AN ASYMPTOTICALLY OPTIMAL APPROXIMATION OF THE CONDITIONAL MEAN CHANNEL ESTIMATOR BASED ON GAUSSIAN MIXTURE MODELS

Michael Koller, Benedikt Fesl, Nurettin Turan, Wolfgang Utschick

Technische Universität München

ABSTRACT

This paper investigates a channel estimator based on Gaussian mixture models (GMMs). We fit a GMM to given channel samples to obtain an analytic probability density function (PDF) which approximates the true channel PDF. Then, a conditional mean estimator (CME) corresponding to this approximating PDF is computed in closed form and used as an approximation of the optimal CME based on the true channel PDF. This optimal estimator cannot be calculated analytically because the true channel PDF is generally not available. To motivate the GMM-based estimator, we show that it converges to the optimal CME as the number of GMM components is increased. In numerical experiments, a reasonable number of GMM components already shows promising estimation results.

***Index Terms*—** conditional mean channel estimation, Gaussian mixture models, expectation-maximization, machine learning, spatial channel model

1. INTRODUCTION

Channel estimation plays an important role in future mobile communications systems, e.g., [1–3]. For many system models, it is known that a conditional mean estimator (CME) yields mean square error (MSE) minimizing channel estimates. However, computing the CME in closed form requires analytic knowledge of the channel probability density function (PDF), which is generally not available. Even if the PDF is given, calculating the CME might not be possible analytically or not be tractable practically. Increasingly, advanced channel models (e.g., [4]) or simulators (e.g., [5, 6]) are used to generate large amounts of realistic channel data and the goal is to design estimation algorithms based on these. Data sets are also available for download [7] or can be obtained via a measurement campaign.

The main idea of this paper is to use the available data to approximate the unknown channel PDF by means of a Gaussian mixture model (GMM). This is motivated by the good approximation properties of GMMs [8] and the increasing need to analyze collected data from measurement campaigns and available channel simulation tools or to characterize them by means of providing optimal estimators. Given a GMM approximation of the true channel PDF, we can analytically compute a corresponding CME, which is then used for the actual channel estimation. We show that this GMM-based estimator converges to the optimal CME as the number of GMM components increases.

In [9–11], e.g., GMMs are used to approximate or model the true channel PDF, e.g., for pilot design or channel clustering. To the best of our knowledge, the obtained GMMs have not been used to approximate the CME asymptotically optimally, which is studied in this work.

This work was supported by the Deutsche Forschungsgemeinschaft under grant UT 36/21.

2. SIGNAL MODEL AND OPTIMAL ESTIMATOR

We consider a single-antenna user who transmits pilots to an N-antennas base station which receives

$$\mathbb{C}^N \ni \boldsymbol{y} = \boldsymbol{h} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{h} \in \mathbb{C}^N$ is the channel and $\boldsymbol{n}$ is additive white Gaussian noise with mean zero and covariance $\boldsymbol{\Sigma} \in \mathbb{C}^{N\times N}$. We denote the noise PDF by $f_{\boldsymbol{n}}$. The channel is assumed to be described by means of a continuous PDF $f_{\boldsymbol{h}}$.

It is well known that the MSE-optimal channel estimator is given by the CME

$$\hat{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h} \mid \boldsymbol{y}] = \int \boldsymbol{h} f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h} \mid \boldsymbol{y}) d\boldsymbol{h}. \tag{2}$$

The conditional PDF of the channel given the observation is

$$f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h} \mid \boldsymbol{y}) = \frac{f_{\boldsymbol{y}|\boldsymbol{h}}(\boldsymbol{y} \mid \boldsymbol{h}) f_{\boldsymbol{h}}(\boldsymbol{h})}{f_{\boldsymbol{y}}(\boldsymbol{y})} = \frac{f_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{h}) f_{\boldsymbol{h}}(\boldsymbol{h})}{f_{\boldsymbol{y}}(\boldsymbol{y})} \tag{3}$$

where $f_{\boldsymbol{y}}$ is the PDF of the receive signal. This paper proposes a method to approximate (2) with the help of GMMs.

3. CHANNEL ESTIMATION WITH GAUSSIAN MIXTURE MODELS

The optimal estimator (2) can generally not be computed analytically, but we can assume to have access to a set $\mathcal{H}_M = \{\boldsymbol{h}_m\}_{m=1}^M$ of channel samples. The following summarizes the proposed procedure to obtain a channel estimator using $\mathcal{H}_M$ and the details can be found in Section 3.1.

Motivated by universal approximation properties of GMMs [8], we fit a GMM $f_{\boldsymbol{h}}^{(K)}$ with K components to the given data $\mathcal{H}_M$ to approximate the unknown channel PDF $f_{\boldsymbol{h}}$. Then, we define a CME $\hat{\boldsymbol{h}}^{(K)}$ as in (2) but we replace $f_{\boldsymbol{h}|\boldsymbol{y}}$ with $f_{\boldsymbol{h}|\boldsymbol{y}}^{(K)}$ which is the conditional density based on the approximation $f_{\boldsymbol{h}}^{(K)}$ instead of the true PDF $f_{\boldsymbol{h}}$. By universal approximation, $f_{\boldsymbol{h}}^{(K)}$ converges to the true channel PDF $f_{\boldsymbol{h}}$ for $K \to \infty$. We show that this implies the convergence of the proposed estimator $\hat{\boldsymbol{h}}^{(K)}$ to the optimal estimator $\hat{\boldsymbol{h}}$ for $K \to \infty$. Lastly, we present how $\hat{\boldsymbol{h}}^{(K)}$, in contrast to $\hat{\boldsymbol{h}}$, can always be expressed in closed form.

3.1. Channel Estimator

By [8, Theorem 5] (universal approximation property of GMMs), there exists a sequence $(f_{\boldsymbol{h}}^{(K)})_{K=1}^{\infty}$ of GMMs which converges to the channel PDF $f_{\boldsymbol{h}}$ in the sense that

$$\lim_{K\to\infty} \|f_{\boldsymbol{h}} - f_{\boldsymbol{h}}^{(K)}\|_\infty = 0 \tag{4}$$

holds. For every K, we define an estimator

$$\hat{\boldsymbol{h}}^{(K)} = \mathrm{E}^{(K)}[\boldsymbol{h} \mid \boldsymbol{y}] = \int \boldsymbol{h} f_{\boldsymbol{h}|\boldsymbol{y}}^{(K)}(\boldsymbol{h} \mid \boldsymbol{y}) d\boldsymbol{h} \tag{5}$$

where, in analogy to (3), we have

$$f_{\boldsymbol{h}|\boldsymbol{y}}^{(K)}(\boldsymbol{h} \mid \boldsymbol{y}) = \frac{f_{\boldsymbol{n}}(\boldsymbol{y}-\boldsymbol{h}) f_{\boldsymbol{h}}^{(K)}(\boldsymbol{h})}{f_{\boldsymbol{y}}^{(K)}(\boldsymbol{y})}. \tag{6}$$

Here, $f_{\boldsymbol{y}}^{(K)}$ is the PDF of the receive signal (1) in the case where the channel $\boldsymbol{h}$ is distributed according to $f_{\boldsymbol{h}}^{(K)}$. With this notation, we have the following theorem.

Theorem 1. *For any $\boldsymbol{y} \in \mathbb{C}^N$, the estimator* (5) *converges to the optimal estimator* (2) *in the sense that it holds:*

$$\lim_{K\to\infty} \|\hat{\boldsymbol{h}} - \hat{\boldsymbol{h}}^{(K)}\| = 0. \tag{7}$$

This theorem gives a strong motivation to use (5) as a channel estimator for some fixed K. Due to space limitations, we only sketch the proof of the theorem in Section 3.2 and provide the details in a future work. Before we turn to the sketch, we explain how to compute (5) in closed form in practice.

A GMM with K components is a PDF of the form

$$f_{\boldsymbol{h}}^{(K)}(\boldsymbol{h}) = \sum_{k=1}^{K} p(k) \mathcal{N}_{\mathbb{C}}(\boldsymbol{h}; \boldsymbol{\mu}_k, \boldsymbol{C}_k). \tag{8}$$

with K *mixing coefficients* $p(k)$ as well as K means $\boldsymbol{\mu}_k \in \mathbb{C}^N$ and covariances $\boldsymbol{C}_k \in \mathbb{C}^{N\times N}$. Here, $\mathcal{N}_{\mathbb{C}}(\boldsymbol{h}; \boldsymbol{\mu}_k, \boldsymbol{C}_k)$ is a Gaussian PDF with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{C}_k$ evaluated at $\boldsymbol{h}$. An expectation-maximization (EM) algorithm in conjunction with given data samples can be used to obtain the parameters in (8). An introduction to GMMs can, e.g., be found in [12].

With a GMM (8) given, it remains to calculate the estimator $\hat{\boldsymbol{h}}^{(K)}$ in (5). To this end, we write (5) as [13]

$$\hat{\boldsymbol{h}}^{(K)} = \sum_{k=1}^{K} p(k \mid \boldsymbol{y}) \, \mathrm{E}^{(K)}[\boldsymbol{h} \mid \boldsymbol{y}, k]. \tag{9}$$

In this expression, $\boldsymbol{h}$ is distributed according to $f_{\boldsymbol{h}}^{(K)}$. This implies that conditioned on one of the components k, the observation $\boldsymbol{y} \mid k$ is a sum of two Gaussian random variables ($\boldsymbol{h} \mid k$ plus noise $\boldsymbol{n}$). Hence, the conditional expectation $\mathrm{E}^{(K)}[\boldsymbol{h} \mid \boldsymbol{y}, k]$ can be computed using the well-known linear minimum mean square error (LMMSE) formula

$$\mathrm{E}^{(K)}[\boldsymbol{h} \mid \boldsymbol{y}, k] = \boldsymbol{C}_k(\boldsymbol{C}_k + \boldsymbol{\Sigma})^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k. \tag{10}$$

Additionally, we obtain the PDF of $\boldsymbol{y}$ as

$$f_{\boldsymbol{y}}^{(K)}(\boldsymbol{y}) = \sum_{k=1}^{K} p(k) \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}; \boldsymbol{\mu}_k, \boldsymbol{C}_k + \boldsymbol{\Sigma}). \tag{11}$$

This is a GMM as well. GMMs generally allow to compute the so-called *responsibilities* by evaluating Gaussian PDFs [12]:

$$p(k \mid \boldsymbol{y}) = \frac{p(k) \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}; \boldsymbol{\mu}_k, \boldsymbol{C}_k + \boldsymbol{\Sigma})}{\sum_{i=1}^{K} p(i) \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}; \boldsymbol{\mu}_i, \boldsymbol{C}_i + \boldsymbol{\Sigma})}. \tag{12}$$

Summarily, plugging (12) and (10) into (9) yields a closed-form expression for the channel estimator $\hat{\boldsymbol{h}}^{(K)}$.

3.2. Proof Sketch

Note that because GMMs can approximate any continuous PDF arbitrarily well [8], there always exists a sequence of GMMs which converges to $f_{\boldsymbol{h}|\boldsymbol{y}}$ for any given observation $\boldsymbol{y}$. In light of this, proving Theorem 1 seems to be a straightforward application of the universal approximation property of GMMs. However, the GMM sequence would then depend on $\boldsymbol{y}$ so that every new observation $\boldsymbol{y}$ would require a new sequence of GMMs to approximate the optimal estimator. This is not desirable. Instead, we only want one sequence $(f_{\boldsymbol{h}}^{(K)})_{K=1}^{\infty}$ of GMMs which converges to $f_{\boldsymbol{h}}$ and we show that this still implies the convergence of $\hat{\boldsymbol{h}}^{(K)}$ to $\hat{\boldsymbol{h}}$ for any given $\boldsymbol{y}$.

Since $\boldsymbol{y}$ is the sum of two random vectors, the PDF $f_{\boldsymbol{y}}$ is the convolution of the PDFs $f_{\boldsymbol{h}}$ and $f_{\boldsymbol{n}}$. Analogously, $f_{\boldsymbol{y}}^{(K)}$ is the convolution of $f_{\boldsymbol{h}}^{(K)}$ and $f_{\boldsymbol{n}}$. A standard argument then shows that (4) implies $\lim_{K\to\infty} \|f_{\boldsymbol{y}} - f_{\boldsymbol{y}}^{(K)}\|_\infty = 0$. Next, we bound the quantity of interest:

$$\|\hat{\boldsymbol{h}} - \hat{\boldsymbol{h}}^{(K)}\| \le \int \|\boldsymbol{h}\| \left| f_{\boldsymbol{h}|\boldsymbol{y}}(\boldsymbol{h} \mid \boldsymbol{y}) - f_{\boldsymbol{h}|\boldsymbol{y}}^{(K)}(\boldsymbol{h} \mid \boldsymbol{y}) \right| d\boldsymbol{h}$$

$$= \int \|\boldsymbol{h}\| |f_{\boldsymbol{n}}(\boldsymbol{y}-\boldsymbol{h})| \left| \frac{f_{\boldsymbol{h}}(\boldsymbol{h})}{f_{\boldsymbol{y}}(\boldsymbol{y})} - \frac{f_{\boldsymbol{h}}^{(K)}(\boldsymbol{h})}{f_{\boldsymbol{y}}^{(K)}(\boldsymbol{y})} \right| d\boldsymbol{h} \tag{13}$$

$$\le \sup_{\boldsymbol{h}\in\mathbb{C}^N} \left| \frac{f_{\boldsymbol{h}}(\boldsymbol{h})}{f_{\boldsymbol{y}}(\boldsymbol{y})} - \frac{f_{\boldsymbol{h}}^{(K)}(\boldsymbol{h})}{f_{\boldsymbol{y}}^{(K)}(\boldsymbol{y})} \right| \int \|\boldsymbol{h}\| f_{\boldsymbol{n}}(\boldsymbol{y}-\boldsymbol{h}) d\boldsymbol{h}. \tag{14}$$

For the equality, we used (3) and (6). The integral in (14) can be shown to be finite for any given $\boldsymbol{y}$ so that Theorem 1 is proved as soon as the supremum in (14) is shown to converge to zero for $K \to \infty$. This takes a bit more effort but is ultimately a consequence of the uniform convergence of $f_{\boldsymbol{h}}^{(K)}$ to $f_{\boldsymbol{h}}$ and of $f_{\boldsymbol{y}}^{(K)}$ to $f_{\boldsymbol{y}}$. Details will be presented in a future work.

4. NUMERICAL EXPERIMENTS

Numerical experiments using two different channel models are presented after a review of related algorithms.

4.1. Channel Models

In all simulations, the noise covariance matrix is $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I} \in \mathbb{C}^{N \times N}$. Generated channels are normalized such that $\mathrm{E}[\|\boldsymbol{h}\|^2] = N$ holds which allows us to define a signal-to-noise ratio (SNR) as $\mathrm{SNR} = 1/\sigma^2$. The number of antennas is $N = 128$, and we generate $190 \cdot 10^3$ training (e.g., for fitting a GMM using an EM algorithm) and $10 \cdot 10^3$ testing samples. Unless stated otherwise, the GMM-based estimator uses $K = 128$ components.

4.1.1. 3GPP

We work with a spatial channel model [4] where channels are modeled conditionally Gaussian: $\boldsymbol{h} \mid \boldsymbol{\delta} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\delta}})$. The random vector $\boldsymbol{\delta}$ collects the angles of arrivals and path gains of the main propagation clusters between a mobile terminal and the base station. The base station employs a uniform linear array (ULA) such that the channel covariance matrix can be computed as $\boldsymbol{C}_{\boldsymbol{\delta}} = \int_{-\pi}^{\pi} g(\theta; \boldsymbol{\delta}) \boldsymbol{a}(\theta) \boldsymbol{a}(\theta)^{\mathrm{H}} d\theta$. Here, $\boldsymbol{a}(\theta) = [1, \exp(j\pi \sin(\theta), \dots, \exp(j\pi(N-1)\sin(\theta)]^{\mathrm{T}}$ is the array steering vector for an angle of arrival θ and g is a power density consisting of a sum of weighted Laplace densities whose standard deviations describe the angle spread of the propagation clusters [4]. For every channel sample, we generate random angles and path gains $\boldsymbol{\delta}$ and then draw the sample as $\boldsymbol{h} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\delta}})$.

4.1.2. QuaDRiGa

Version 2.4 of the QuaDRiGa channel simulator [5,6] is used to generate channel samples for the numerical experiments. We simulate an urban macrocell single carrier scenario at a frequency of 2.53 GHz. The base station is equipped with a ULA with 128 "3GPP-3D" antennas and the mobile terminals employ an "omni-directional" antenna. The base station is placed at a height of 25 meters and covers a sector of 120°. The minimum and maximum distances between the mobile terminals and the base station are 35 meters and 500 meters, respectively. In 80% of the cases, the mobile terminals are located indoors at different floor levels, whereas the mobile terminals' height is 1.5 meters in the case of outdoor locations.

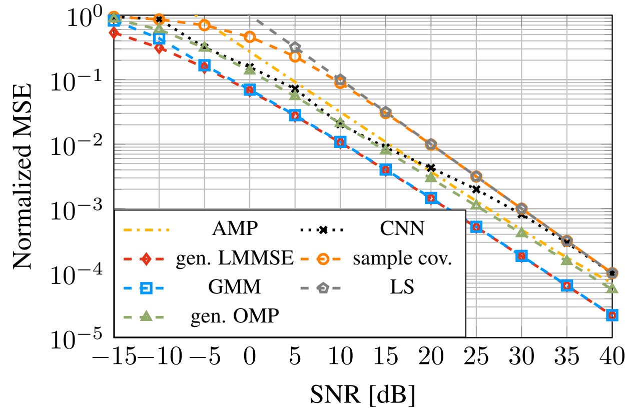

Fig. 1. Channel estimation using the model from Section 4.1.1 with one propagation cluster.

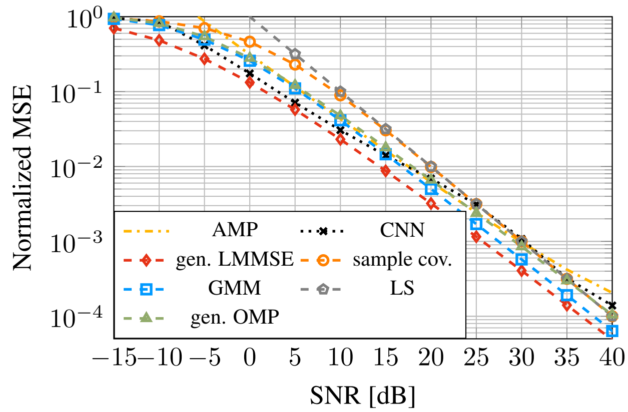

Fig. 2. Channel estimation using the model from Section 4.1.1 with three propagation clusters.

QuaDRiGa models channels as $\boldsymbol{h} = \sum_{\ell=1}^{L} \boldsymbol{g}_\ell e^{-2\pi j f_c \tau_\ell}$, where ℓ is the path number, and the number of multi-path components L depends on whether there is line of sight (LOS), non-line of sight (NLOS), or outdoor-to-indoor (O2I) propagation: $L_{\text{LOS}} = 37$, $L_{\text{NLOS}} = 61$, or $L_{\text{O2I}} = 37$. The carrier frequency is denoted by f_c and the ℓ-th path delay by τ_ℓ. The coefficients vector $\boldsymbol{g}_\ell$ consists of one complex entry for each antenna pair, which comprises the attenuation of a path, the antenna radiation pattern weighting, and the polarization [14]. As described in the QuaDRiGa manual, the generated channels are post-processed to remove the path gain.

4.2. Related Channel Estimators

Two obvious baseline channel estimators are the least squares (LS) estimator which simply computes $\hat{\boldsymbol{h}}_{\text{LS}} = \boldsymbol{y}$ in our case and the LMMSE formula based on the sample covariance matrix. For the latter, we use all $M = 190 \cdot 10^3$ training data to estimate a sample covariance matrix $\boldsymbol{C} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{h}_m \boldsymbol{h}_m^{\mathrm{H}}$ and then estimate channels as $\hat{\boldsymbol{h}}_{\text{sample cov.}} = \boldsymbol{C}(\boldsymbol{C} + \boldsymbol{\Sigma})^{-1} \boldsymbol{y}$.

In compressive sensing, the channel is assumed to be (approximately) sparse such that we have $\boldsymbol{h} \approx \boldsymbol{D}\boldsymbol{s}$ for a sparse vector $\boldsymbol{s} \in \mathbb{C}^L$. The *dictionary* $\boldsymbol{D} \in \mathbb{C}^{N \times L}$ is typically an oversampled discrete Fourier transform matrix (e.g., [15]). Compressive sensing algorithms then recover an estimate $\hat{\boldsymbol{s}}$ of $\boldsymbol{s}$ assuming $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{s} + \boldsymbol{n}$ and estimate the channel as $\hat{\boldsymbol{h}} = \boldsymbol{D}\hat{\boldsymbol{s}}$. One algorithm is orthogonal matching pursuit (OMP) [16–18] which we employ with $L = 4N$. OMP needs to know the sparsity order. Since order estimation is a difficult problem, we avoid it via a genie-aided approach: OMP gets access to the true channel to choose the optimal sparsity. This yields a performance bound for OMP. A more elaborate compressive sensing algorithm is approximate message passing (AMP) [19, 20] which does not need to know the sparsity order. We achieved the best results with $L = 2N$.

A convolutional neural network (CNN)-based channel estimator was introduced in [21]. The authors exploit knowledge about the ULA geometry and about the 3GPP channel model in order to derive its architecture. We use this CNN as described in [21] for both channel models introduced in Section 4.1. The CNN is always trained on data that corresponds to the channel model on which it is tested afterwards. The activation function is the rectified linear unit and we use the input transform based on the $2N \times 2N$ Fourier matrix, cf. [21, Equation (43)].

4.3. Numerical Results

Figs. 1 and 2 show the estimation performance in terms of normalized MSE when the channel model from Section 4.1.1 is used. Since for every channel, a random covariance matrix $\boldsymbol{C}_{\boldsymbol{\delta}}$ is generated and the channel is then drawn according to $\boldsymbol{h} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\delta}})$, we can use this channel covariance matrix to compute LMMSE channel estimates as $\hat{\boldsymbol{h}}_{\text{gen. LMMSE}} = \boldsymbol{C}_{\boldsymbol{\delta}}(\boldsymbol{C}_{\boldsymbol{\delta}} + \boldsymbol{\Sigma})^{-1}\boldsymbol{y}$ to provide a performance bound for all algorithms. We call this estimator "genie LMMSE" because the true $\boldsymbol{C}_{\boldsymbol{\delta}}$ is used which is not available in practice.

In Fig. 1, we investigate channels with one propagation cluster. It is interesting to see that the GMM-based estimator performs almost as well as the genie LMMSE estimator. In the mid-SNR range, the two compressive sensing algorithms are approximately equally good. In Fig. 2, we have three propagation clusters. A first observation is the strong performance of the CNN estimator in the mid-SNR range. Note that we can generally not expect any estimator to reach the genie LMMSE curve because it has more channel knowledge. In the higher SNR-range, the GMM-based estimator seems to be the only algorithm still outperforming LS estimation.

In Fig. 3, we concentrate on the QuaDRiGa channel model described in Section 4.1.2 where the channel covariance matrices and therefore the genie LMMSE curve are no longer available. Here, the two compressive sensing algorithms behave not as similarly as they did in the previous experiments. Additionally, their performance is not as convincing. A reason might be that the channels now are not sparse enough. The CNN estimator shows again a good performance and overall the GMM-based estimator can compete with it or is better.

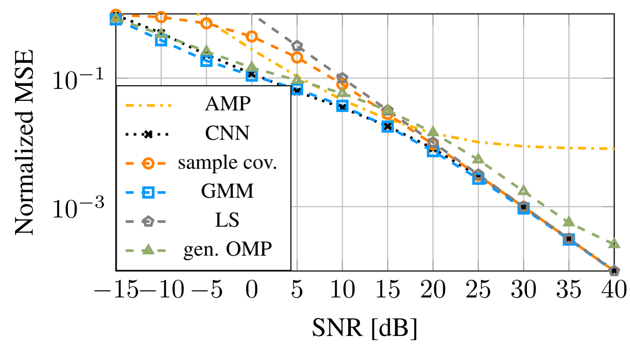

Fig. 3. Channel estimation using the model from Section 4.1.2.

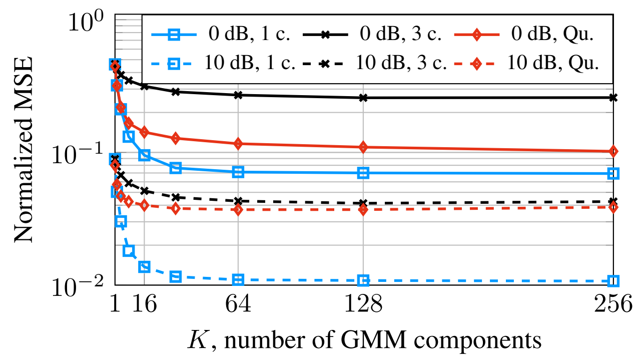

Fig. 4. Varying K for two different SNRs. Channel models with one cluster (1 c.), three clusters (3 c.), QuaDRiGa (Qu.).

Fig. 4 displays the last experiment where we analyze the effect of varying the number K of GMM components. The expected improvement of the MSE as K increases can be seen. For $K = 1$, the GMM-based estimator shows a performance almost identical to the sample covariance matrix-based estimator because the sample covariance matrix is exactly the maximum likelihood estimate in this case. We emphasize that for larger K, more training data is in general necessary to reach the GMM's performance limit because the GMM has more parameters for larger K. This reflects a typical trade-off between performance and complexity, which may also depend on the amount of available training data.

5. OUTLOOK

We have presented preliminary results of the GMM-based estimator in a simple channel estimation scenario that is representative of future applications for characterizing measurement campaigns or simulation tools. In a future work, a rigorous proof of Theorem 1 will be presented and we plan to analyze in particular the convergence behavior of the estimator. Also, we intend to generalize the proof to multiple-input multiple-output channel estimation where a pilot matrix is present.

6. REFERENCES

[1] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G. Larsson, Thomas L. Marzetta, Ove Edfors, and Fredrik Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] Daniel C. Araújo, André L. F. de Almeida, Johan Axnäs, and João C. M. Mota, "Channel estimation for millimeter-wave very-large MIMO systems," in *2014 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Sept. 2014, pp. 81–85.

[3] Mathew K. Samimi and Theodore S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, July 2016.

[4] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations," Tech. Rep. 25.996 (V16.0.0), 3rd Generation Partnership Project (3GPP), July 2020.

[5] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, 2014.

[6] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa: Quasi deterministic radio channel generator, user manual and documentation," Tech. Rep., Fraunhofer Heinrich Hertz Institute, v2.2.0, 2019.

[7] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. of Inf. Theory and Appl. Workshop (ITA)*, San Diego, CA, Feb. 2019, pp. 1–8.

[8] T. Tin Nguyen, Hien D. Nguyen, Faicel Chamroukhi, and Geoffrey J. McLachlan, "Approximation by finite mixtures of continuous density functions that vanish at infinity," *Cogent Math. Statist.*, vol. 7, no. 1, pp. 1750861, 2020.

[9] Chao-Kai Wen, Shi Jin, Kai-Kit Wong, Jung-Chieh Chen, and Pangan Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[10] Yupeng Li, Jianhua Zhang, Pan Tang, and Lei Tian, "Clustering in the wireless channel with a power weighted statistical mixture model in indoor scenario," *China Commun.*, vol. 16, no. 7, pp. 83–95, July 2019.

[11] Yujie Gu and Yimin D. Zhang, "Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2334–2346, May 2019.

[12] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.

[13] John T. Flam, Saikat Chatterjee, Kimmo Kansanen, and Torbjörn Ekman, "On MMSE estimation: A linear model under Gaussian mixture statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3840–3845, July 2012.

[14] Martin Kurras, Sida Dai, Stephan Jaeckel, and Lars Thiele, "Evaluation of the spatial consistency feature in the 3GPP geometry-based stochastic channel model," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2019, pp. 1–6.

[15] Ahmed Alkhateeb, Geert Leus, and Robert W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *2015 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2015, pp. 2909–2913.

[16] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[17] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems, and Computers*, Nov. 1993, pp. 40–44 vol.1, ISSN: 1058-6393.

[18] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[19] David L. Donoho, Arian Maleki, and Andrea Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Inf. Theory Workshop on Inf. Theory (ITW 2010, Cairo)*, Jan. 2010, pp. 1–5.

[20] Arian Maleki, Laura Anitori, Zai Yang, and Richard G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, July 2013.

[21] David Neumann, Thomas Wiese, and Wolfgang Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, June 2018.