Dealing with the Mobility Problem of Massive MIMO using Extended Prony's Method

Haifan Yin[†], Haiquan Wang[⋆], Yingzhuang Liu[†], David Gesbert[§]

[†]School of Electronic Information and Communications, Huazhong University of Science and Technology, 430074 Wuhan, China
[⋆]School of Communications Engineering, Hangzhou Dianzi University, 310018 Hangzhou, China
[§]Communications Systems Department, EURECOM, 06904 Biot, France
Email: yin@hust.edu.cn, tx_wang@hdu.edu.cn, liuyz@mail.hust.edu.cn, gesbert@eurecom.fr

Abstract—**Massive MIMO is a key technology for 5th generation (5G) mobile communications. The large excess of base station (BS) antennas brings unprecedented spectral efficiency. However, during the initial phase of industrial testing, a practical challenge arises which undermines the actual deployment of massive MIMO and is related to mobility. In fact, testing teams reported that in moderate-mobility scenarios, e.g., 30 km/h of UE speed, the performance may drop 50% compared to the low-mobility scenario, a problem not foreseen by theoretical papers on the subject. In order to deal with this challenge, we propose a Prony-based angular-delay domain (PAD) prediction method, which is built on exploiting the angle-delay-Doppler structure of the multipath. Our theoretical analysis shows that when the number of base station antennas and the bandwidth are large, the prediction error of our PAD algorithm converges to zero for any UE velocity level, provided that only two accurate enough previous channel samples are available. Simulation results show that under the realistic channel model of 3GPP in rich scattering environment, our proposed method even approaches the performance of stationary scenarios where the channels do not vary at all.**

Index Terms—**massive MIMO, mobility, 5G, channel aging, channel prediction, Prony's method**

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) introduced in [1], is one of the key enablers of the 5G cellular systems. Compared to traditional MIMO with fewer BS antennas, massive MIMO can offer superior spectral efficiency and energy efficiency [2] at least in theory. The growing orthogonality between channel vectors of different UEs allows the base station to reject interference by inexpensive precoding methods, providing that Channel State Information (CSI) is known at base station. CSI acquisition is known to be a formidable problem in massive MIMO. An example of CSI acquisition issue is the pilot contamination problem. A rich body of literature has addressed this problem. The solutions vary from angular/amplitude domain discrimination [3] [4] [5], pilot coordination [3], multi-cell minimum mean square error (M-MMSE) [6] [7], etc.

Despite the technology hype and great expectations behind massive MIMO, some of the latest field trials have been more than disappointing when it comes to actual system performance (see [8] for example). In particular it appeared that CSI acquisition can be severely affected in mobility scenarios. This is related to the time-varying nature of wireless channel which itself limits its coherence time. In practical cellular networks, the processing delay at the base station is inevitable because of the highly sophisticated 5G protocol, scheduling, resource allocation, encoding/decoding, and channel training under UE power constraint. This implies that even in moderate-mobility scenarios, the processing delay can end up being larger than the coherence time, making it essentially unusable for multiuser beamforming. Channel aging effect is particularly harmful to massive MIMO, which relies on high precision CSI to achieve large multiplexing gains. It was for instance observed in industrial settings, that with a typical CSI delay of 4 milliseconds, the moderate-mobility scenario at 30 km/h leads to as much as 50% of the performance reduction versus in low-mobility scenario at 3 km/h. Some information theoretic efforts exploiting severely delayed CSI have been demonstrated to be possible but never tested out in practical 5G contexts [9], [10]. The effects of channel aging under a simple autoregressive model of the channel time variations were studied in [11]. A linear finite impulse reponse (FIR) Wiener predictor was proposed which could partially overcome the channel aging effect at the expense of high computational complexity.

In this paper, we revisit the problem of CSI acquisition by combining it with channel prediction algorithms. We propose a novel Prony-based angular-delay domain channel prediction algorithm by exploiting the structural information of the multipath channel. The large number of BS antennas and the large bandwidth in 5G lead to higher resolution in both angular domain and delay domain. Our idea consists in exploiting this high resolution regime specifically. In practice the approach involves projecting the channel into an angular-delay grid, then capturing the channel variations in the angular-delay domain. Moreover, we propose to adopt here Prony's method, traditionally used in the context of spectral analysis, for its ability to predict a uniformly sampled signal composed of damped complex exponentials. This feature turns out to be useful in the massive MIMO context because the training signal in 5G are normally periodic and the channel can be regarded as a sum of complex exponentials with each one corresponding to a path response having a Doppler component. We analyze the asymptotic performance of our PAD algorithm

and prove that as the number of base station antennas and the bandwidth increase, the channel prediction error converges to zero, provided that only two accurate enough channel samples are available. Finally, since in practice, current channel estimates are noisy, we improve the performance of the PAD method by combining it with a denoising method using an adaptation of Tufts-Kumaresan's method [12].

Simulations under the clustered delay line (CDL) channel model of 3GPP [13] shows that our proposed method at 60 km/h of UE speed is very close to the ideal case of a stationary setting. To the best of our knowledge, the study of channel prediction under such a realistic model of wideband massive MIMO has never been done before, and the high spatial-frequency resolution of 5G has never been fully exploited to solve the mobility challenge.

Notations: We use boldface to denote matrices and vectors. Specifically, $\mathbf{I}$ denotes the identity matrix. $(\mathbf{X})^T$, $(\mathbf{X})^*$, and $(\mathbf{X})^H$ denote the transpose, conjugate, and conjugate transpose of a matrix $\mathbf{X}$ respectively. $(\mathbf{X})^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{X}$. $\text{tr}\{\cdot\}$ denotes the trace of a square matrix. $\|\cdot\|_F$ stands for the Frobenius norm. $\mathbb{E}\{\cdot\}$ denotes the expectation. $\mathbf{X} \otimes \mathbf{Y}$ is the Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$. $\text{vec}(\mathbf{X})$ is the vectorization of the matrix $\mathbf{X}$. $\text{diag}\{\mathbf{a_1}, ..., \mathbf{a_N}\}$ denotes a diagonal matrix or a block diagonal matrix with $\mathbf{a_1}, ..., \mathbf{a_N}$ at the main diagonal. $\triangleq$ is used for definition.

II. CHANNEL MODELS

For ease of exposition, we consider an arbitrary UE in a certain cell. We assume the antennas at the base station form a uniform planar array (UPA) with N_v rows and N_h columns as in commercial systems[1]. Denote the number of antennas at the base station as N_t and the number of antennas at the UE as N_r. It is clear that $N_t = N_v N_h$. The network operates in TDD mode and the UL and DL occupy the same bandwidth, which consists of N_f subcarriers with spacing $\triangle f$. The channel is composed of P multipath, with each path having a certain angle, delay, Doppler, and complex amplitude. We denote the elevation departure angle, azimuth departure angle, elevation arrival angle, and azimuth arrival angle of the p-th path as $\theta_{p,\text{ZOD}}$, $\phi_{p,\text{AOD}}$, $\theta_{p,\text{ZOA}}$, and $\phi_{p,\text{AOA}}$ respectively. The DL channel at a certain time t and a subcarrier with frequency f is denoted as $\mathbf{H}(f,t) \in \mathbb{C}^{N_r \times N_t}$. According to [13], the channel between the s-th base station antenna and the u-th UE antenna is modeled as

$$h_{u,s}(f,t) = \sum_{p=1}^{P} \beta_p e^{\frac{j2\pi \hat{r}_{\text{rx},p}^T \bar{d}_{\text{rx},u}}{\lambda_0}} e^{\frac{j2\pi \hat{r}_{\text{tx},p}^T \bar{d}_{\text{tx},s}}{\lambda_0}} e^{-j2\pi f \tau_p} e^{j\omega_p t}, \quad (1)$$

where β_p and τ_p are the complex amplitude and the delay of the p-th path respectively. λ_0 is the wavelength of center frequency. $\hat{r}_{\text{rx},p}$ is the spherical unit vector with azimuth arrival angle $\phi_{p,\text{AOA}}$ and elevation arrival angle $\theta_{p,\text{ZOA}}$:

$$\hat{r}_{\text{rx},p} \triangleq \begin{bmatrix} \sin\theta_{p,\text{ZOA}} \cos\phi_{p,\text{AOA}} \\ \sin\theta_{p,\text{ZOA}} \sin\phi_{p,\text{AOA}} \\ \cos\theta_{p,\text{ZOA}} \end{bmatrix}. \quad (2)$$

Likewise, $\hat{r}_{\text{tx},p}$ is the spherical unit vector defined as:

$$\hat{r}_{\text{tx},p} \triangleq \begin{bmatrix} \sin\theta_{p,\text{ZOD}} \cos\phi_{p,\text{AOD}} \\ \sin\theta_{p,\text{ZOD}} \sin\phi_{p,\text{AOD}} \\ \cos\theta_{p,\text{ZOD}} \end{bmatrix}. \quad (3)$$

$\bar{d}_{\text{rx},u}$ is the u-th UE antenna's location vector which contains the 3D cartesian coordinate. Similarly, $\bar{d}_{\text{tx},s}$ is the location vector of the s-th base station antenna. The last exponential term $e^{j\omega_p t}$ is the Doppler of the p-th path, where t denotes time. ω_p is defined as $\omega_p \triangleq \hat{r}_{\text{rx},p}^T \bar{v}/\lambda_0$, where $\bar{v}$ is the UE velocity vector with speed v, travel azimuth angle ϕ_v, and travel elevation angle θ_v:

$$\bar{v} \triangleq v[\ \sin\theta_v \cos\phi_v \quad \sin\theta_v \sin\phi_v \quad \cos\theta_v\ ]^T. \quad (4)$$

An illustration of the coordinate system is shown in Fig. 1.
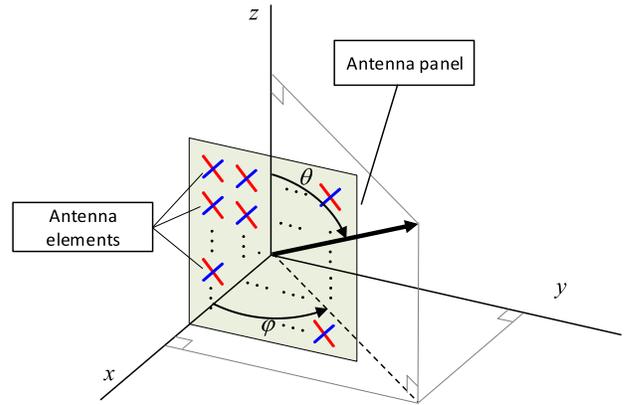
Fig. 1. Definition of the coordinate system.

Let $\mathbf{h}_u(f,t) \in \mathbb{C}^{1 \times N_t}$ denote the channel between all base station antenna and the u-th UE antenna at time t and frequency f. We write the channels at all N_f subcarriers in a matrix form:

$$\mathbf{H}_u(t) \triangleq [\ \mathbf{h}_u^T(f_1,t) \quad \mathbf{h}_u^T(f_2,t) \quad \cdots \quad \mathbf{h}_u^T(f_{N_f},t)\ ], \quad (5)$$

where f_i is the frequency of the i-th $(1 \leq i \leq N_f)$ subcarrier.

III. DEALING WITH MOBILITY PROBLEM

A. The mobility problem of massive MIMO

As is well known, channel time variability can create inter-user interference induced by a precoder which is computed based on aging CSI. This impediment can be mitigated by anticipating the future channel variations. While predicting the future fading state of a wireless channel is a very challenging problem, the accounting of the specific space-time structure of the channel which arises in a broadband context (as in 5G) opens fresh perspectives for improvement.

B. A review of Prony's method

Prony's method proposed by Gaspard Riche de Prony in 1795 is a useful tool to analyze a uniformly sampled signal composed of a number of damped complex exponentials [14] and extract valuable information (e.g., the amplitudes and frequencies of the exponentials) which can be used for prediction.

A review of this method is given below. Suppose we have K samples of data $y(k)$ which consist of N exponentially damped signals:

$$y(k) = \sum_{n=1}^{N} \beta_n e^{(-\alpha_n + j2\pi f_n)k}, 0 \le k \le K-1, \quad (6)$$

where α_n (positive) and f_n $(1 \le n \le N)$ are the pole damping factor and pole frequency respectively. β_n $(1 \le n \le N)$ is the complex amplitude. Note that in the special case of channel prediction, $y(k)$ can be regarded as the uniformly sampled channel estimate. Define the following polynomial:

$$P_0(z) \triangleq \prod_{n=1}^{N}(z - e^{s_n}) = \sum_{n=0}^{N} p_n z^n, z \in \mathbb{C}, \quad (7)$$

where $s_n = -\alpha_n + j2\pi f_n$ for $n = 1, \cdots, N$. It is clear that $p_N = 1$ and $e^{s_n}, (n = 1, \cdots, N)$ are zeros of $P_0(z)$. For an arbitrary $m \in \mathbb{N}$, one has

$$\sum_{n=0}^{N} p_n y(n+m) = \sum_{n=0}^{N} p_n \sum_{l=1}^{N} \beta_l e^{s_l(n+m)}$$
$$= \sum_{l=1}^{N} \beta_l e^{s_l m}\left(\sum_{n=0}^{N} p_n e^{s_l n}\right) \overset{a}{=} 0, \quad (8)$$

where $\overset{a}{=}$ is due to the fact that $e^{s_l} (l = 1, \cdots, N)$ are zeros of $P_0(z)$. Eq.(8) implies that the following homogeneous linear difference equation is fulfilled:

$$\sum_{n=0}^{N-1} p_n y(n+m) = -y(N+m), m \in \mathbb{N}. \quad (9)$$

Thus, we may obtain the coefficients p_n with the $2N$ sampled data by solving

$$\begin{bmatrix} y(0) & y(1) & \cdots & y(N) \\ y(1) & y(2) & \cdots & y(N+1) \\ \vdots & \vdots & \vdots & \vdots \\ y(N-1) & y(N) & \cdots & y(2N-1) \end{bmatrix}\begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_N \end{bmatrix} = \mathbf{0}. \quad (10)$$

Or equivalently, the following linear equations:

$$\mathbf{Y}\mathbf{p} = -\mathbf{h}, \quad (11)$$

where $\mathbf{Y}$ is a square Hankel matrix

$$\mathbf{Y} \triangleq \begin{bmatrix} y(0) & y(1) & \cdots & y(N-1) \\ y(1) & y(2) & \cdots & y(N) \\ \vdots & \vdots & \vdots & \vdots \\ y(N-1) & y(N) & \cdots & y(2N-2) \end{bmatrix}, \quad (12)$$

$$\mathbf{p} \triangleq [\ p_0 \quad p_1 \quad \cdots \quad p_{N-1}\ ]^T, \quad (13)$$

$$\mathbf{h} \triangleq [\ y(N) \quad y(N+1) \quad \cdots \quad y(2N-1)\ ]^T. \quad (14)$$

The least squares solution to Eq. (11) is given by

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{Y}\mathbf{p} + \mathbf{h}\|^2 = -\mathbf{Y}^\dagger \mathbf{h}. \quad (15)$$

Note that we may need $K \ge 2N$ samples to obtain all the coefficients $p_n, n = 0, \cdots, N-1$.

C. Prony-based angular-delay domain channel prediction

As shown in Eq. (1), the channel is composed of P paths, each path has a Doppler term $e^{j\omega_p t}, p = 1, \cdots, P$. In practice, the number of paths can be large, e.g., several hundreds in rich scattering environment, which makes the prediction difficult. In order to cope with this problem, we propose a Prony-based angular-delay domain (PAD) channel prediction algorithm. The main idea is to convert the channel into another domain where the Doppler terms of different paths are less intertwined with each other. We choose this domain in such a way that it reflects the geometry of the antenna array and the wideband delay response structure of the channel. Denote a discrete Fourier transform (DFT) matrix of size $N \times N$ by

$$\mathbf{W}(N) \triangleq \frac{1}{\sqrt{N}}\begin{bmatrix} \omega^{0\cdot 0} & \omega^{0\cdot 1} & \cdots & \omega^{0(N-1)} \\ \omega^{1\cdot 0} & \omega^{1\cdot 1} & \cdots & \omega^{1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(N-1)\cdot 0} & \omega^{(N-1)\cdot 1} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix},$$

where $\omega \triangleq e^{-2\pi j/N}$. Since UPA antenna array is considered, we define a spatial orthogonal basis $\mathbf{S}$ as

$$\mathbf{S} \triangleq \mathbf{W}(N_h) \otimes \mathbf{W}(N_v), \quad (16)$$

where N_h and N_v are the number of columns and the number of rows of antennas on the UPA respectively. Note that we assume the antenna numbering in $\mathbf{h}_u(f, t)$ is first columns then rows. The frequency orthogonal basis is defined as $\mathbf{F} \triangleq \mathbf{W}(N_f)$. We convert the channel Eq. (5) into angular-delay domain with the orthogonal basis of $\mathbf{S}$ and $\mathbf{F}$.

$$\mathbf{G}_u(t) \triangleq \mathbf{S}^H \mathbf{H}_u(t)\mathbf{F}. \quad (17)$$

$\mathbf{G}_u(t)$ is in fact the orthogonal projection of the channel on the angular-delay grid of size $N_t \times N_f$. Due to limited angular spread as well as the limited delay spread of the wireless propagating environment [3] [15], the channel in angular-delay domain is sparse. In other words, most of the elements in matrix $\mathbf{G}_u(t)$ are very close to zero. As a result we may ignore the insignificant elements in $\mathbf{G}_u(t)$ and focus on the predictions of the significant ones. Let $\mathbf{g}_u(t_l) = \text{vec}(\mathbf{G}_u(t_l))$ and $\tilde{\mathbf{g}}_u(t_l)$ be the re-arranged $\mathbf{g}_u(t_l)$ with its absolute values in non-increasing order. The number of non-negligible angular-frequency positions N_s is defined as

$$N_s \triangleq \arg\min_{N_s}\left\{\sum_{u=1}^{N_r}\sum_{l=0}^{L}\sum_{n=1}^{N_s}|\tilde{\mathbf{g}}_u(t_l, n)|^2 \ge \gamma \sum_{u=1}^{N_r}\sum_{l=0}^{L}\|\mathbf{G}_u(t_l)\|_F^2\right\},$$

where $\tilde{\mathbf{g}}_u(t_l, n)$ is the n-th entry of $\tilde{\mathbf{g}}_u(t_l)$, γ is a positive value that is close to 1. The physical meaning of γ is the ratio between the sum power of significant elements and the total power of the channel. An empirical value of γ can be 0.99. Note that N_s is normally much smaller than the size of the matrix $\mathbf{G}_u(t)$: $N_s \ll N_t N_f$. Thus by ignoring the insignificant elements, we may greatly reduce the computational complexity in channel prediction. We use $g_{u,n}(t), (n = 1, \cdots, N_s)$ to denote the n-th significant entry, which is located at the $r(n)$-th row and the $c(n)$-th column of $\mathbf{G}_u(t)$. The channel can be approximated as

$$\bar{\mathbf{H}}_u(t) = \sum_{n=1}^{N_s} g_{u,n}(t)\mathbf{s}_{r(n)}\mathbf{f}_{c(n)}^H, \quad (18)$$

where $\mathbf{s}_i$ is the i-th column of $\mathbf{S}$ and $\mathbf{f}_i$ is the i-th column of $\mathbf{F}$. Assuming the CSI delay $T_d = N_d \triangle T$, $N_d \in \mathbb{N}$, we seek to predict the channel $\mathbf{H}_u(t_{L+N_d})$ using the known samples $\mathbf{H}_u(t_0), \mathbf{H}_u(t_1), \cdots, \mathbf{H}_u(t_L)$. Without loss of generality, we assume L is odd and let the order of the predictor $N = (L+1)/2$. For a certain $n, 1 \le n \le N_s$, we may obtain the Prony coefficients by solving

$$\mathcal{G}(u,n)\mathbf{p}(u,n) = -\mathbf{g}(u,n), \qquad (19)$$

where $\mathcal{G}(u,n) \triangleq$

$$\begin{bmatrix} g_{u,n}(t_0) & g_{u,n}(t_1) & \cdots & g_{u,n}(t_{N-1}) \\ g_{u,n}(t_1) & g_{u,n}(t_2) & \cdots & g_{u,n}(t_N) \\ \vdots & \vdots & \vdots & \vdots \\ g_{u,n}(t_{N-1}) & g_{u,n}(t_N) & \cdots & g_{u,n}(t_{2N-2}) \end{bmatrix} \qquad (20)$$

$$\mathbf{p}(u,n) \triangleq [\ p_0(u,n) \quad \cdots \quad p_{N-1}(u,n)\ ]^T, \qquad (21)$$

$$\mathbf{g}(u,n) \triangleq [\ g_{u,n}(t_N) \quad g_{u,n}(t_{N+1}) \quad \cdots \quad g_{u,n}(t_{2N-1})\ ]^T.$$

The least squares estimate of $\mathbf{p}(u,n)$ is

$$\hat{\mathbf{p}}(u,n) = -\mathcal{G}^\dagger(u,n)\mathbf{g}(u,n). \qquad (22)$$

The prediction of $g_{(u,n)}(t_{L+1})$ is given by

$$\hat{g}_{u,n}(t_{L+1}) = -\mathbf{g}(u,n,L)\hat{\mathbf{p}}(u,n), \qquad (23)$$

where $\mathbf{g}(u,n,L) \triangleq [\ g_{u,n}(t_{L-N+1}) \quad \cdots \quad g_{u,n}(t_L)\ ]$. When $N_d > 1$, we may repeat computing Eq. (23) $N_d - 1$ times and update $\mathbf{g}(u,n,L)$ each time by removing the first column and appending the previous predict to the last column, until we obtain the prediction $\hat{g}_{u,n}(t_{L+N_d})$.

The method is summarized in Algorithm 1.

Algorithm 1 PAD channel prediction method
___
1: Compute the angular-delay domain channel $\mathbf{G}_u(t_l)$ for $l = 0, \cdots, L$ and $u = 1, \cdots, N_r$ according to Eq. (17).
2: Find the significant values $g_{u,n}(t_l)$ and their positions $r(n)$ and $c(n)$ for $u = 1, \cdots, N_r, n = 1, \cdots, N_s, l = 0, \cdots, L$.
3: **for** $u = 1, \cdots, N_r$
4:    **for** $n = 1, \cdots, N_s$
5:    Compute the least squares estimate of the Prony coefficients as in Eq. (22);
6:    Repeat Eq. (23) N_d times to compute the prediction $\hat{g}_{(u,n)}(t_{L+N_d})$;
7:    **end for**
8:    Reconstruct the channel prediction at t_{L+N_d} as in Eq. (18) with $g_{u,n}(t)$ replaced by $\hat{g}_{(u,n)}(t_{L+N_d})$;
9: **end for**
___

D. Performance analysis of the PAD algorithm

The asymptotical performance of our PAD algorithm is now analyzed. In order to facility the analysis, we make a mild technical assumption that for any two paths $p \ne q$, at least one of the three properties are non-identical: their elevation departure angles, their azimuth departure angles, and their delays. We denote the noisy observation of $\mathbf{H}_u(t)$ as $\hat{\mathbf{H}}_u(t)$ and the predicted channel at t_{L+N_d} as $\widehat{\mathbf{H}}_u(t_{L+N_d})$. Our main theoretical result is shown in Theorem 1.

Theorem 1 *For an arbitrary delay $N_d \in \mathbb{N}^+$ and any UE velocity level, the asymptotic performance of the PAD algorithm yields:*

$$\lim_{N_v, N_h, N_f \to \infty} \frac{\left\| \widehat{\mathbf{H}}_u(t_{L+N_d}) - \mathbf{H}_u(t_{L+N_d}) \right\|_F^2}{\left\| \mathbf{H}_u(t_{L+N_d}) \right\|_F^2} = 0, \qquad (24)$$

under the condition that two channel samples are accurate enough, i.e., $\forall k = L - 1, L$,

$$\lim_{N_v, N_h, N_f \to \infty} \frac{\left\| \tilde{\mathbf{H}}_u(t_k) - \mathbf{H}_u(t_k) \right\|_F^2}{\left\| \mathbf{H}_u(t_k) \right\|_F^2} = 0. \qquad (25)$$

Proof: See the full-length version [16]. $\square$
Note that condition Eq. (25) may require some non-linear signal processing techniques. See [5] as an example of how this condition can be fulfilled for a multi-cell massive MIMO scenario in the presence of pilot contamination.

E. Dealing with channel estimation error

The channel estimate at the base station is always corrupted by noise, which undermines the performance of Prony's method. Thus we propose to deal with noise with a supplementary method, which relies on the subspace structure of the channel sample matrix and the second-order long-term statistics of the noisy channel samples. It consists in the following two ingredients

1) Tufts-Kumaresan's method: The main idea of the Tufts-Kumaresan's method [12] is to apply singular value decomposition (SVD) to the sample matrix, i.e., Eq. (12) or Eq. (20), of the linear prediction equations, and then remove the contributions of small singular values. Taking the estimate of $\mathbf{p}(u,n)$ for example, the SVD of $\mathcal{G}(u,n)$ can be written as:

$$\mathcal{G}(u,n) = \mathbf{U}(u,n)\mathbf{\Sigma}(u,n)\mathbf{V}^H(u,n) \qquad (26)$$

$$\approx \mathbf{U}_s(u,n)\mathbf{\Sigma}_s(u,n)\mathbf{V}_s^H(u,n), \qquad (27)$$

where $\mathbf{\Sigma}_s(u,n)$ only contains the significant singular values of $\mathcal{G}(u,n)$. The Tufts-Kumaresan's estimate of the Prony coefficients are given by

$$\hat{\mathbf{p}}_{\text{tk}}(u,n) = -\mathbf{V}_s(u,n)\mathbf{\Sigma}_s^{-1}(u,n)\mathbf{U}_s^H(u,n)\mathbf{g}(u,n). \qquad (28)$$

Note that $\mathbf{\Sigma}_s(u,n)$ can be obtained in a way that the minimum number of singular values satisfies

$$\text{tr}\{\mathbf{\Sigma}_s(u,n)\} \ge \gamma_{\text{tk}} \text{tr}\{\mathbf{\Sigma}(u,n)\}, \qquad (29)$$

where the threshold γ_{tk} is no greater than 1, i.e., $\gamma_{\text{tk}} = 0.99$.

2) Channel denoising with statistical information: The noisy channel samples between all base station antennas and the u-th UE antenna at time t and frequency f can be modeled as $\tilde{\mathbf{h}}_u(f,t) \in \mathbb{C}^{1 \times N_t}$:

$$\tilde{\mathbf{h}}_u(f,t) = \mathbf{h}_u(f,t) + \mathbf{n}_u(f,t), \tag{30}$$

where $\mathbf{h}_u(f,t)$ is the accurate channel and $\mathbf{n}_u(f,t)$ is the independent and identically distributed (i.i.d.) complex Gaussian noise with zero-mean and covariance σ_n^2. It is easy to obtain the covariance matrix of the noisy channel at the base station:

$$\tilde{\mathbf{R}} = \mathbb{E}\left\{\tilde{\mathbf{H}}^H(f,t)\tilde{\mathbf{H}}(f,t)\right\}, \tag{31}$$

where the expectation is taken over time, frequency, or both. $\tilde{\mathbf{H}}(f,t)$ is defined as

$$\tilde{\mathbf{H}}(f,t) \triangleq [\ \tilde{\mathbf{h}}_1^T(f,t) \quad \tilde{\mathbf{h}}_2^T(f,t) \quad \cdots \quad \tilde{\mathbf{h}}_{N_r}^T(f,t)\ ]^T. \tag{32}$$

From Eq. (30) we have

$$\tilde{\mathbf{R}} = \mathbf{R} + N_r \sigma_n^2 \mathbf{I}, \tag{33}$$

where

$$\mathbf{R} = \mathbb{E}\left\{\mathbf{H}^H(f,t)\mathbf{H}(f,t)\right\}, \tag{34}$$

with $\mathbf{H}(f,t)$ being the accurate counterpart of $\tilde{\mathbf{H}}(f,t)$. Due to the large number of BS antennas and the limited scattering environment, the channel covariance matrix $\mathbf{R}$ has a low-rankness property [3] [15], which means a fraction of the eigenvalues of $\mathbf{R}$ are very close to zero. Thus we may exploit this property to have an estimate of the power of noise. The eigen-decomposition of $\tilde{\mathbf{R}}$ is written as $\tilde{\mathbf{R}} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{U}}^H$ where $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}\{\sigma_1,...,\sigma_{N_t}\}$. The estimate of the noise power σ_n^2 can be obtained by averaging the smallest eigenvalues of $\tilde{\mathbf{R}}$. A linear filter $\mathcal{W}$ can be derived for channel denoising purpose:

$$\mathcal{W} = \arg\min_{\mathcal{W}} \mathbb{E}\left\{\|\tilde{\mathbf{H}}(f,t)\mathcal{W} - \mathbf{H}(f,t)\|_F^2\right\}. \tag{35}$$

The solution is given by Proposition 1.

Proposition 1 *The linear solution to the optimization problem of Eq. (35) yields*

$$\mathcal{W} = \tilde{\mathbf{U}}\mathbf{D}\tilde{\mathbf{U}}^H, \tag{36}$$

where $\mathbf{D}$ is a diagonal matrix with its i-th ($i = 1, \cdots, N_t$) diagonal entry being $\frac{\sigma_i - N_r\hat{\sigma}_n^2}{\sigma_i}$.

Proof: The derivation is based on the linear minimum mean square error (LMMSE) criterion and is straightforward. $\square$

Note that the covariance matrix is computed based on samples of all N_r UE antennas, since the scattering environments experienced by all co-located N_r antennas is very similar. The denoising filter $\mathcal{W}$ can also be built for each UE antenna, however with higher complexity.

IV. NUMERICAL RESULTS

This section contains simulation results of our proposed channel prediction schemes. The simulation parameters are listed in Table I. Since we adopt the CDL-A channel model, the number of multipath is 460, i.e., for each UE, there are 23 clusters of multipath with each cluster containing 20 rays. Consider a typical setting of 5G at 3.5 GHz with 30 kHz of subcarrier spacing and normal cyclic prefix. Each slot, with a duration of 0.5 ms, contains 14 OFDM symbols [17]. We assume that UE sends one SRS signal in each slot, which means one channel sample is available every 0.5 ms. The

TABLE I
BASIC SIMULATION PARAMETERS

Carrier frequency	3.5 GHz
Subcarrier spacing	30 kHz
Bandwidth	20 MHz (51 RBs)
Number of UEs	8
BS antenna configuration	$(\underline{M}, \underline{N}, \underline{P}, \underline{M}_g, \underline{N}_g) = (2, 8, 2, 1, 1)/(4, 8, 2, 1, 1)$, $(dH, dV) = (0.5, 0.8)\lambda$, the polarization angles are $\pm 45°$
UE antenna configuration	$(\underline{M}, \underline{N}, \underline{P}, \underline{M}_g, \underline{N}_g) = (1, 1, 2, 1, 1)$, the polarization angles are $0°$ and $90°$
Channel model	CDL-A
Delay spread	300 ns
DL precoder	Eigen zero-forcing (EZF) [18]
UE receiver	MMSE-IRC
CSI delay	4 ms
Number of paths	460

tuple $(\underline{M}, \underline{N}, \underline{P}, \underline{M}_g, \underline{N}_g)$ in Table I means the antenna array is composed of $\underline{M}_g\underline{N}_g$ panels of UPAs with $\underline{M}_g$ being the number of panels in a column and $\underline{N}_g$ the number of panels in a row. Furthermore, each antenna panel has $\underline{M}$ rows and $\underline{N}$ columns of antenna elements in the same polarization. The number of polarizations is always $\underline{P} = 2$. Thus the total number of antennas is $\underline{M} \times \underline{N} \times \underline{P} \times \underline{M}_g \times \underline{N}_g$ for a certain BS or UE. We consider 20 MHz of bandwidth where one channel estimate per each resource block (RB) is available in frequency domain. Note that although in this section the speeds of all UEs are assumed equal, our algorithms work well in a more realistic setting that multiple UEs move at different speeds. This is because the prediction is performed on a per UE basis.

We first ignore the channel estimation error and plot the performance of spectral efficiency as a function of SNR. We show the performance of PAD algorithm with 60 km/h of UE speed in Fig 2. The performances of different UE speeds without channel prediction are also added as reference curves. The curves labeled by "FIR Wiener" are obtained by the use of a classical linear predictor based on AR modeling of channel variations (for instance as proposed by [11]). We may observe from Fig. 2 that in a noise-free channel estimation setting, our proposed algorithms nearly approach the ideal case where the channel is stationary. It is interesting to note that our PAD algorithm outperforms the low-mobility setting of 3 km/h without prediction. Note that the FIR Wiener predictor gives only moderate prediction gains. In fact it is not performing as well because models that account for multipath space-time

structure (such as the one in [13]) do not necessarily conform with the simple AR(1) channel aging model.
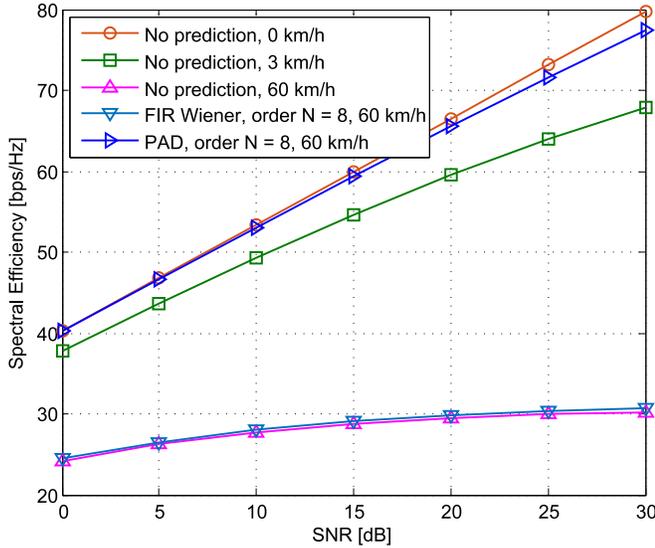
Fig. 2. The spectral efficiency vs. SNR, $N_t = 64$, noise-free channel samples.

Now the channel estimation error is taken into consideration, assuming the ratio between the channel power and the power of estimation noise is 20 dB. we plot the performances of our PAD algorithm with the denoising methods given by Sec. III-E in Fig. 3. Our proposed PAD algorithm with
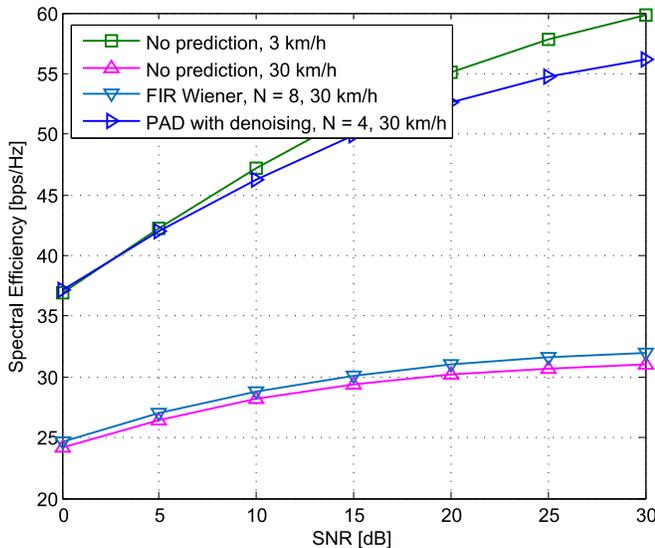
Fig. 3. The spectral efficiency vs. SNR, $N_t = 32$, noisy channel samples.

denoising methods in moderate mobility scenario of 30 km/h is very close to the low-mobility scenario of 3 km/h.

V. CONCLUSIONS

In this paper we addressed the practical challenge of massive MIMO - the mobility problem. We proposed Prony-based angular-delay domain channel prediction method which is

based on the specific angle-delay-Doppler structure of the channel and relies on the high spatial-frequency resolution in 5G massive MIMO. Our theoretical analysis proves that the proposed PAD method is able to achieve asymptotically error-free prediction, provided that only two accurate channel samples are available. We also proposed a supplementary channel denoising method for PAD algorithm to deal with channel sample error. Simulation results show that in moderate mobility scenarios, our proposed methods achieve nearly ideal performance of stationary setting.

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] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[4] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[5] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, 2016.

[6] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: exploiting all pilots for interference suppression," *EURASIP J. Wirel. Comm.*, vol. 2017, no. 1, Jun 2017.

[7] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[8] W. Wang, W. Liang, B. Li, L. Gu, J. Sheng, P. Qiu, J. Wang, and Y. Wang, "Field trial on TDD massive MIMO system with polar code," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017, pp. 1–6.

[9] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, "The degrees of freedom region of temporally correlated MIMO networks with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 494–514, Jan. 2014.

[10] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[11] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[12] R. Kumaresan and D. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 6, pp. 833–840, 1982.

[13] 3GPP, *Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 15)*. Technical Report TR 38.901, available: http://www.3gpp.org.

[14] G. R. B. Prony, "Essai expéimental et analytique sur les lois de la dilatabilité de fluides éastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool à différentes températures," *Journal de L'École Polytechnique*, vol. 1, no. 2, pp. 24–76, 1795.

[15] 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.

[16] H. Yin, H. Wang, Y. Liu, and D. Gesbert, "Addressing the curse of mobility in massive MIMO with Prony-based angular-delay domain channel predictions," *submitted*, 2019. [Online]. Available: http://arxiv.org/abs/1912.11330

[17] 3GPP, *NR; Physical Channels and Modulation (Release 15)*. Technical Specification TS 38.211, available: http://www.3gpp.org.

[18] L. Sun and M. R. McKay, "Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, Oct. 2010.