Estimation of Angular Power Spectrum Using Multikernel Adaptive Filtering

Eiji Ninomiya[*], *Masahiro Yukawa*[*], *Renato L. G. Cavalcante*[†], and *Lorenzo Miretti*[†‡]
[*] Department of Electronics and Electrical Engineering, Keio University, Yokohama, Japan
E-mail: ninomiya@ykw.elec.keio.ac.jp, yukawa@elec.keio.ac.jp, Tel: +81(0)45-566-1521
[†] Fraunhofer Heinrich Hertz Institute, Berlin, Germany
E-mail: renato.cavalcante@hhi.fraunhofer.de, lorenzo.miretti@hhi.fraunhofer.de
[‡] Technical University of Berlin, Berlin, Germany

***Abstract*—This paper addresses the problem of estimating the angular power spectrum (APS) of massive multiple input multiple output wireless channels. Estimating the APS is useful, for instance, for simplifying the downlink channel estimation problem in frequency division duplex systems. We propose an efficient online algorithm that estimates the APS from the channel spatial covariance matrix. The proposed algorithm approximates the APS as a sum of Gaussian functions and leverages the framework of multikernel adaptive filtering.**

***Index Terms*—Massive MIMO, angular power spectrum, covariance matrix, multikernel adaptive filtering.**

I. Introduction

The angular power spectrum (APS) estimation arises in important massive multiple input multiple output (MIMO) applications, such as estimation of the downlink (DL) covariance matrix [1], [2], pilot decontamination [3], and localization [4]. Many of these studies estimate the APS from measurements of the spatial covariance matrix, which is an inverse problem that cannot be solved analytically in general. This has been addressed in [1], [2] using the projection methods based on some prior knowledge about the APS, but estimation with high accuracy remains a major challenge to be tackled.

In this paper, we propose an efficient APS estimation scheme based on the multikernel adaptive filtering [5] with Gaussian functions. This approach is motivated by the empirical observation that typical APS can be expressed as the sum of few realtively smooth bell-shaped functions. The Gaussian model can express smooth functions efficiently, and this suggests that our model may fit the APS estimation problem. The multikernel adaptive filteringing algorithm updates the parameters (weights, scales, and centers) of the Gaussian functions by exploiting knowledge of the uplink (UL) channel covariance matrix. It also reduces the number of Gaussian functions used in our model, thereby yielding sparse estimates. Simulations show that (i) the proposed scheme yields highly accurate estimates of APS from noisy covariance matrices as well as array response vectors, and (ii) it leads to accurate estimation of the DL covariance matrix in massive MIMO channels in the frequency division duplex (FDD) mode.

E. Ninomiya and M. Yukawa were supported by JST SICORP Grant Number JPMJSC20C6, Japan. R. L. G. Cavalcante and L. Miretti acknowledge the financial support of the Federal Ministry of Education and Research of Germany (BMBF) under grant 01DR21009.

The rest of the paper is organized as follows. After stating the APS estimation problem in Section II, we present the proposed algorithm, and relation to prior work in Section III. The simulation results in terms of APS and DL channel covariance matrix estimation error are presented in Section IV, followed by conclusions in Section V.

Notation: The vector spaces of real, complex, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{N}$. $\mathbb{R}_+$ denotes positive real space. We use boldface to denote vectors and matrices. $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote respectively the transpose and Hermitian transpose. Throughout this paper, $(\cdot)^{\mathrm{u}}$ and $(\cdot)^{\mathrm{d}}$ denote the UL and DL systems, respectively.

II. Problem Statement

For simplicity, we consider the MIMO channel between a single-antenna user equipment (UE) and a base station (BS) with $N \in \mathbb{N}$ antennas in a 2D (azimuth-only) scenario [1], [2], [6]. The UL covariance matrix is given by

$$\mathbf{R}^{\mathrm{u}} = \int_{-\pi/2}^{\pi/2} \rho^{\star}(\theta)\mathbf{a}^{\mathrm{u}}(\theta)\mathbf{a}^{\mathrm{u}}(\theta)^{\mathsf{H}} d\theta, \tag{1}$$

where $\rho^{\star} : [-\pi/2, \pi/2] \to \mathbb{R}_+$ is the APS, which determines the average received/transmitted power per unit angle,

$$\mathbf{a}^{\mathrm{u}}(\theta) = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{j2\pi\frac{d}{\lambda}\sin\theta} & \dots & e^{j2\pi\frac{d}{\lambda}(N-1)\sin\theta} \end{bmatrix}^{\mathsf{T}} \tag{2}$$

is the response vector of the uniform linear array (ULA) with the antenna spacing $d \in \mathbb{R}$, and wavelength $\lambda \in \mathbb{R}$.

In this paper, we seek to estimate $\rho^{\star}$ satisfying (1), given $\mathbf{R}^{\mathrm{u}}$ and $\mathbf{a}^{\mathrm{u}}(\theta)$. In the next section, we propose a method to solve this problem. Previous methods employ algorithms that estimate the APS from the observed covariance matrix $\mathbf{R}^{\mathrm{u}}$ [1], [2]. However, this generally requires solving an ill-posed inverse problem. Therefore, we propose a new method to estimate the APS with high accuracy by modeling the APS as the sum of Gaussian functions, using prior knowledge about the APS.

III. Proposed Methods

First, we present the adopted APS approximated model using Gaussian functions. Second, we present time-varying cost functions which will be reduced in an online fashion by

978-616-590-477-3

multikernel adaptive filters. Third, we explain the dictionary construction scheme which is a part of the multikernel adaptive filtering algorithm. Fourth, the update rule for the model parameters is explained. Finally, the relation to the existing studies is presented.

A. APS modeling

Let us define a dictionary of some prespecified size p as the following set of functions:

$$\mathcal{D} := \{k(\cdot; c_j, \xi_j)\}_{j=1,\cdots,p}, \tag{3}$$

where $k(\cdot; c_j, \xi_j) : [-\pi_2, \pi_2] \to \mathbb{R}_+ : \theta \mapsto \exp(-(\theta - c_j)^2/\xi_j)$ is the Gaussian function with centers $(c_j)_{j=1,\cdots p}$ and scales $(\xi_j)_{j=1,\cdots p}$. (The dictionary construction will be discussed in Section III-C.) Our APS model is then given as follows:

$$\hat{\rho}(\theta) = \sum_{j=1}^{p} \alpha_j k(\theta; c_j, \xi_j), \tag{4}$$

where $(\alpha_j)_{j=1,\cdots p} \subset \mathbb{R}_+^p$ are the weights of the Gaussians. Note here that we pose nonnegativity on α_j to ensure nonnegativity of $\hat{\rho}$. The modeling in (4) relies on the fact that the APS is typically a sparse and smooth function.

Due to the symmetric (Hermitian) structure of the covariance matrix, let $\boldsymbol{r}^{\mathrm{u}} \in \mathbb{R}^{N_{\boldsymbol{r}^{\mathrm{u}}}}$ denote a vectorization of the upper triangular parts of $\Re\{\mathbf{R}^{\mathrm{u}}\}$ and $\Im\{\mathbf{R}^{\mathrm{u}}\}$, where $N_{\boldsymbol{r}^{\mathrm{u}}} := N(2N+1)$, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ represent the real and imaginary parts, respectively.[1] Then, the mth component of $\boldsymbol{r}^{\mathrm{u}}$ is given by

$$r_m^{\mathrm{u}} = \int_{-\pi/2}^{\pi/2} \rho^\star(\theta) g_m^{\mathrm{u}}(\theta) d\theta \quad m = 1, 2, \cdots, N_{\boldsymbol{r}^{\mathrm{u}}}, \tag{5}$$

where $g_m^{\mathrm{u}} : [-\pi/2, \pi/2] \to \mathbb{R}$ is the mth component of the analogous vectorization of the matrix $\mathbf{a}^{\mathrm{u}}(\theta)\mathbf{a}^{\mathrm{u}}(\theta)^{\mathsf{H}}$.

Replacing the true APS $\rho^\star$ in (5) by its estimate $\hat{\rho}$ in (4), the corresponding estimate $\hat{r}_m^{\mathrm{u}}$ of r_m^{u} is given as follows:

$$\begin{aligned} \hat{r}_m^{\mathrm{u}} &= \int_{-\pi/2}^{\pi/2} \hat{\rho}(\theta) g_m^{\mathrm{u}}(\theta) d\theta \\ &\approx \frac{\pi}{M} \sum_{i=1}^{M} \left(\sum_{j=1}^{p} \alpha_j k(\theta_i; c_j, \xi_j) \right) g_m^{\mathrm{u}}(\theta_i), \end{aligned} \tag{6}$$

where $\theta_1, \theta_2, \cdots, \theta_M \in [-\pi/2, \pi/2]$ are sample points. Now, the set of our kernel parameters to learn is $\beta := \{(\alpha_j, \xi_j, c_j)\}_{j=1,\cdots,p} \in \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}^p$. Let $\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \cdots, \alpha_p]^{\mathsf{T}} \in \mathbb{R}_+^p$. Given the dictionary $\mathcal{D}$, our cost function to minimize is given as follows:

$$J(\beta) := \sum_{m=1}^{N_{\boldsymbol{r}^{\mathrm{u}}}} (\hat{r}_m^{\mathrm{u}} - r_m^{\mathrm{u}})^2 + \lambda \Omega(\boldsymbol{\alpha}), \tag{7}$$

where $\Omega(\boldsymbol{\alpha})$ is a sparsity promoting regularizer with the regularization parameter $\lambda > 0$.

[1] If the matrix $\mathbf{R}^{\mathrm{u}}$ has a Toeplitz structure, the size of $\boldsymbol{r}^{\mathrm{u}}$ could be reduced further to $2N$ by stacking the real and imaginary parts of only the first column, for instance.

B. Time varying cost function

We presented the basic idea of our APS modeling in the previous subsection. In practice, the number, as well as the locations and widths, of peaks in the APS is unknown. The dictionary and its size are therefore updated during the learning process. The time-dependent dictionary of time-variable size p_n is defined as

$$\mathcal{D}_n := \{k(\cdot; c_{j,n}, \xi_{j,n})\}_{j=1,\cdots,p_n}, \quad n \in \mathbb{N}, \tag{8}$$

where $c_{j,n} \in \mathbb{R}$ and $\xi_{j,n} > 0$ are the center and scale parameters at time n, respectively. The estimate at time n is then defined accordingly as follows:

$$\hat{\rho}_n(\theta) := \sum_{j=1}^{p_n} \alpha_{j,n} k(\theta; c_{j,n}, \xi_{j,n}), \quad j = 1, \cdots, p_n. \tag{9}$$

Since the locations, widths, and heights of peaks of the APS $\rho^\star$ are unknown, the parameters $\alpha_{j,n}$, $c_{j,n}$, and $\xi_{j,n}$ need to be need to be learned from the measurements, or more specifically, from the covariance information $r_1^{\mathrm{u}}, r_2^{\mathrm{u}}, \cdots, r_{N_{\boldsymbol{r}^{\mathrm{u}}}}^{\mathrm{u}}$. The estimate $\hat{r}_{m,n}^{\mathrm{u}}$ is given by

$$\hat{r}_{m,n}^{\mathrm{u}} := \frac{\pi}{M} \sum_{i=1}^{M} \left(\sum_{j=1}^{p_n} \alpha_j k(\theta_i; c_j, \xi_j) \right) g_m^{\mathrm{u}}(\theta_i). \tag{10}$$

We define the set of kernel parameters at time n as $\beta_n := \{(\alpha_j, \xi_j, c_j)\}_{j=1,\cdots,p_n} \in \mathbb{R}_+^{p_n} \times \mathbb{R}_+^{p_n} \times \mathbb{R}^{p_n}$. Let $\boldsymbol{\alpha} := [\alpha_{1,n}, \alpha_{2,n}, \cdots, \alpha_{p_n,n}]^{\mathsf{T}}$. Then, given the dictionary $\mathcal{D}_n$ of size p_n, the cost function is given by

$$\min_{\beta_n} J_n(\beta_n) := \sum_{m=1}^{N_{\boldsymbol{r}^{\mathrm{u}}}} (\hat{r}_{m,n}^{\mathrm{u}} - r_m^{\mathrm{u}})^2 + \lambda \Omega_n(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} \in \mathbb{R}^{p_n}, \tag{11}$$

where $\Omega_n(\boldsymbol{\alpha}) := \sum_{j=1}^{p_n} \omega_{j,n} |\alpha_j| + \iota_{\mathbb{R}_+^{p_n}}(\boldsymbol{\alpha})$, and $\lambda > 0$ is the regularization parameter. Here, $\sum_{j=1}^{p_n} \omega_{j,n} |\alpha_j|$ is the weighted ℓ_1 norm with weights $\omega_{j,n} > 0$, and the indicator function $\iota_{\mathbb{R}_+^{p_n}}(\boldsymbol{\alpha})$ returns 0 if $\boldsymbol{\alpha} \in \mathbb{R}_+^{p_n}$, and it returns $+\infty$ if $\boldsymbol{\alpha} \notin \mathbb{R}_+^{p_n}$, guaranteeing nonnegativity of the weights. The regularizer Ω_n sparsifies the vector $\boldsymbol{\alpha}_n$, and those dictionary elements associated with zero weights will be discarded to reduce the dictionary size.

Multikernel adaptive filtering algorithm updates the parameters based on instantaneous measurements. Since the number of "data" to be used for learning is finite (which is $N_{\boldsymbol{r}^{\mathrm{u}}}$ to be specific), we use those data periodically. Our "instantaneous" cost is thus given as follows:

$$J_n^{\mathrm{inst}}(\beta_n) := (\hat{r}_{m(n),n}^{\mathrm{u}} - r_{m(n)}^{\mathrm{u}})^2 + \lambda \Omega_n(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} \in \mathbb{R}^{p_n}, \tag{12}$$

where $m(n) := (n \mod N_{\boldsymbol{r}^{\mathrm{u}}}) + 1 \in \{1, 2, \cdots, N_{\boldsymbol{r}^{\mathrm{u}}}\}$.

C. Dictionary construction

The formulation in (11) is nonconvex, meaning that a typical iterative solution would depend on the initial value of the parameter β_n. To mitigate sensitivity to the initial condition, we use the multiscale screening method [5] which builds the dictionary by conducting novelty tests hierarchically with multiple "layers" corresponding to different "scales" of Gaussian kernel. The first layer has the coarsest "lens" corresponding to the largest scale, and the last layer has the finest one corresponding to the smallest scale. The low frequency component of the function is captured at the first layer, and the middle to high frequency components are gradually extracted in the subsequent layers. The novelty test at each layer involves two conditions: (i) the coherence condition [7], and (ii) the error condition. The new function is regarded to be novel (i) if it is sufficiently different from any of the dictionary elements, and (ii) if the estimation error is sufficiently large. The coherence condition is given by [7]

$$\text{coherrence}^{(q)} := \max_{j=1,\cdots,p_n} \left| k(\theta; c_j, \xi_{\text{init}}^{(q)}) \right| \leq \delta^{(q)}, \quad q \in \mathcal{Q},$$

where $\delta^{(q)} \in (0,1)$ is some prespecified threshold. The concrete steps are presented below.

1) Let $\xi_{\text{init}}^{(q)}, q \in \mathcal{Q} := \{1, 2, \cdots, Q\}$ be the initial kernel scales, where $\xi_{\text{init}}^{(1)} \geq \xi_{\text{init}}^{(2)} \geq \cdots \xi_{\text{init}}^{(Q)} > 0$.
2) The Gaussian function $k(\cdot, \theta; \xi_{\text{init}}^{(1)})$ centered at the new input θ for $\rho(\theta)$ enters the dictionary $\mathcal{D}_n$ selectively if the novelty condition at the first layer (at the coarsest level) is satisfied. Go to the second layer otherwise.
3) At the second layer, $k(\cdot, \theta; \xi_{\text{init}}^{(2)})$ enters the dictionary if the novelty condition at the second coarsest level is satisfied. Go to the third layer otherwise.
4) Repeat this procedure until the Qth (final) layer unless the function enters the dictionary at some intermediate layer.

D. Update of weights, scales, and centers

To derive the update equation of $\boldsymbol{\alpha}_n$, we rewrite $\hat{r}_{m,n}^{\text{u}}$ as

$$\hat{r}_{m,n}^{\text{u}} = \left\langle \boldsymbol{\alpha}, \frac{\pi}{M} \sum_{i=1}^{M} \boldsymbol{k}_n(\theta_i) g_m^{\text{u}}(\theta_i) \right\rangle, \tag{13}$$

where $\boldsymbol{k}_n := [k(\theta; c_{1,n}, \xi_{1,n}), \cdots, k(\theta; c_{p_n,n}, \xi_{p_n,n})]^{\mathsf{T}}$. An application of the adaptive proximal forward backward splitting algorithm [8], [?] to the instantaneous cost function J_n^{inst} yields the following update equation:

$$\boldsymbol{\alpha}_{n+1} := T\Big\{ \text{prox}_{\lambda\Omega_n} \Big(\boldsymbol{\alpha}_n + \mu^{(\alpha)} (r_{m(n)}^{\text{u}} - \boldsymbol{\alpha}_n^{\mathsf{T}} \boldsymbol{\kappa}_{m(n),n}) \boldsymbol{\kappa}_{m(n),n} \Big) \Big\}, \tag{14}$$

where $\boldsymbol{\kappa}_{m(n),n} := \frac{\pi}{M} \sum_{i=1}^{M} \boldsymbol{k}_n(\theta_i) g_{m(n)}^{\text{u}}(\theta_i)$, the operator T removes zero components to reduce the dictionary size. The proximity operator $\text{prox}_{\lambda\Omega_n} : \mathbb{R}^{p_n} \to \mathbb{R}^{p_n}$ is defined as $\text{prox}_{\lambda\Omega_n}(\boldsymbol{x}) := \arg\min_{\boldsymbol{y} \in \mathbb{R}^{p_n}} (\Omega_n(\boldsymbol{y}) + \frac{1}{2\lambda} \|\boldsymbol{x} - \boldsymbol{y}\|^2), \boldsymbol{x} \in \mathbb{R}^{p_n}$. The ith component is $[\text{prox}_{\lambda\Omega_n}(\boldsymbol{x})]_i = \max\left\{1 - \frac{\lambda_{\omega_{i,n}}}{|x_i|}, 0\right\} x_i$. Using Proposition 1 of [9], we have $\text{prox}_{\lambda\Omega_n}(\boldsymbol{x}) = P_{\mathbb{R}_+^{p_n}}(\boldsymbol{x} - \lambda \mathbf{1})$ for $\boldsymbol{x} \in \mathbb{R}_+^{p_n}$, where $\mathbf{1}$ is the vector of ones. Here, the projection operator $P_{\mathbb{R}_+^{p_n}}$ to the nonnegative orthant $\mathbb{R}_+^{p_n}$ ensures nonnegativity of the weight vector $\boldsymbol{\alpha}_n$.

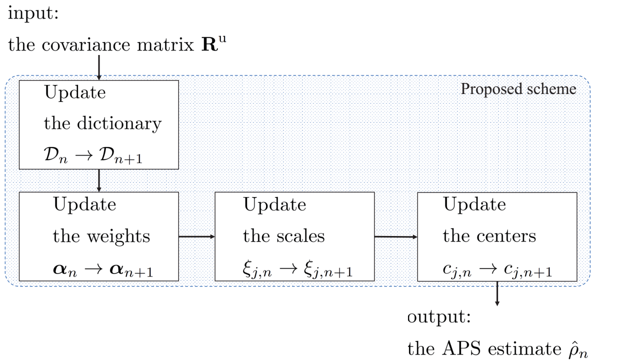

Fig. 1: A block diagram of the proposed scheme.

We also explain the update of the scale and the center. The update of the scale parameter $\xi_{j,n}$ is motivated by suppressing the cost J_n on $\mathbb{R}_+$. To update the scale parameter $\xi_{j,n}$ to reduce the cost function J_n within $\mathbb{R}_+$, the mirror descent method [10] is employed with the negative entropy, leading to the following update equation:

$$\xi_{j,n} = \arg\min_{\xi \in \mathbb{R}_+} \left\{ \left\langle \xi, \frac{\partial J_n(\beta_n)}{\partial \xi_j} \right\rangle + \frac{B_\phi(\xi \| \xi_{j,n})}{\mu_{j,n}^{(\xi)}} \right\}, \tag{15}$$

where $B_\phi(\xi \| \xi_{j,n}) := \phi(\xi) - \phi(\xi_{j,n}) - \langle \nabla\phi(\xi_{j,n}), \xi - \xi_{j,n} \rangle$ is the Bregman divergence associated with the continuous convex function $\phi(x) := x \log x - x, x > 0$, and $\mu_{j,n}^{(\xi)} = \xi_{j,n} \mu^{(\eta)}$ for some small constant $\mu^{(\eta)} > 0$ is the stepsize parameter. The partial differential in (15) is given by

$$\frac{\partial J_n(\beta_n)}{\partial \xi_j} = -\frac{2 e_n \alpha_{j,n} |\theta - c_{j,n}|^2 k(\theta; c_{j,n}, \xi_{j,n})}{\xi_{j,n}^2}. \tag{16}$$

The update equation of the kernel centers $c_{j,n}$ is given by

$$c_{j,n+1} = c_{j,n} - \mu^{(c)} \frac{\partial J_n(\beta_n)}{\partial c_j}, \tag{17}$$

where $\mu^{(c)} > 0$ is the stepsize parameter and

$$\frac{\partial J_n(\beta_n)}{\partial c_j} = -\frac{4 e_n \alpha_{j,n} k(\theta; c_{j,n}, \xi_{j,n})(\theta - c_{j,n})}{\xi_{j,n}}. \tag{18}$$

The computational complexity of the proposed scheme at each iteration is $O(N^2 p_n)$. A block diagram of the proposed scheme is shown in Fig. 1.

E. Relation to prior work

Previous studies [1], [2], [6] characterize a solution of the ill-posed inverse problems stated in Section II as a common point of multiple closed convex sets, which is computed by a projection method. This method has a limitation in estimating a "peaky" APS when the number of antennas cannot be large. In contrast, the proposed method estimates such a peaky APS efficiently as shown in the following section.

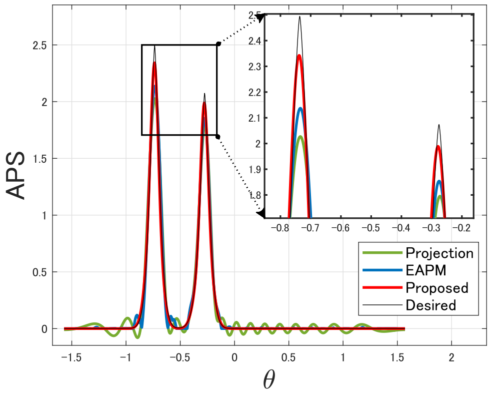

Fig. 2: The APS estimation results.

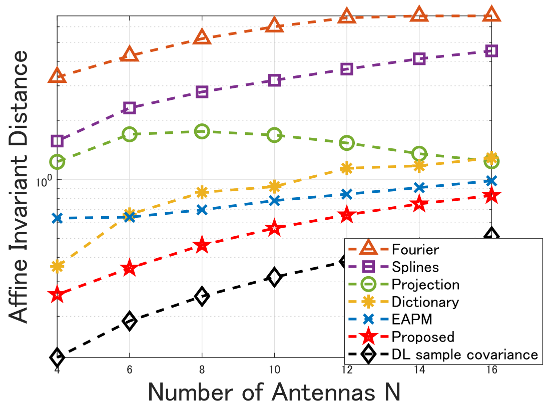

Fig. 4: A comparison of different DL covariance estimators.

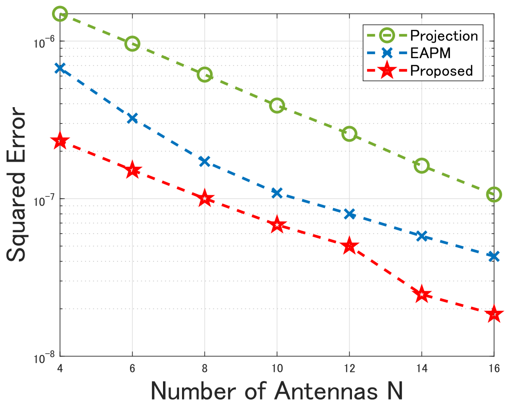

Fig. 3: A comparison of the APS estimators: the estimation errors vs the number of BS antennas N.

IV. Simulation Results

We show the efficacy of the proposed algorithm for the APS estimation for different numbers of antennas. In addition, we show the estimation results of the DL covariance matrix $\mathbf{R}^{\mathrm{d}}$ in massive MIMO channels in the FDD mode measure by the metric affine invariant distance. The true APS is modeled as

$$\rho(\theta) = \sum_{q=1}^{Q} \frac{\alpha_q}{\sqrt{2\pi\Delta_q^2}} \exp\left(\frac{|\theta - \phi_q|^{\frac{3}{2}}}{\Delta_q^2}\right), \tag{19}$$

where $Q = \{1, 2, 3, 4, 5\}$ is clusters of scatterers surrounding the BS and the UE, α_q are the weights satisfying $\sum_{q=1}^{Q} \alpha_q = 1$, ϕ_q is the mean from $[-\pi/3, \pi/3]$, and Δ_q^2 is distributions from $[3°, 5°]$. The domain of θ is divided into one million segments.

The ULA wavelength is $\lambda = c/f$, where $c = 3.0 \times 10^8$ m/s is the light velocity and f is the frequency. The UL/DL frequencies $f_{\mathrm{UL}}, f_{\mathrm{DL}}$ are 1.8 GHz and 1.9 GHz, respectively. The antenna spacing d is set to a half of the UL wavelength.

Experiment 1: APS Estimation

Fig. 2 shows the result of the APS estimation. It is seen that the proposed method yields accurate APS estimates around the peak. Fig. 3 shows the APS estimation errors across the number of antennas used in practical MIMO channels, $N = 4$ to 16 (step 2). We randomly generate $\rho(\theta)$ satisfying (19) with different weights, scales, and centers and take an average of 300 times. The initial scales for each kernel are $\xi_{\mathrm{init}}^{(1)} = 1, \xi_{\mathrm{init}}^{(2)} = 0.1, \xi_{\mathrm{init}}^{(3)} = 0.01$, the center step size is $\mu^{(c)} = 5.0 \times 10^{-3}$, the scale step size is $\mu^{(\xi)} = 0.1$, the weight step size is $\mu^{(\alpha)} = 0.1$, the regularization parameter is $\lambda = 1.0 \times 10^{-5}$. As can be seen in the figure, the proposed method shows high accuracy in estimating the APS. In particular, the results show that the APS estimation outperforms the existing methods [1] for $N = 8$ and 16. We finally mention that, in our additional experiments with different angles of arrival (i.e., different locations of peaks), there were no significant differences in estimation performance.

Experiment 2: Estimation of DL Covariance Matrix $\mathbf{R}^{\mathrm{d}}$

Next, we show the results of estimating the DL covariance matrix $\mathbf{R}^{\mathrm{d}}$ in the applied application in massive MIMO channel in the FDD mode. The estimate DL covariance matrix $\hat{\mathbf{R}}^{\mathrm{d}}$ is expressed as

$$\hat{\mathbf{R}}^{\mathrm{d}} = \int_{-\pi/2}^{\pi/2} \hat{\rho}(\theta)\mathbf{a}^{\mathrm{d}}(\theta)\mathbf{a}^{\mathrm{d}}(\theta)^{\mathsf{H}} d\theta, \tag{20}$$

where $\mathbf{a}^{\mathrm{d}}(\theta)$ is the DL array response formula. A comparison of the results of the $\mathbf{R}^{\mathrm{d}}$ calculation using $\hat{\rho}$ estimated by our method and a conventional method is shown in Fig. 4. The error between the estimated covariance matrix $\hat{\mathbf{R}}$ and true one $\mathbf{R}$ is defined by $\|\log(\mathbf{R}^{\frac{1}{2}}\hat{\mathbf{R}}^{-1}\mathbf{R}^{\frac{1}{2}})\|_F$, where $\log(\boldsymbol{A}) := \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(\boldsymbol{A} - \boldsymbol{I}_n)^k$. This is the affine invariant distance in the Riemannian space of positive semidefinite matrices [11], [12]. Fig. 4 shows that the proposed method outperforms the compared methods for all array sizes.

V. Conclusion

In this paper, we proposed an algorithm achieving high accuracy by modeling APS as a sum of Gaussian function and using a multikernel adaptive filtering. In order to avoid excessive dictionary size, the algorithm guaranteed nonnegativity by removing zero and negative weights of the Gaussian function. We showed that the proposed method estimated locally arriving APS with a better system than conventional methods. As a major application, the estimation of the DL covariance matrix in massive MIMO systems operating in the the FDD mode was presented and its effectiveness was successfully demonstrated. An investigation of using different types of antennas (omnidirectional, dipole, etc.) is left as a future work.

VI. References

[1] L. Miretti, R. L. G. Cavalcante, and S. Stanczak, "Channel covariance conversion and modelling using infinite dimensional Hilbert spaces," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 69, 2021, pp. 3145–3159.

[2] R. L. G. Cavalcante, L. Miretti, and S. Stańczak, "Error Bounds for FDD Massive MIMO Channel Covariance Conversion with Set-Theoretic Methods," in *IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7.

[3] R. L. G. Cavalcante and S. Stańczak, "Channel Covariance Estimation in Multiuser Massive MIMO Systems with an Approach Based on Infinite Dimensional Hilbert Spaces," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5180–5184.

[4] A. Decurninge, L. G. Ordóñez, P. Ferrand, H. Gaoning, L. Bojie, Z. Wei, and M. Guillaud, "CSI-based Outdoor Localization for Massive MIMO: Experiments with a Learning Approach," in *15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–6.

[5] M. Takizawa and M. Yukawa, "Online Learning with Self-tuned Gaussian Kernels: Good Kernel-initialization by Multiscale Screening," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4863–4867.

[6] N. Agrawal, R. L. G. Cavalcante, and S. Stańczak, "Adaptive Estimation of Angular Power Spectra for Time-Varying MIMO Channels," in *IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021, pp. 96–100.

[7] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online Prediction of Time Series Data With Kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.

[8] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 3734–3737.

[9] M. Yukawa and H. Kagami, "Supervised nonnegative matrix factorization via minimization of regularized Moreau-envelope of divergence function with application to music transcription," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 2041–2066, 2018, special Issue on Recent advances in machine learning for signal analysis and processing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016003217306336

[10] A. S. Nemirovskij and D. B. Yudin, "Problem complexity and method efficiency in optimization," 1983.

[11] A. Decurninge, M. Guillaud, and D. T. M. Slock, "Channel Covariance Estimation in Massive MIMO Frequency Division Duplex Systems," in *IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.

[12] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.