

Low-Rank Approximations for Spatial Channel Models

Thomas Wiese, Lorenz Weiland, Wolfgang Utschick

Fachgebiet Methoden der Signalverarbeitung, Technical University of Munich, Germany

Email: {thomas.wiese, lorenz.weiland, utschick}@tum.de

Abstract—We analyze the problem of approximating vectors that are superpositions of many closely spaced steering vectors. Such vectors appear in realistic models for wireless communication channels and describe single clusters of scatterers. We question the practice of using a dictionary of steering vectors in order to find a sparse approximation. Alternative dictionaries can be obtained from the Karhunen-Loève expansion of the channel vectors or they can be learned from observed channel realizations. Furthermore, it is possible to restrict the allowed combinations of dictionary elements, thereby reducing the complexity of algorithms that find an approximation. We provide simulation results and discuss the performance of the various approaches.

Index Terms—Direction-of-arrival (DOA) estimation, Spatial channel models, Dictionary learning

I. INTRODUCTION

Wireless communication systems with large numbers of antennas are currently being investigated for use in future standards [1]. The dimension of the channel vector or matrix grows along with the number of antennas. Thereby, the geometric structure of the channel vector is revealed, which is that of a superposition of a number of propagation paths as suggested by preliminary measurement campaigns [2], [3]. The question arises whether for these systems, classical minimum mean squared error (MMSE) or least-squares (LS) channel estimation, both of which are agnostic to any kind of structure in the channel vector, can be improved upon by incorporating structural prior information.

Models used for *simulating* channels use superpositions of many propagation paths that are due to several localized clusters of scatterers. At least at the base station side, which is typically situated at an exposed location, all scatterers within the same cluster have the same angle to within few degrees, e.g., ± 5 degrees for urban macro cells and ± 2 degrees for urban micro cells in the spatial channel model (SCM) as defined by the ETSI 3rd Generation Partnership Project (3GPP) [4]. On the other hand, models used for *estimating* channels, use superpositions of a few distinct propagation paths, each of which is described by its angle and delay, see, e.g., [5]–[7]. In both cases, the resulting channel vector exhibits a low-dimensional structure, which can possibly be exploited to improve channel estimation.

There are many different ways to describe this low-dimensional structure and then there are different algorithms that find a representation of the channel vector given such a structure. The standard example is to assume that the channel can be described with few columns of an oversampled discrete

Fourier transform (DFT) matrix, i.e., few steering vectors that correspond to the angles of either the propagation paths or the centers of the clusters of scatterers. However, this is neither the only way to describe this low-dimensional structure, nor is it necessarily the best way. For example, let the channel vector \mathbf{h} be generated by 20 equal-power sub-paths with random coefficients at angles between -4.3101 and 4.3101 degrees (the urban microcell scenario in [4]). Perform a Karhunen-Loève expansion (KLE) of \mathbf{h} and only retain the k strongest components. By definition of the KLE, an approximation of a realization of a channel vector in the subspace spanned by those k vectors yields a smaller error (in the mean) than an approximation in the subspace spanned by any (a priori fixed) set of k DFT vectors. In this simple case, where clusters only occur around zero degrees, there is, thus, a more appropriate dictionary than the DFT dictionary.

The goal of the present paper is to compare different forms of sparse representations suitable for use with 3GPP spatial channel models. In particular, we investigate whether sparse combinations of columns of the oversampled DFT matrix are efficient to describe channels that are generated according to the clustered-scatterers channel model or if there are other, more efficient dictionaries.

II. PROBLEM FORMULATION

We use bold-face lower-case letters for vectors and bold-face upper-case letters for matrices. Let

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} [1 \quad \exp(i\pi \sin \theta) \quad \dots \quad \exp(i\pi(N-1) \sin \theta)]^T$$

denote the normalized steering vector, i.e., the signal recorded by a uniform linear array (ULA) of N sensors with inter-element spacing $\lambda/2$ at a given time-instant if a source in the far-field of the array and at an angle θ transmits a harmonic signal with wavelength λ . Our goal is to find an efficient sparse representation of a channel vector

$$\mathbf{h} = \sum_{\theta \in \Theta} y_{\theta} \mathbf{a}(\theta) \in \mathbb{C}^N \quad (1)$$

which is given as the superposition of steering vectors with angles in the set Θ . We assume that the set Θ is large, i.e., there are many propagation paths, but localized, i.e., all angles are relatively similar. For example, the angles in Θ can be obtained by drawing P samples from a Laplace distribution with mean δ and standard deviation σ and the coefficients y_{θ}

are selected as unit modulus with random phase. This model is used in the 3GPP spatial channel model [4]. The parameters σ , and P depend on the chosen scenario, e.g., urban macro cell ($P = 20$, $\sigma = 5^\circ$), urban microcell ($P = 20$, $\sigma = 2^\circ$), etc. The distribution of the cluster centers δ depends on the cell geometry, e.g., on the number of sectors per cell.

A dictionary (or code book) is a fat matrix $\mathbf{D} \in \mathbb{C}^{N \times M}$ of unit-norm vectors, e.g., an oversampled DFT matrix. For a given dictionary \mathbf{D} and channel vector \mathbf{h} , the optimal k th order approximation error is given by

$$\varepsilon_0(\mathbf{D}, \mathbf{h}) = \min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{h} - \mathbf{D}\mathbf{x}\|_2^2 \quad \text{s.t. } \|\mathbf{x}\|_0 \leq k \quad (2)$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of \mathbf{x} . Then, a dictionary performs well for approximating channel vectors \mathbf{h} , if the expected value of the optimal approximation error

$$\mathbb{E}_{\mathbf{h}}[\varepsilon_0(\mathbf{D}, \mathbf{h})] \quad (3)$$

is small. The expectation is with respect to the random variable \mathbf{h} , i.e., with respect to the set Θ and the coefficients y_θ .

Our first question is whether for a given parametrization of the distribution of the channel vector \mathbf{h} and a given sparsity parameter k , there is a dictionary \mathbf{D} of the same size as an oversampled DFT matrix, but which has a smaller mean approximation error. In Sec. III, we construct a dictionary from the Karhunen-Loève expansion of \mathbf{h} and we learn a dictionary from observed channel realizations.

However, as it is computationally prohibitive to find the optimal k -sparse approximation of \mathbf{h} , we need to replace (2) with an approximation

$$\varepsilon(\mathbf{D}, \mathbf{h}) = \|\mathbf{h} - \mathbf{D}\hat{\mathbf{x}}(\mathbf{h}, \mathbf{D})\|_2^2 \quad (4)$$

where $\hat{\mathbf{x}}(\mathbf{h}, \mathbf{D})$ are the coefficients of a suboptimal k -sparse approximation of \mathbf{h} in the dictionary \mathbf{D} as found, e.g., by the OMP algorithm or the IHT algorithm [8], [9].

The sparsity constraint $\|\mathbf{x}\|_0 \leq k$ can also be written as a union-of-subspaces constraint $\mathbf{x} \in \Sigma_k$, where Σ_k are all k -dimensional subspaces of \mathbb{C}^M spanned by k canonical basis vectors. The vector \mathbf{h} is then approximated in the union of subspaces given as the image of Σ_k under \mathbf{D} . Our second question is whether there is another union-of-subspaces model Γ_k , for which the approximation error $\varepsilon_0(\Gamma_k, \mathbf{h})$ is small (in the mean). In Sec. IV, we show how to obtain a different union-of-subspaces model if we restrict the allowed combinations of dictionary elements, for example, by allowing only adjacent steering vectors.

III. ALTERNATIVE DICTIONARIES

In this section, we present two approaches by which alternative dictionaries are obtained. First, we show how the KLE of the channel vector can be calculated for a fixed cluster center δ and a small standard deviation of the sub-paths σ . We obtain an alternative dictionary by keeping only the M_2 strongest components and repeating this process for M_1 different cluster centers so that the total dictionary is composed of $M_1 M_2$ elements. Second, we show how dictionary learning

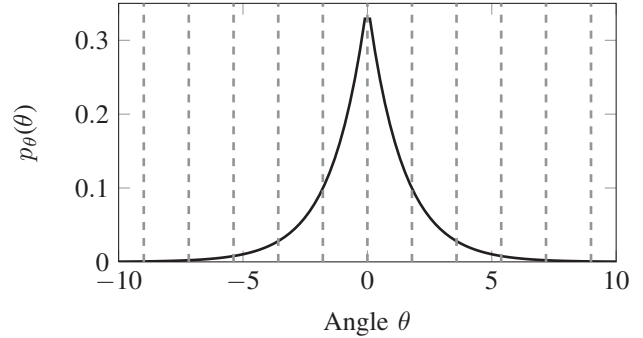


Figure 1. Probability density function of a Laplace random variable with 2° standard deviation. The dashed lines show the grid points of an orthogonal DFT matrix with 64 columns

can be used to find a dictionary that is (sub-)optimally adapted to observed channel vectors.

A. Dictionary derived from the Karhunen-Loève expansion

Let θ be a Laplace random variable with mean 0° and standard deviation σ with a probability density function p_θ as shown in Fig. 1 (for $\sigma = 2^\circ$). Such a random variable describes the angular distribution of the sub-paths belonging to the same cluster. We calculate the KLE of a vector \mathbf{h} distributed as in (1). Let $y_\theta = \exp(i\pi\varphi)/\sqrt{P}$ be a unit-modulus complex number with uniformly distributed phase $\varphi \sim \mathcal{U}_{[-1,1]}$. Clearly, $\mathbb{E}[y_\theta] = 0$, $\mathbb{E}[|y_\theta|^2] = P^{-1}$, and $\mathbb{E}[y_\theta y_{\theta'}^*] = 0$ for independent variables y_θ and $y_{\theta'}$. It follows that $\mathbb{E}[\mathbf{h}] = 0$ and the covariance matrix is given by

$$\begin{aligned} \text{Cov}[\mathbf{h}] &= \mathbb{E}[\mathbf{h}\mathbf{h}^H] = \sum_{\theta, \theta' \in \Theta} \mathbb{E}[y_\theta y_{\theta'}^*] \mathbb{E}[\mathbf{a}(\theta)\mathbf{a}(\theta')^H] \\ &= P^{-1} \sum_{\theta \in \Theta} \mathbb{E}[\mathbf{a}(\theta)\mathbf{a}(\theta)^H] \\ &= \mathbb{E}[\mathbf{a}(\theta)\mathbf{a}(\theta)^H]. \end{aligned}$$

We calculate the element on the n th off-diagonal of this matrix analytically: Let $b = \sigma[\text{rad}]/\sqrt{2}$ denote the scale parameter of the Laplace distribution of θ . We obtain

$$\begin{aligned} \mathbb{E}[\mathbf{a}(\theta)\mathbf{a}(\theta)^H]_{m, m+n} &= \frac{1}{N} \int p_\theta(\theta) \exp(-i\pi n \sin \theta) d\theta \\ &= \frac{1}{2Nb} \int \exp(-|\theta|/b - i\pi n \sin \theta) d\theta \\ &\approx \frac{1}{2Nb} \int \exp(-|\theta|/b - i\pi n \theta) d\theta \\ &= (N(1 + (bn\pi)^2))^{-1} \end{aligned}$$

where we used $\sin \theta \approx \theta$ for small θ . This approximation is very accurate for small σ , e.g., $\sigma = 2^\circ$, because $\exp(-|\theta|/b)$ is rapidly decreasing, see Fig. 1.

The cumulative sum of ordered eigenvalues of the covariance matrix of \mathbf{h} for $N = 64$ antennas is shown in Fig. 2. It can be seen that a random channel vector \mathbf{h} can be well approximated by a low-dimensional subspace. By taking the M_2 eigenvectors corresponding to the strongest eigenvalues of the covariance

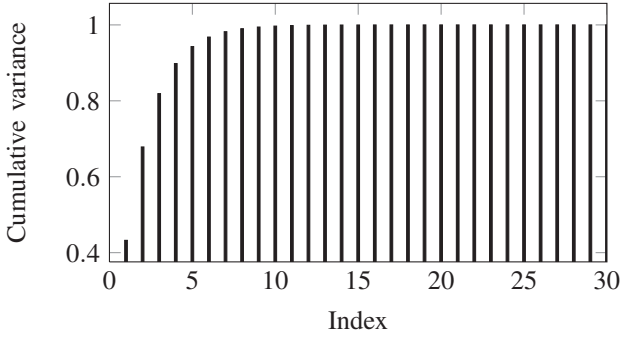


Figure 2. Cumulative variance for the 30 largest principal components

Algorithm 1 Generation of SCM dictionary

- 1) **Input:** Standard deviation σ , Number of cluster centers M_1 , Dimension of approximating subspace M_2
 - 2) Calculate eigendecomposition from covariance matrix with entries $(N(1 + (bn\pi)^2))^{-1}$ at n th off-diagonal, where $b = \sigma[\text{rad}]/\sqrt{2}$
 - 3) Let \mathbf{D}_0 be the matrix composed of eigenvectors corresponding to the M_2 largest eigenvalues
 - 4) Set $d_m = -1 + (2m - 1)/M_1$ (uniform grid in $[-1, 1]$)
 - 5) Set $\mathbf{D}_m = \text{diag}(\mathbf{a}(\text{asin}(d_m))) \cdot \mathbf{D}_0$ (rotate \mathbf{D}_0 to center d_m)
 - 6) Set $\mathbf{D} = [\mathbf{D}_1 \ \mathbf{D}_2 \ \dots \ \mathbf{D}_{M_1}] \in \mathbb{C}^{N \times M_1 M_2}$
-

matrix, we obtain dictionary vectors for a cluster of scatterers centered at 0° and with standard deviation σ .

We obtain the complete dictionary by rotating this principal subspace, which is centered at 0° , to a total of M_1 grid points. For reasons of symmetry we choose equi-spaced grid points between -1 and 1 in the sine-space of the angle. The algorithm is described in Alg. 1. We refer to the matrix \mathbf{D} , which is the output of the algorithm, as the SCM-dictionary, i.e., the dictionary designed according to the SCM.

B. Learning the best dictionary

Both the dictionary composed of steering vectors and that generated from basis vectors of the KLE of the channel vector rely on the geometric channel model and this could result in a performance degradation if the true channel does not follow the model. A dictionary that is learned from a large number of actual channel realizations, however, does not suffer such a drawback. We thus briefly describe the K-SVD algorithm [10], which can be used to find a dictionary that is (sub-)optimally adapted to some observed channel realizations.

Let $\mathbf{h}_i \in \mathbb{C}^N$ for $i = 1, \dots, T$ denote a number of observed channel realizations, which are generated according to some unknown model, e.g., actual measurements. Let k denote the desired approximation order. The dictionary learning problem is given as

$$\min_{\mathbf{X}, \mathbf{D}} \frac{1}{T} \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F^2 \quad \text{s.t. } \|\mathbf{x}_i\|_0 \leq k, \quad i = 1, \dots, T \quad (5)$$

where $\|\cdot\|_F$ denotes the Frobenius norm and where the columns of the matrix \mathbf{H} are realizations \mathbf{h}_i and where \mathbf{x}_i denotes the i th column of \mathbf{X} . This is the same optimization problem as before, but this time we also optimize with respect to the dictionary. Moreover, the expectation with respect to the unknown distribution of \mathbf{h} is replaced by the large sample average, i.e., $T^{-1} \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F^2 \approx \mathbb{E}_{\mathbf{h}} [\|\mathbf{h} - \mathbf{D}\mathbf{x}\|_2^2]$.

The optimization problem (5) is very hard to solve and, thus, tackled with an alternating optimization algorithm. First, we restrict the problem to dictionaries \mathbf{D} with normalized columns $\|\mathbf{d}_n\|_2 = 1$ and note that this scaling is absorbed by the \mathbf{X} matrix. Then, the variables \mathbf{D} and \mathbf{X} are updated in an alternating manner. If \mathbf{D} is fixed, a sparse coding \mathbf{x}_i for the i th column of \mathbf{H} is found by using an algorithm that approximately solves

$$\min_{\mathbf{x}} \|\mathbf{h}_i - \mathbf{D}\mathbf{x}\|_2^2 \quad \text{s.t. } \|\mathbf{x}\|_0 \leq k$$

e.g., the OMP algorithm [9]. Then, the columns of \mathbf{D} are updated one after another while only performing minor updates to \mathbf{X} . Let $\mathbf{Z} = \mathbf{X}^T$ so that the n th column \mathbf{z}_n of \mathbf{Z} is the n th row of \mathbf{X} . We can write $\mathbf{H} - \mathbf{D}\mathbf{X}$ as

$$\mathbf{H} - \mathbf{D}\mathbf{X} = \mathbf{H} - \sum_{n \neq j} \mathbf{d}_n \mathbf{z}_n^T - \mathbf{d}_j \mathbf{z}_j^T = \mathbf{E}_j - \mathbf{d}_j \mathbf{z}_j^T$$

where the matrix $\mathbf{E}_j = \mathbf{H} - \sum_{n \neq j} \mathbf{d}_n \mathbf{z}_n^T$ does not depend on \mathbf{d}_j . Let $\mathbf{E}_j = \sum_q \sigma_q \mathbf{u}_q \mathbf{v}_q^H$ denote the singular value decomposition of \mathbf{E}_j with $\sigma_q \geq 0$ and \mathbf{u}_q and \mathbf{v}_q the left and right singular vectors, respectively. Then, the minimization of (5) with respect to the j th column of \mathbf{D} can be written as

$$\min_{\mathbf{d}_j} \frac{1}{T} \left\| \sum_q \sigma_q \mathbf{u}_q \mathbf{v}_q^H - \mathbf{d}_j \mathbf{z}_j^T \right\|_F^2.$$

and is solved by selecting $\mathbf{d}_j = \mathbf{u}_1$ and $\mathbf{z}_j^T = \sigma_1 \mathbf{v}_1^H$, where σ_1 is the maximal singular value. Thus, the j th column of the matrix \mathbf{D} is updated by \mathbf{u}_1 and the j th row of the matrix \mathbf{X} is updated by $\sigma_1 \mathbf{v}_1^H$. The algorithm can be initialized with an oversampled DFT matrix, $\mathbf{D}_0 = \mathbf{D}_{\text{DFT}}$, or with the dictionary \mathbf{D}_{SCM} found with Alg. 1.

IV. ALTERNATIVE UNION-OF-SUBSPACES MODELS

Thus far, we have argued that channel vectors \mathbf{h} can be approximated in the image under \mathbf{D} of Σ_k . However, if the channel \mathbf{h} is really generated from a single cluster of paths, we can also try to approximate \mathbf{h} , for example, with k adjacent steering vectors. Or, if we are working with the SCM dictionary, we could restrict the search for an optimal approximation to linear combinations of dictionary vectors that all correspond to the same cluster center.

Thus, let $\Phi_b \subset \mathbb{C}^N$ denote a k -dimensional subspace for $b = 1, \dots, B$ and $\Gamma_k = \cup_{b=1, \dots, B} \Phi_b$ be the union of these subspaces. We can then approximate \mathbf{h} in Γ_k and evaluate the optimal approximation error

$$\varepsilon_0(\Gamma_k, \mathbf{h}) = \min_{\hat{\mathbf{h}} \in \Gamma_k} \|\mathbf{h} - \hat{\mathbf{h}}\|_2^2.$$

If we choose Φ_b as the image under \mathbf{D} of k basis vectors $\mathbf{e}_i \in \mathbb{C}^N$, where i varies in an index set I_b , and if we

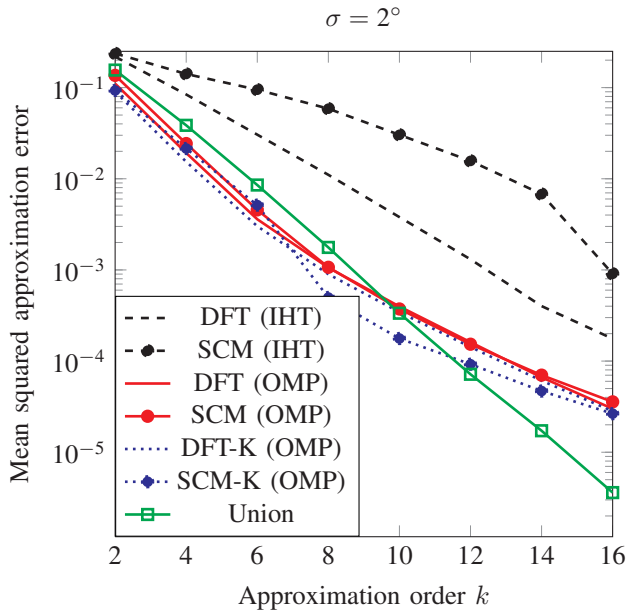


Figure 3. Mean squared approximation errors for different orders k of the various algorithms for channel vectors with $\sigma = 2^\circ$

take the union over all possible index sets I_b of size k , we recover the original approximation problem. This set consists of $\binom{N}{k}$ distinct subspaces, which is why this problem is hard to solve. However, if we reduce the allowed combinations of dictionary vectors, we can reduce the total number of subspaces. This is desirable for several reasons. First, the complexity of the optimization problem is reduced significantly and it may become possible to find an optimal solution instead of a suboptimal one. Second, by imposing more structure on the problem, i.e., by allowing less subspaces in which an approximation is searched, we can achieve a de-noising effect by approximating \mathbf{h} in Γ_k .

We use the k strongest vectors $\mathbf{u}_1(\delta), \dots, \mathbf{u}_k(\delta)$ of the KLE of \mathbf{h} for a given cluster center δ and let δ vary on a grid. But we do not allow combinations of dictionary elements that correspond to different cluster centers. We obtain a union-of-subspaces model

$$\Gamma_k = \cup_{\delta \in \text{grid}} \text{span}\{\mathbf{u}_1(\delta), \dots, \mathbf{u}_k(\delta)\}$$

which is adapted to the KLE of \mathbf{h} . In this model, the number of subspaces in Γ_k corresponds to the number B of grid points for δ , which may be much smaller than $\binom{N}{k}$.

Finding the best approximating subspace, i.e., the optimal δ , is done by correlating \mathbf{h} with all subspaces: Let \mathbf{P}_b denote the orthogonal projector onto $\text{span}\{\mathbf{u}_1(\delta_b), \dots, \mathbf{u}_k(\delta_b)\}$. Then, the optimal approximation error is given by

$$\varepsilon_0(\Gamma, \mathbf{h}) = \min_{b=1, \dots, B} \|\mathbf{h} - \mathbf{P}_b \mathbf{h}\|_2^2.$$

This is completely analogous to finding a direction of arrival by beamforming, i.e., calculating the projection of \mathbf{h} onto the steering vector $\mathbf{a}(\theta)$ for all angles θ on a grid.

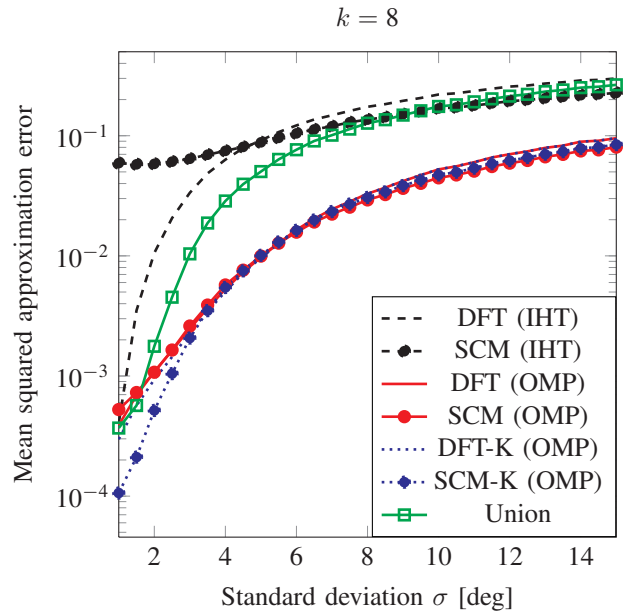


Figure 4. Mean squared approximation errors of the various algorithms based on channels with varying σ and where $k = 8$ was used as the approximation order

This approximation method can be generalized to multiple cluster centers with the fusion frame formalism developed in [11].

V. SIMULATION RESULTS

We use Monte Carlo estimates to calculate the quantities $\mathbb{E}_{\mathbf{h}}[\varepsilon(\mathbf{D}, \mathbf{h})]$ and $\mathbb{E}_{\mathbf{h}}[\varepsilon_0(\Gamma_k, \mathbf{h})]$ for the different dictionaries \mathbf{D} presented in Sec. III and the union-of-subspaces model Γ_k presented in Sec. IV. The antenna array used is always a ULA with $N = 64$ antennas. All curves are based on 10 000 independent realizations for the channel vectors \mathbf{h} for any given combination of parameters.

The dictionary of steering vectors (which we refer to as DFT in the plot legends) is an eight times oversampled DFT matrix. This corresponds to a uniform grid for $\sin \theta$ with a spacing of $2/(8N)$ between consecutive grid points. The dictionary derived from the KLE of the spatial channel model (which we refer to as SCM) is constructed according to Alg. 1 with $M_1 = N$, $M_2 = 8$ and $\sigma = 2^\circ$. Thus, both dictionaries are of the same size.

We then use both dictionaries to initialize the dictionary learning algorithm. Our implementation closely follows the description in [10], i.e., in the update of the j th dictionary element, we only use columns of the matrix $\mathbf{H} - \mathbf{D}\mathbf{X}$ that correspond to training data that actually use the j th element. If this element is unused, it is replaced by the normalized training signal \mathbf{h}_i that is currently least well approximated. We run the algorithm for ten outer iterations (the inner iterations are the sequential updates of the columns of the dictionary) and with training data composed of 100 000 independent realizations of the channel vector \mathbf{h} according to the SCM as in (1) with $\sigma = 2^\circ$, $P = 20$, and δ uniformly distributed on $[-90^\circ, 90^\circ]$.

Once every outer iteration, we use the OMP algorithm to compute a new sparse coding matrix \mathbf{X} . We obtain different K-SVD dictionaries for each approximation order k used by the OMP algorithm. Thus, the K-SVD dictionaries are tailored to the approximation order k . The resulting dictionaries are referred to as DFT-K and SCM-K.

The grid used for the cluster centers in the union-of-subspaces model is the same as that used to construct the eight times oversampled DFT matrix, i.e., the grid for $\sin \delta$ is given as $\{-1, -1 + 2/(8N), \dots, 1 - 2/(8N)\}$. Then, as for the SCM dictionary, we use the k principal components of the KLE for a cluster centered at 0° as the subspace for $\delta = 0^\circ$ and rotate this subspace to all other grid points. We thus obtain $8N$ distinct k -dimensional subspaces, in which we can calculate the optimal approximation (this method is referred to as Union).

We use two different algorithms, OMP and IHT, that find a suboptimal approximation of \mathbf{h} in the respective dictionaries. The IHT algorithm is run with a step size parameter $\mu = 0.1$ and for 100 iterations without any convergence criterion. We then perform a final least squares step on the final index set.

In our first experiment, we compare the rate of decay of the different approximation errors as we increase the approximation order k . In our second experiment, we test the robustness of the different dictionaries, which are all designed for an SCM with a standard deviation $\sigma = 2^\circ$, with respect to different σ of the SCM used to generate the channel vectors \mathbf{h} .

From the results, which are shown in Figures 3 and 4, we see that between the eight times oversampled DFT dictionary and the corresponding SCM dictionary, there is not much of a difference regarding the approximation error if the OMP algorithm is used. In contrast, the IHT algorithm is more sensitive to the choice of dictionary and appears to yield better results with the DFT dictionary. The results for the union-of-subspaces model are interesting. The set of subspaces in which the channels are approximated is considerably smaller than those used by the OMP and IHT algorithms. Yet, if the true model coincides with the model used to calculate the KLE (as in Fig. 3), the approximation is almost as good as that found by OMP for small orders k and even better for large orders k . However, as can be seen from Fig. 4, the performance of this approach degrades more quickly as the model used to generate the channels deviates from that used to generate the KLE.

VI. CONCLUSION

We conclude from our simulation results that the dictionary of steering vectors is a good choice for approximating vectors that are generated according to the 3GPP spatial channel model with a single cluster. However, given that the learned dictionaries perform about as good as the DFT dictionary and in view of the uncertainty about whether the 3GPP SCM accurately models real communication channels, the K-SVD dictionary might result to be a good practical choice. Compared to the DFT dictionary, which is well understood and widely adopted, the SCM dictionary has the drawback of being dependent on the channel model, which could be a problem in practice if the true model does not coincide with the assumed model.

REFERENCES

- [1] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.
- [2] T. S. Rappaport, Y. Qiao, J. I. Tamir, J. N. Murdock, and E. Ben-Dor, "Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems (invited paper)," in *IEEE Radio and Wireless Symposium (RWS)*, Jan. 2012, pp. 151–154.
- [3] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
- [4] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations (release 12)," 3rd Generation Partnership Project (3GPP), TR 25.996 V12.0.0, 2014.
- [5] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [6] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
- [7] D. C. Araujo, A. L. F. de Almeida, J. A. Xnas, and J. C. M. Mota, "Channel estimation for millimeter-wave very-large MIMO systems," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, Sep. 2014, pp. 81–85.
- [8] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, Dec. 2008.
- [9] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [11] P. Boufounos, G. Kutyniok, and H. Rauhut, "Sparse recovery from combined fusion frame measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3864–3876, Jun. 2011.