Enhancing the Estimation of mm-Wave Large Array Channels by Exploiting Spatio-Temporal Correlation and Sparse Scattering

Saeid Haghighatshoar*, Giuseppe Caire*
*Communications and Information Theory Group, Technische Universität Berlin
Emails: saeid.haghighatshoar@tu-berlin.de, caire@tu-berlin.de

***Abstract*—In order to cope with the large path-loss exponent of mm-Wave channels, a high beamforming gain is needed. This can be achieved with small hardware complexity and high hardware power efficiency by Hybrid Digital-Analog (HDA) beamforming, where a very large number $M \gg 1$ of antenna array elements requires only a relatively small $m \ll M$ number of A/D converters and modulators/demodulators. As such, the estimation of mm-Wave MIMO channels must deal with two specific problems: 1) high Doppler, due to the large carrier frequency; 2) impossibility of observing directly the M-dimensional channel vector at the antenna array elements, due to the mentioned HDA implementation. In this paper, we consider a novel scheme inspired by recent results on gridless multiple measurement vectors problem in compressed sensing, that is able to exploit the inherent mm-Wave channel sparsity in the angular domain in order to cope with both the above problems simultaneously. Our scheme uses past pilot-symbol observations in a window of length T in order to estimate a low-dimensional subspace that approximately contains the channel vector at the current time. This subspace information can be used directly, in order to separate users in the spatial domain, or indirectly, in order to improve the estimate of the user channel vector from the current pilot-symbol observation.**

1 Introduction

Millimeter wave (mm-Wave) communication is a promising technology for the next generation of WLANs and outdoor cellular systems [1, 2]. In order to cope with the large path-loss exponent of mm-Wave channels, a high beamforming gain is needed. While large antenna arrays can be implemented with a small form factor due to the small wavelength, it is clear that conventional all-digital baseband processing as proposed for large MIMO systems at lower frequencies [3–5] is not a suitable solution here. In fact, because of the large signal bandwidth available at mm-Waves, the demodulation and quantization of the signal at each antenna array element would require an enormous A/D front-end bit-rate, with corresponding unacceptable hardware power consumption. For this reason, a promising approach for mm-Wave communication is the Hybrid Digital-Analog (HDA) beamforming, where the beamforming function is achieved in two stages. The first stage uses as analog reconfigurable beamforming network operating in the RF domain, and achieves beamforming gain and some coarser multiuser interference rejection while reducing the signal dimension from $M \gg 1$ (number of antenna array elements) to some $m \ll M$ (number of RF chains and A/D converters). The second stage, processes the m-dimensional baseband signal in the digital domain in order to achieve further multiuser MIMO spatial multiplexing gain [6, 7].

For multiuser spatial multiplexing, the base station needs to estimate the M-dimensional channel vectors of all the users. Channel estimation for mm-Wave MIMO channels must deal with two specific problems: 1) potentially rapid variations of the small-scale fading coefficients, due to the large carrier frequency; 2) impossibility of observing directly the M-dimensional channel vectors of the users at the antenna array elements, due to the mentioned HDA implementation. Fortunately, mm-Wave channels have a special feature that helps to cope with both the above problems simultaneously, namely, the resulting channel vectors are typically very sparse in the angular domain, since only the Line-of-Sight path and/or a few dominant multipath components convey significant power.

In this paper, we consider a novel scheme inspired by recent results on gridless multiple measurement vectors problem in compressed sensing, that exploits the inherent mm-Wave channel sparsity in the angular domain in order to cope with both the above problems. In this scheme, we exploit the past pilot-symbol observations in a window of length T in order to estimate a low-dimensional subspace that approximately contains the channel vector at the current time slot. This subspace information can be used directly, to separate users in the spatial domain, or indirectly, to improve the estimate of the user channel vector in the current time slot. Simulations show very encouraging preliminary results, and in particular confirm that the channel subspace information obtained over a window of past measurements provides significant improvements with respect to the conventional "one-shot" techniques, that estimate the channel vectors by using only the current pilot observation.

Notations: We denote vectors by boldface small letters (e.g., $\mathbf{x}$), matrices by boldface capital letters (e.g., $\mathbf{X}$), scalar constant by non-boldface letters (e.g., x or X), and sets by calligraphic letters (e.g., $\mathcal{X}$). The i-th element of a vector $\mathbf{x}$ and the (i, j)-th element of a matrix $\mathbf{X}$ will be denoted by $[\mathbf{x}]_i$ and $[\mathbf{X}]_{i,j}$ respectively. We denote the Hermitian and the transpose of a matrix $\mathbf{X}$ by $\mathbf{X}^{\mathsf{H}}$ and $\mathbf{X}^{\mathsf{T}}$, respectively. The same notation is used for vectors and scalars. We use $\mathbb{T}_+$ for the space of Hermitian semi-definite Toeplitz matrices. For an $\mathbf{x} \in \mathbb{C}^M$, we denote by $\mathbb{T}(\mathbf{x})$ a Hermitian Toeplitz matrix whose first column is $\mathbf{x}$. We always use $\mathbf{I}$ for the identity matrix, where the dimensions may be explicitly indicated for the sake of clarity (e.g., $\mathbf{I}_p$ denotes the $p \times p$ identity matrix). For an integer $k \in \mathbb{Z}$, we use the shorthand notation $[k]$ for the set of non-negative integers $\{0, 1, 2, \ldots, k-1\}$, where the set is empty if $k < 0$.

2 Model and Problem Statement

2.1 Channel Model

Motivated by mm-Wave channel measurements and models [2], we consider a simple propagation model for the wireless scattering channel in which the transmission between a single-antenna user and the M-antenna base-station array occurs through $p \ll M$ multipath components (see Fig. 1). The base-station is equipped with a *Uniform Linear Array* (ULA), with spacing $d = \frac{\lambda}{2\sin(\theta_{\max})}$ between its elements, with λ being the wave-length, and scans the angular range $[-\theta_{\max}, \theta_{\max}]$ for some $\theta_{\max} \in (0, \pi/2)$. We denote by $\mathbf{a}(\theta) \in \mathbb{C}^M$ the array response for the AoA $\theta \in [-\theta_{\max}, \theta_{\max}]$, whose k-th component, is given by $[\mathbf{a}(\theta)]_k = e^{jk\frac{2\pi d}{\lambda}\sin(\theta)} = e^{jk\pi\frac{\sin(\theta)}{\sin(\theta_{\max})}}$. We

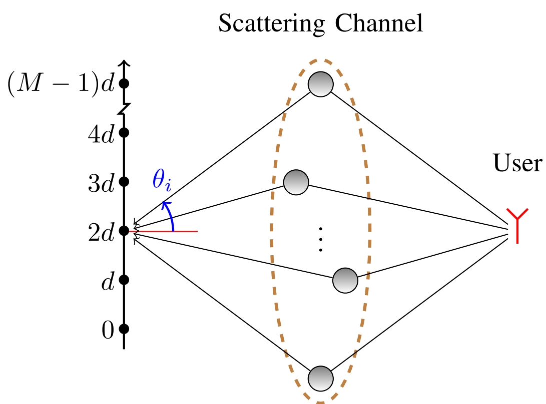

Fig. 1: Scattering channel with discrete angle of arrivals.

consider a discrete-time model, where the channel vector of a user at time t is given by

$$\mathbf{h}[t] = \sum_{\ell=1}^{p} w_\ell[t]\mathbf{a}(\theta_\ell), \tag{1}$$

where θ_ℓ denotes the angle-of-arrival (AoA) of the ℓ-th multipath component and where $w_\ell[t]$ is the corresponding small-scale fading coefficient, assumed $\sim \mathcal{CN}(0, \sigma_\ell^2)$. According to the well-known Wide-Sense Stationary Uncorrelated Scattering (WSSUS) model, the coefficients $w_\ell[t]$ are WSS processes with respect to t and mutually uncorrelated with respect to ℓ. The general wisdom of multiuser MIMO considers "one-shot" or "instantaneous" estimation [3]. This consists of partitioning the slot into a training phase and a data transmission phase. The channel vectors are estimated during the training phase, and these estimates are used in the data transmission phase. In compliance with most of the recent "massive MIMO" literature [5], we assume Time-Division Duplexing (TDD) and channel reciprocity [4], such that the channel vectors of the users are estimated during a training phase, in which orthogonal (uplink) pilot symbols are transmitted by the users to the base-station. The resulting estimates are used in data transmission phase to receive data streams transmitted simultaneously by the users to the base-station (uplink), or to transmit multiple data streams from the base-station to the users (downlink). In both cases, the data streams are separated in the spatial domain by linear beamforming (spatial multiplexing).

As anticipated in the introduction, in mm-Wave channels the "instantaneous" channel estimation may suffer from the fact that the mm-Wave channels change rapidly in time. Therefore, the ability of the beamformer to eliminate the multiuser interference in the spatial domain may be impaired by the "channel aging" phenomenon, i.e., by the time the channel estimate is used, the channel has already significantly changed. In addition, due to the discussed HDA implementation of the base-station front-end, the whole M-dimensional received signal in correspondence of the uplink pilot symbols cannot be fully observed. Rather, only an m-dimensional projection (or "sketch") through the analog beamforming network (consisting of m separate RF chains) is available.

While the channel vectors may change rapidly in time (up to the limit of having i.i.d. channels across different time slots), the WSS assumption implies that the scattering geometry, expressed by the AoA's $\{\theta_\ell\}_{\ell=1}^p$ and the multipath component strengths $\{\sigma_\ell^2\}_{\ell=1}^p$, remains invariant for a very large number of slots. This is justified by the fact that the "small-scale fading" channel gains $w_\ell[t]$ go through a full phase cycle when the distance between transmitter and receiver varies by one wavelength (e.g., 1cm at 30 GHz), whereas AoAs and path strengths change only when the "large-scale" geometry of the propagation between the transmitter and the receiver significantly changes.[1]

In this work, for the sake of clarity, we focus on the channel estimation problem of an individual user from uplink pilot symbols sent periodically with a period τ, and accumulated in an observation window of T slots, thus, in total there are $\nu = \frac{T}{\tau}$ training samples (see Fig. 2). The received signal at the

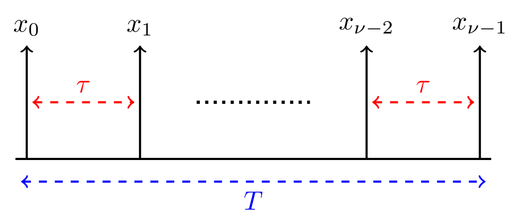

Fig. 2: Periodic pilot transmission for channel estimation.

i-th training period, $i \in [\nu]$, is given by $\mathbf{y}_i = \mathbf{h}_i + \mathbf{n}_i$, where $\mathbf{h}_i = \mathbf{h}[i\tau]$ denotes the random channel vector of the user (at time $t = i\tau$), and where $\mathbf{n}_i \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ is the additive white Gaussian noise of the array. We define the training *signal-to-noise-ratio* (SNR) by $\mathsf{snr} = \sum_{\ell=1}^p \sigma_\ell^2/\sigma^2$. Once an estimate of the channel vector $\mathbf{h}_i$ is available, it is used in the data transmission phase of the current slot $\mathcal{T}_i = [i\tau, (i+1)\tau - 1]$ to calculate the beamformer for the base-station receiver (uplink) and/or the base-station transmitter (downlink).

[1] Strictly speaking, according to the widely accepted Wide-Sense Stationary Uncorrelated Scattering (WSSUS) model, the second-order statistics of the channel vector process are time-invariant, implying that AoAs and signal strengths are strictly constant in time. As a mater of fact, the WSSUS model is a local approximation, with coherence time much larger than the small-scale fading coherence time.

2.2 One-Shot Sparse Channel Estimation

Since by assumption we have $p \ll M$, the channel vector $\mathbf{h}_i$, $i \in [\nu]$, has a sparse representation in the continuous dictionary $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in [-\theta_{\max}, \theta_{\max}]\}$ consisting of the array responses for different AoAs θ, with the sparsity being $\frac{p}{M} \ll 1$. Classical compressed sensing (CS) methods [8, 9] can be used to estimate $\mathbf{h}_i$ via a few, say $m \ll M$, linear projections of the received signal $\mathbf{y}_i$ rather than the whole components thereof. This feature is well-suited for the HDA front-end implementation that supports a number of RF chains and A/D converters much smaller than the number of array elements. Let us denote the $m \times M$ measurement projection matrix by $\mathbf{B}$, where we assume that the rows of $\mathbf{B}$ are orthonormal[2]. Also let $\mathbf{x}_i = \mathbf{B}\mathbf{y}_i = \mathbf{B}(\mathbf{h}_i + \mathbf{n}_i)$, $i \in [\nu]$, be the resulting m-dimensional projections. To recover the sparse signal $\mathbf{h}_i$, we use the atomic-norm denoising algorithm [10]

$$\widehat{\mathbf{h}}_i = \arg\min \|\mathbf{h}\|_{\mathcal{A}} \text{ s.t. } \|\mathbf{x}_i - \mathbf{B}\mathbf{h}\|^2 \leq \epsilon, \tag{2}$$

where $\epsilon \approx m\sigma^2$ is an estimate of the noise power, and where $\|\mathbf{h}\|_{\mathcal{A}}$ denotes the atomic norm of $\mathbf{h}$ with respect to the continuous dictionary of the array vectors $\mathcal{A}$, defined by

$$\|\mathbf{h}\|_{\mathcal{A}} = \inf \Big\{ \sum_{\ell} c_\ell : c_\ell \geq 0, \text{and} \quad \exists\, (\theta_\ell, \phi_\ell) \text{ s.t. } \mathbf{h} = \sum_{\ell} c_\ell e^{j\phi_\ell} \mathbf{a}(\theta_\ell) \Big\}. \tag{3}$$

In general, finding a closed-form formula or even efficiently computing the atomic norm of a vector in a given dictionary is a challenging task, and different methods have been proposed for its approximation [10]. However, for the dictionary $\mathcal{A}$, it has been shown that the atomic norm can be efficiently computed via *semi-definite programming* (SDP) [11]. This results in the following SDP for estimating the sparse channel vector $\mathbf{h}_i$:

$$\widehat{\mathbf{h}}_i = \underset{\mathbf{h}\in\mathbb{C}^M, \mathbf{v}\in\mathbb{C}^M, \gamma\in\mathbb{R}_+}{\arg\min} \ \text{tr}[\mathbb{T}(\mathbf{v})] + \gamma \text{ s.t.} \quad \begin{bmatrix} \mathbb{T}(\mathbf{v}) & \mathbf{h} \\ \mathbf{h}^{\mathsf{H}} & \gamma \end{bmatrix} \succeq \mathbf{0}, \|\mathbf{x}_i - \mathbf{B}\mathbf{h}\|^2 \leq \epsilon, \tag{4}$$

where $\mathbb{T}(\mathbf{v})$ denotes an $M \times M$ Hermitian Toeplitz matrix whose first column is $\mathbf{v}$, and where $\epsilon = m\sigma^2$ is an estimate of the noise power.

In this paper, we will use optimization (4) as the *one-shot* sparse channel estimation algorithm since it uses only the observation $\mathbf{x}_i$ on the current slot i and does not exploit the previous training samples in a window of duration ν consisting of $\big\{\mathbf{x}_j : j \in \{i-\nu, i-\nu+1, \ldots, i-1\}\big\}$.

2.3 Time Variation of the Channel Vectors

For the sake of simplicity, we assume that the multipath component coefficients evolve according to first order Markov processes given by

$$w_\ell[t] = \alpha_\ell\, w_\ell[t-1] + \sigma_\ell \sqrt{1-\alpha_\ell^2}\; i_\ell[t], \tag{5}$$

where $i_\ell[t]$ is the innovation process for $w_\ell[t]$, which is a Gaussian process with a covariance $\mathbb{E}[i_\ell[t] i_{\ell'}[t']] = \delta_{\ell,\ell'}\delta_{t,t'}$, and where α_ℓ is the coefficient of first order auto-regression filter, which should be inside the unit circle to have a stable filter, i.e., $|\alpha_\ell| < 1$. To obtain a stationary process, we assume that $w_\ell[0] \sim \mathcal{CN}(0, \sigma_\ell^2)$ is initialized with the first realization of the channel gain for the ℓ-th scatterer. In this case, $w_\ell[t]$ generated by (5) is a stationary Gaussian process for all $t \geq 0$, whose auto-correlation function is given by

$$r_\ell[\Delta] = \mathbb{E}\big[w_\ell[t+\Delta] w_\ell[t]^*\big] = \sigma_\ell^2\, \alpha_\ell^{|\Delta|}. \tag{6}$$

For simplicity, we assume that $\alpha_\ell = \alpha$ is the same for all ℓ, and $\alpha \in [0, 1)$ is real-valued and positive. Since $\mathbf{h}_i$ is obtained by sampling $\mathbf{h}[t]$ every τ seconds, the matrix-valued auto-correlation function of $\mathbf{h}_i$ is given by

$$\mathbb{E}[\mathbf{h}_i \mathbf{h}_{i'}^{\mathsf{H}}] = \beta^{|i-i'|} \sum_{\ell=1}^{p} \sigma_\ell^2 \mathbf{a}(\theta_\ell)\mathbf{a}^{\mathsf{H}}(\theta_\ell) \tag{7}$$

where $\beta = \alpha^\tau$. Without loss of generality, we shall consider a measurement window $[\nu] = \{0, \ldots, \nu-1\}$ of ν slots, and look at the transmitter/receiver operations in slot $\mathcal{T}_\nu = [\nu\tau, (\nu+1)\tau - 1]$. Therefore, the measurement window is referred to as a block of "past observations", while the measurement at slot ν is the "current observation". We define the *coherence time* (or the settling time) of the channel by $\tau_c = \frac{1}{\log(1/\alpha)}$. We consider three idealized cases of interest:

1) When $T \ll \tau_c$, the channel process is almost constant over a time significantly larger than T. It follows that the channel on the current slot is approximately identical to the channel over the whole past observation window. In this case, predicting the channel on the current slot from the past window is expected to be very effective.
2) When $\tau \ll \tau_c \approx T$, the channel varies significantly over the past observation window, but remains approximately constant over each slot. Hence, one-shot estimation over the current slot yields an accurate estimate in high SNR. However, since channel estimation is performed *before* beamforming (in fact, it is used to calculate the beamformer) in mm-Wave communication it is reasonable to expect that estimation occurs in low SNR (without array beamforming gain). Hence, we are interested in using the past observation window to *improve* the one-shot estimation of the current channel.
3) When $\tau_c \approx \tau$, the channel process varies significantly over a slot (i.e., it is nearly i.i.d. over different slots). In this case, one-shot estimation is ineffective due to channel aging over the current slot, especially in the donwlink case. Nevertheless, we can learn the channel dominant subspace, i.e., the linear span of the atoms that best represent the channel over the past observation window, and still be able to separate the users in the signal space based only on subspace information. This is effective when such channel subspaces are low-dimensional, as is the case for mm-Wave channels [7].

2.4 Exploiting Past Measurements

In order to illustrate the fact that both sparsity in the AoA domain and time correlation can be used to improve channel

[2]Since $\mathbf{B}$ is the projection matrix corresponding to the RF beamforming receiver, it can be designed to satisfy row orthonormality.

estimation, we consider two extremes of cases 1) and 3) said above. In the first case, the channel is exactly constant over a an interval much larger than T, i.e., $\mathbf{h}[i\tau] = \mathbf{h}_i = \mathbf{h}_0$ for $i \in [\nu]$. Hence, by simply averaging the training observations $\mathbf{x}_i$ for $i \in [\nu]$, we obtain

$$\overline{\mathbf{x}} = \frac{1}{\nu} \sum_{i\in[\nu]} \mathbf{x}_i = \mathbf{B} \left(\mathbf{h}_0 + \frac{1}{\nu} \sum_{i\in[\nu]} \mathbf{n}_i \right). \quad (8)$$

Applying the one-shot sparse estimator (4) to (8), we obtain an estimate of $\mathbf{h}_\nu \approx \mathbf{h}_0$ with an improvement in the observation SNR by a factor of ν. Furthermore, because of the strong correlation in time, the system does not even need to exploit the observation on the current slot (this would only improve the SNR by a marginal factor of $(1 + 1/\nu)$). This means that, for highly time-correlated channel dynamics, channel prediction can be effectively exploited.

Now consider the opposite extreme case, where the channel gains are i.i.d. over the sequence of slots. Let us consider the sample covariance estimator $\widehat{\mathbf{C}}_x = \frac{1}{\nu} \sum_{i\in[\nu]} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$. By the consistency of the sample covariance, for sufficiently large ν, we have

$$\widehat{\mathbf{C}}_x \approx \mathbf{B}\mathbf{C}_h\mathbf{B}^{\mathsf{H}} + \sigma^2\mathbf{B}\mathbf{B}^{\mathsf{H}} = \mathbf{B}\mathbf{C}_h\mathbf{B}^{\mathsf{H}} + \sigma^2\mathbf{I}_m, \quad (9)$$

where we have assumed that the rows of $\mathbf{B}$ are orthonormal. In our previous work [12], we showed that it is possible to exploit the angular sparsity and the underlying Toeplitz structure of $\mathbf{C}_h$ (for the ULA), such that the p-dimensional signal subspace that contains $\mathbf{h}_i$ with probability 1, namely, $\text{Span}\{\mathbf{a}(\theta_\ell) : \ell = 1, \dots, p\}$, be efficiently estimated when the projection matrix $\mathbf{B}$ has only $m \approx 2\sqrt{M}$ rows. As a matter of fact, it is sufficient to let $\mathbf{B}$ have a single non-zero element equal to 1 in each row, such that $\mathbf{B}$ induces a subsampling of the array elements (antenna selection) in coprime locations. In particular, ν of the order $\sim 50 - 100$ samples seems to be sufficient to precisely estimate this subspace for moderate SNR values around $\mathsf{snr} \sim 0 - 10$ dB. Let $\mathbf{U}$ be the $M \times p$ tall unitary matrix whose columns are basis of the estimated signal subspace. We can improve the one-shot estimate the channel vector $\mathbf{h}_\nu = \mathbf{h}[\nu\tau]$ on the current slot by solving the following least-square problem

$$\widehat{\mathbf{w}}_\nu = \underset{\mathbf{w}\in\mathbb{C}^p}{\arg\min} \|\mathbf{x}_\nu - \mathbf{B}\mathbf{U}\mathbf{w}\|^2, \quad (10)$$

from which we can estimate the channel vector by $\widehat{\mathbf{h}}_\nu = \mathbf{U}\widehat{\mathbf{w}}_\nu$. If the power of the signal $\mathbf{h}_\nu$ is not uniformly distributed in different directions spanned by the columns of $\mathbf{U}$, this estimate can be further improved by weighted least-squares.

In this case, when the channel varies so fast that even the aging over a single slot yields too much degradation of the beamforming performance, the multiuser interference can still be managed by exploiting only the subspace information rather than the instantaneous estimate $\widehat{\mathbf{h}}_\nu$. For example, the interference from a user with channel vector $\mathbf{h}_\nu$ can be eliminated by projecting onto the orthogonal complement of its p-dim subspace. The drawback is that, compared with the projection on the orthogonal complement of $\widehat{\mathbf{h}}_\nu$, which wastes only 1 degree of freedom, one wastes p degrees of freedom for zero-forcing a specific user. However, this results in a negligible loss when $p \ll M$, especially when a whole group of users spanning roughly the same subspace can be zero-forced simultaneously [6, 7].

It is seen that, in both extreme cases of channel time dynamics, the window of past observations provides very useful information that can be exploited at the base-station receiver (uplink) or transmitter (downlink). In Section 3, we propose an algorithm that uses the training samples $\mathbf{h}_i$, $i \in [\nu]$, to find an estimate of the p-dim signal subspace $\mathbf{U}$, which would be exploited in the ν-th training period. When this information is used to enhance the channel estimation on the current slot, we evaluate the performance of our algorithm by looking at the correlation coefficient between the true and the estimated channel vector, i.e., $\eta(\mathbf{h}_\nu, \widehat{\mathbf{h}}_\nu) = \frac{|\langle \mathbf{h}_\nu, \widehat{\mathbf{h}}_\nu \rangle|}{\|\mathbf{h}_\nu\| \|\widehat{\mathbf{h}}_\nu\|}$. When we use the subspace information to reject interference, we shall look at the normalized residual signal power, given by $\mu(\mathbf{h}_\nu, \mathbf{U}) = \frac{\mathbf{h}_\nu^{\mathsf{H}}(\mathbf{I}_M - \mathbf{U}\mathbf{U}^{\mathsf{H}})\mathbf{h}_\nu}{\|\mathbf{h}_\nu\|^2}$.

3 Algorithm for Subspace Estimation

As a robust algorithm for subspace estimation, we use a variant of RMMV (reduced multiple-measurement vector) algorithm that we proposed in [12]. The main motivation for this algorithm comes from the *multiple measurement vectors (MMV)* problem in compressed sensing. We will briefly explain the MMV problem and why it gives a suitable formulation for subspace estimation in our case. We will also briefly explain the motivation for using RMMV algorithm for extracting the signal subspace.

Consider the channel vectors $\mathbf{h}_i$, $i \in [\nu]$, belonging to an observation window of size $T = \nu\tau$. As we explained in Section 2.1, we assume that the scattering geometry of the user remains invariant inside this window. This implies that, no matter how the channel dynamics (slowly or quickly varying), the channel vectors of the user inside the window have a sparse representation in the continuous dictionary $\mathcal{A}$ consisting of array responses for different AoA $\theta \in [-\theta_m, \theta_m]$. In particular, all the channel vectors $\mathbf{h}_i$, $i \in [\nu]$, have the same support in $\mathcal{A}$, which is given by the AoA $\{\theta_\ell\}_{\ell=1}^p$. This implies that not only every individual channel vector is sparse over $\mathcal{A}$, but also all the channel vectors together have a joint (group) sparsity structure. This problem has been vastly studied in the compressed sensing literature and it has been shown that exploiting the joint sparsity can further boost the performance, e.g., reduce the number of required measurements (see [13–15] and the references therein).

Different algorithms have been proposed in the literature for exploiting the joint sparsity such as greedy algorithms [13], convex optimization with a regularization to promote the joint sparsity [14], subspace methods [15], and more recent off-grid variants [16, 17]. In this paper, similar to the one-shot estimation problem (4), we will focus on atomic norm denoising for estimating the jointly sparse channel vectors from the collection of sketches $\mathbf{x}_i$, $i \in [\nu]$, where the joint sparsity is incorporated by considering the new dictionary

$$\mathcal{D} = \{\mathbf{a}(\theta)\mathbf{b}^{\mathsf{H}} : \theta \in [-\theta_{\max}, \theta_{\max}], \mathbf{b} \in \mathbb{C}^\nu\}. \quad (11)$$

This approach has been used in [16, 17], where it has been shown that, similar to the one-shot variant (4), the atomic norm denoising can be formulated as an SDP. However, the constraints of this SDP have dimension $(M+\nu) \times (M+\nu)$, which increases by increasing the number of samples. As a result, the computational complexity is quite high even for moderate values $M \approx 64$ and number of samples $\nu \approx 100$.

In [12], we proposed the RMMV algorithm, which has pretty the same performance as the SDP proposed in [16, 17] but its computational complexity does not increase with the sample size ν. This algorithm first computes the sample covariance matrix $\widehat{\mathbf{C}}_x$, its *singular value decomposition* (SVD) given by $\widehat{\mathbf{C}}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{H}}$, and the low-dimensional data given by $\widetilde{\mathbf{X}} = \mathbf{U}\mathbf{\Lambda}$. It is not difficult to check that $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{V}_m$, where $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\nu-1}]$ is the matrix of the whole sketches, with the SVD $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}$, and where $\mathbf{V}_m$ is the $\nu \times m$ matrix consisting of the first m columns of $\mathbf{V}$. Note that $\widetilde{\mathbf{X}}$ is an $m \times m$ matrix, whose dimension depends on the dimension of the sketches rather than the number of observations ν. It is not difficult to see that, similar to the columns of $\mathbf{X}$, the columns of $\widetilde{\mathbf{X}}$ still keep their MMV format, i.e., they have the same support over the projected dictionary given by $\mathbf{B}\mathcal{D} = \{\mathbf{B}\,\mathbf{a}(\theta)\mathbf{b}^{\mathsf{H}} : \theta \in [-\theta_{\max}, \theta_{\max}], \mathbf{b} \in \mathbb{C}^{\nu}\}$. The RMMV algorithm is obtained by applying the atomic norm denoising to the low-dimensional data $\widetilde{\mathbf{X}}$, and has the following SDP formulation [12]:

$$\mathbf{C}_y^* = \underset{\mathbf{T}\in\mathbb{T}_+, \mathbf{W}\in\mathbb{C}^{m\times m}}{\arg\min} \ \mathsf{Tr}(\mathbf{B}\mathbf{T}\mathbf{B}^{\mathsf{H}}) + \mathsf{Tr}(\mathbf{W})$$
$$\text{subject to} \begin{bmatrix} \mathbf{B}\mathbf{T}\mathbf{B}^{\mathsf{H}} & \widetilde{\mathbf{X}} \\ \widetilde{\mathbf{X}}^{\mathsf{H}} & \mathbf{W} \end{bmatrix} \succeq \mathbf{0}, \qquad (12)$$

where $\mathbb{T}_+$ denotes the space of all $M \times M$ Hermitian Toeplitz matrices, and where $\mathbf{C}_y^*$ is an estimate of the sample covariance of the whole data $\mathbf{y}$, whose dominant subspace gives an estimate of the signal subspace of $\mathbf{C}_h$ (the covariance matrix of the channel vectors).

4 Simulations

In this section, we assess the performance our proposed algorithm via numerical simulations. We use τ as in Section 2.1 for the period of training symbols, and τ_c for the coherence time of the channel. When $\tau \approx \tau_c$, the resulting channel vectors are approximately independent from each other, whereas when $\tau \ll \tau_c$, the channel vectors are fully correlated.

Channel Model. We consider a simple model for the channel consisting of $p = 3$ multipath components with the AoAs $\{0, +20, -20\}$ degrees, and with equal power in each component.

Array Model and Sampling Scheme. For simulation, we use an array with $M = 64$ antennas. We take $m = 16$ orthogonal sketches of the array input signal, thus, the sampling ratio is $\rho = \frac{m}{M} = 0.25$. We use an $m \times M$ random binary sampling matrix $\mathbf{B}$, which selects m array elements randomly (random antenna selection). In particular, each row of $\mathbf{B}$ has only one 1 is a random antenna location, and has 0 elsewhere.

Window Size. We use a window of size $\nu = 50$, where the signal subspace or the channel vector $\mathbf{h}_\nu$ at the last instant ν is estimated from all the channel vectors $\mathbf{h}_i$, $i \in [\nu]$.

Performance Metric. We consider two performance metrics as explained in Section 2.4. When the goal is to use the past observations to enhance the channel estimation on the current slot, we use the correlation coefficient between the true and the estimated channel vector, i.e., $\eta(\mathbf{h}_\nu, \widehat{\mathbf{h}}_\nu) = \frac{|\langle \mathbf{h}_\nu, \widehat{\mathbf{h}}_\nu \rangle|}{\|\mathbf{h}_\nu\| \|\widehat{\mathbf{h}}_\nu\|}$, and plot the CCDF (complementary cumulative distribution function) of $20\log_{10}[1/\eta(\mathbf{h}_\nu, \widehat{\mathbf{h}}_\nu)]$. Fig. 3 shows the simulation results for this case.

When we use the subspace information to reject interference, we consider normalized residual signal power, given by $\mu(\mathbf{h}_\nu, \mathbf{U}) = \frac{\mathbf{h}_\nu^{\mathsf{H}}(\mathbf{I}_M - \mathbf{U}\mathbf{U}^{\mathsf{H}})\mathbf{h}_\nu}{\|\mathbf{h}_\nu\|^2}$, where $\mathbf{U}$ is the estimated signal subspace, and plot the CCDF of $10\log_{10}[1/\mu(\mathbf{h}_\nu, \mathbf{U})]$ as a performance measure. Fig. 4 shows the resulting performance.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *Access, IEEE*, vol. 1, pp. 335–349, 2013.

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

[4] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. ACM, 2012, pp. 53–64.

[5] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive mimo for next generation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, 2014.

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

[7] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. on Sel. Areas on Commun. (JSAC)*, vol. 32, no. 6, pp. 1239–1255, 2014.

[8] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

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

[10] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[11] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5987–5999, 2013.

[12] S. Haghighatshoar and G. Caire, "Channel vector subspace estimation from low-dimensional projections," *arXiv preprint arXiv:1509.07469*, 2015.

[13] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[14] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[15] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3613–3641, 2012.

[16] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *arXiv preprint arXiv:1408.2242*, 2014.

[17] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *arXiv preprint arXiv:1405.6585*, 2014.

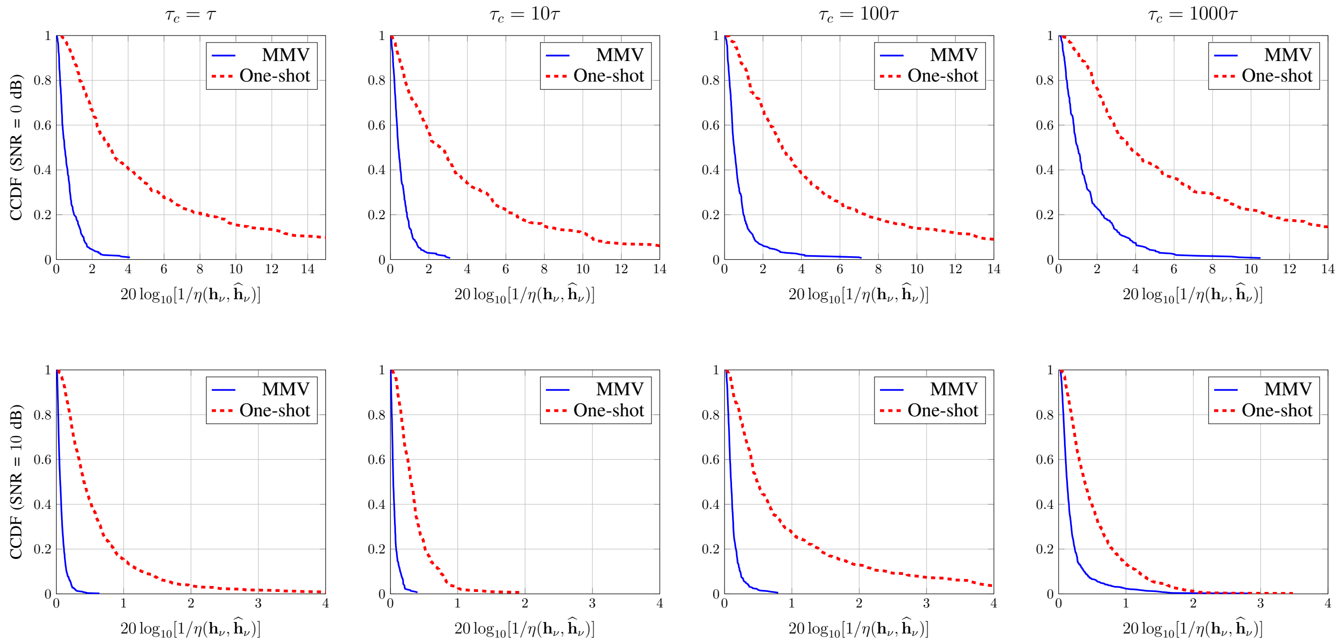

Fig. 3: Comparing the performance of MMV method with the traditional One-shot channel estimation for different SNR and different channel coherence time τ_c. Window size $\nu = 50$, number of array elements $M = 64$, dimension of the sketches $m = 16$, and sampling scheme is *random antenna selection* (the sampling matrix $\mathbf{B}$ is a binary matrix with only one 1 in each row).

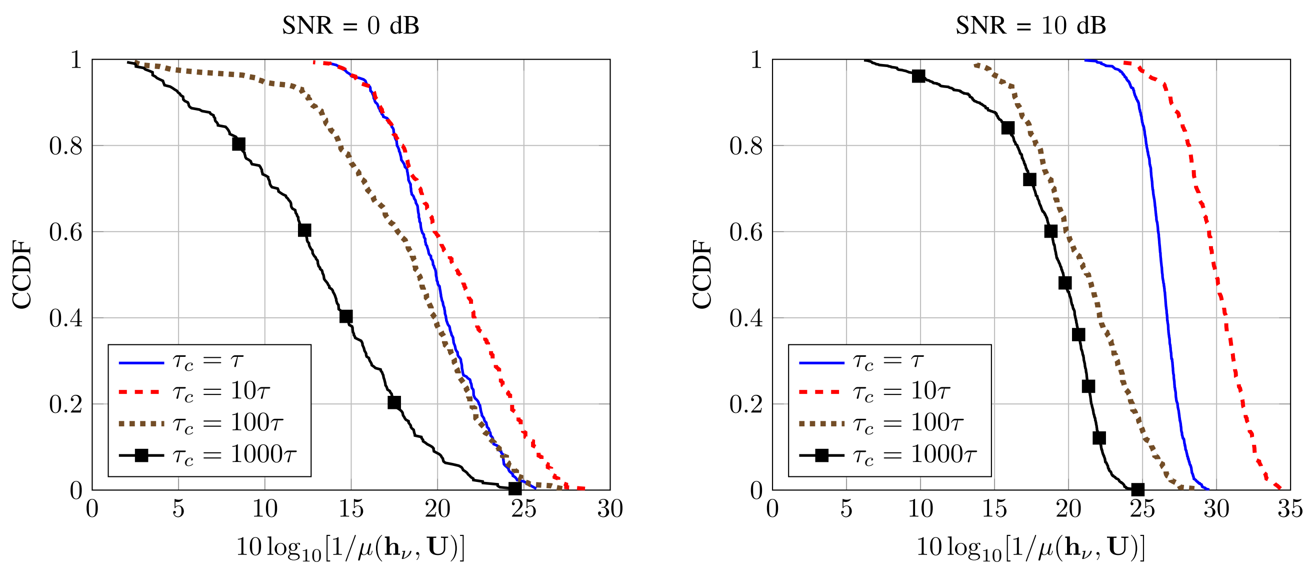

Fig. 4: The fraction of the power of the $\mathbf{h}_\nu$ rejected by projecting onto the estimated subspace for different values of SNR and for different values of channel coherence time τ_c.