. 1 shows the NMSEs of the estimators against SNR when the training overhead $T = 40$. The G-ObMP outperforms the others, and the LS performs the worst. Comparing the G-OMP and KLT-OMP, the former performs better than the latter indicating that use of a redundant dictionary is beneficial to OMP based channel estimation. It is interesting to see that the G-OMP performs better than the MMSE in high SNR regime, even though the G-OMP does not need any information about the signal. By exploiting the sparsity of the channel, the grid-based methods can perform better than the conventional LS/MMSE estimators.

Fig. 2 compares the NMSEs against the training overhead when SNR = 30dB. Again, the G-ObMP performs the best and the LS performs the worst; the KLT-OMP performs worse than the G-OMP and the MMSE; the G-OMP behaves better than the MMSE only when $T > 35$. These results indicate that the training overhead can be reduced by employing the proposed ObMP.

V. Conclusion

The ObMP based channel estimation scheme for FDD massive MIMO downlink was proposed. The CS problem for estimating the channel was formulated and solved by the ObMP algorithm. The oblique operator has been designed using the probability that each angular grid coincides with the AoD, which is evaluated from the channel covariance matrix. The pilot vectors are designed so that the MSE of the oracle channel estimate is minimized under the transmission power constraint. It is shown through computer simulation that the proposed method outperforms the existing methods including G-OMP, KLT-OMP, and LS/MMSE estimators.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp 3590-3600, Nov. 2010.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742-757, Oct. 2014.

[4] P. W. C. Chan, E. S. Lo, R. R. Wang, E. K. S. Au, V. K. N. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, "The evolution path of 4G networks: FDD or TDD?," *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42-50, Dec. 2006.

[5] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802-814, Oct. 2014.

[6] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp 787-801, Oct. 2014.

[7] X. Rao, and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261-3271, Jun. 2014.

[8] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," in *Proc. IEEE*, vol. 98, no. 6, pp. 1058-1076, Jun. 2010.

[9] B. Lee, J. Choi, J. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback reduction for FDD-based massive MIMO systems," in *Proc. IEEE Int. Conf. on Commun.*, Sydney, Australia, Jun. 2014.

[10] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Paris, France, Apr. 2012.

[11] Y.-G. Lim and C.-B. Chae, "Compressed channel feedback for correlated massive MIMO systems," in *Proc. IEEE Int. Conf. on Commun.*, Sydney, Australia, Jun. 2014.

[12] J. Joung and S. Sun, "SCF: Sparse channel-state-information feedback using Karhunen-Loève transform," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014.

[13] K. Lee, Y. Bresler, and M. Junge, "Oblique pursuits for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 6111-6141, Sep. 2013.

[14] Y. Chi, and R. Calderbank, "Knowledge-enhanced matching pursuit," in *Proc. IEEE Int. Conf. Acoust., Speed, Signal Process.*, Vancouver, Canada, May 2013.

[15] J. Lee, G. Gil, and Yong H. Lee, "Exploiting spatial sparsity for estimation channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014.

[16] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016-2029, Apr. 2013.

[17] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441-6463, Oct. 2013.

[18] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[19] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695-5702, Dec. 2007.

[20] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300-4312, Sep. 2011.

[21] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887-2898, Jun. 2013.