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 estimating channel vectors that are superpositions of many closely spaced steering vectors. Such vectors describe single clusters of scatterers and appear in realistic channel models. We question the practice of approximating such vectors as superpositions of few steering vectors in the context of sparse representations. We investigate if, instead of using such oversampled DFT dictionaries in recovery algorithms, performance gains are possible by using other dictionaries that are more adapted to channel vectors.**

***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 micros cells in the 3GPP model [4]. On the other hand, models used for *estimating* channels, use superpositions of 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 in 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 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 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 (SCM). 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

Let

$$a(\theta) = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \exp[i\pi \sin\theta] & \cdots & \exp[i\pi(N-1)\sin\theta] \end{pmatrix}^T$$

denote the normalized steering vector, i.e., the signal recorded by an 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

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

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 SCM model [4]. The distribution of the parameters δ, σ, and P depends on the chosen scenario, e.g., urban macro cell ($P = 20, \delta \sim \mathcal{U}[-40°, 40°]$, $\sigma = 5°$), urban microcell ($P = 20, \delta \sim \mathcal{U}[-40°, 40°]$, $\sigma = 2°$), etc.

A dictionary is a fat matrix $D \in \mathbb{C}^{N\times M}$ of unit-norm vectors, e.g., an oversampled DFT matrix. For a given dictionary, we are interested in evaluating

$$E\left[\min_x \|h - Dx\|_2^2 \quad \text{s.t. } \|x\|_0 \leq k\right] \tag{2}$$

where the expectation is with respect to the random variable h, i.e., the set Θ and the coefficients y_θ. By $\|x\|_0$ we denote the cardinality of the vector x. By solving the optimization problem within the expectation operator, we obtain the best k-sparse representation of h in the given dictionary D and by taking the expectation we obtain the mean approximation error of the channel vector for the best k-sparse approximation in the given dictionary.

Our first question is whether for a given parametrization of the distribution of the channel vector h and a given sparsity parameter k, there is a dictionary D other than an oversampled DFT matrix, which has a smaller mean approximation error. However, as it is computationally prohibitive to find the optimal k-sparse approximation of h, we need to replace (2) with an approximation

$$E\left[\|h - Dx(h, D)\|_2^2\right] \tag{3}$$

where $x(h, D)$ are the coefficients of a suboptimal k-sparse approximation of h in the dictionary D as found, e.g., by the OMP algorithm, the IHT algorithm, or ℓ_1-minimization [8]–[10].

The sparsity constraint $\|x\|_0 \leq k$ can also be written as a union-of-subspaces constraint $x \in \Sigma_k$, where Σ_k are all k-dimensional subspaces of $\mathbb{C}^M$ spanned by k canonical basis vectors. The vector h is then approximated in the union of subspaces given as the image of Σ_k under D. Our second question is whether there is another union-of-subspaces model Γ_k, for which the expected approximation error

$$E\left[\min_{\hat{h}} \|h - \hat{h}\|_2^2 \quad \text{s.t. } \hat{h} \in \Gamma_k\right]$$

is small. An example is the fusion frame formalism [11], where $\hat{h}$ is assumed to admit an approximation by k/k_0 vectors, each of which lies in a k_0-dimensional subspace.

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 σ. We obtain an alternative dictionary by keeping only the M_2 strongest components and repeating this process for M_1 different cluster centers. Second, we show how dictionary learning can be used to find a dictionary that is (sub-)optimally adapted to observed channel vectors.

A. Dictionary derived from Karhunen-Loève expansion

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

$$\begin{aligned}\mathrm{Var}[h] = E[hh^H] &= \sum_{\theta,\theta' \in \Theta} E[y_\theta y_{\theta'}^*] E[a(\theta)a(\theta')^H] \\ &= \frac{1}{P}\sum_{\theta\in\Theta} E[a(\theta)a(\theta)^H] \\ &= E[a(\theta)a(\theta)^H]\,.\end{aligned}$$

We calculate the element on the nth 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

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

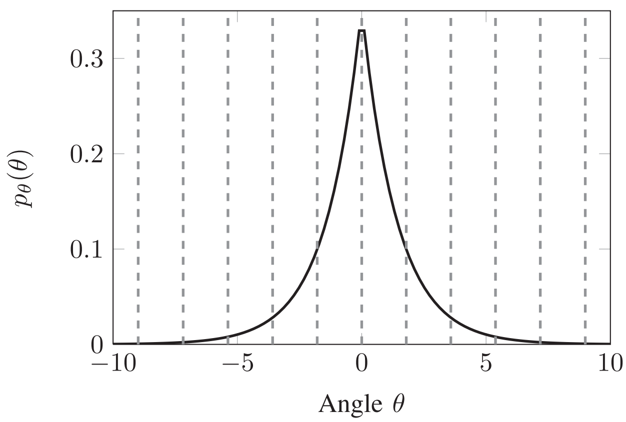

Fig. 1. Probability density function of a Laplace random variable with 2° standard deviation. The dashed lines show the grid points of an orthogonal DFT matrix with 64 columns

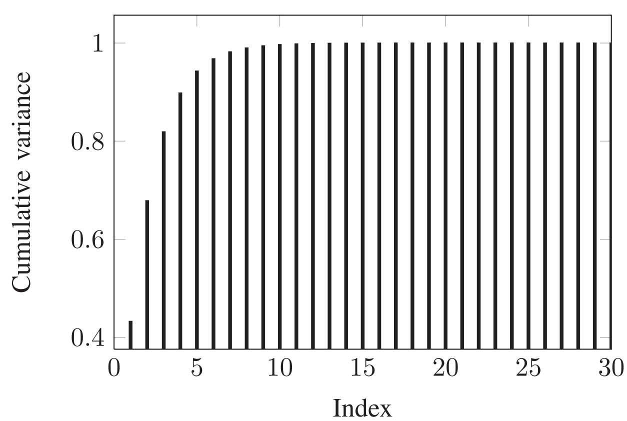

Fig. 2. Cumulative variance for the 30 largest principal components

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 nth off-diagonal, where $b = \sigma[\text{rad}]/\sqrt{2}$
3) Let 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 $D_m = \text{diag}(a(\text{asin}(d_m))) \cdot D_0$ (rotate D_0 to center d_m)
6) Set $D = [D_1\ D_2\ \ldots\ D_{M_1}]$

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

The cumulative sum of ordered eigenvalues of the covariance matrix of h for $N = 64$ antennas is shown in Fig. 2. It can be seen that a random channel vector h can be well approximated by a low-dimensional subspace. By taking the, e.g., eigenvectors corresponding to the eight strongest eigenvalues of the covariance matrix, we obtain vectors for a dictionary corresponding to 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 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

In this section, we plan to describe the K-SVD algorithm [12], which finds a suboptimal solution to the problem

$$\min_{X,D} \|H - DX\|_F^2 \quad \text{s.t. } \|x_i\|_0 \leq k\,,$$

where the columns of the matrix H are realizations of channel vectors according to the SCM and where x_i denotes the $i$the column of X. Here, the dictionary D is subject to optimization and the goal is to find a dictionary that is capable of concisely describing the observed data H. A suboptimal solution is found with an alternating optimization algorithm. The algorithm can be initialized with an oversampled DFT matrix, $D_0 = D_{\text{DFT}}$.

IV. Alternative Subspaces

In this section, we plan to describe alternative low-rank descriptions of the channel vector. An example is given by the fusion frame formalism [11].

V. Simulation Results

We provide Monte Carlo estimates of the mean approximation error (3) for the different methods of obtaining a low-rank approximation of channel vectors h that are generated according to the SCM. In this extended abstract, the results are limited to a comparison between the SCM dictionary and the DFT dictionary and for the OMP and IHT algorithms. We plan to add simulations for learned dictionaries as well as for other sparse approximation algorithms, e.g., ℓ_1-minimization. Moreover, we plan to add results obtained from using different forms of low-rank approximations, e.g., the fusion frame formalism, where the channel vector is approximated by few low-dimensional subspaces (instead of k one-dimensional subspaces).

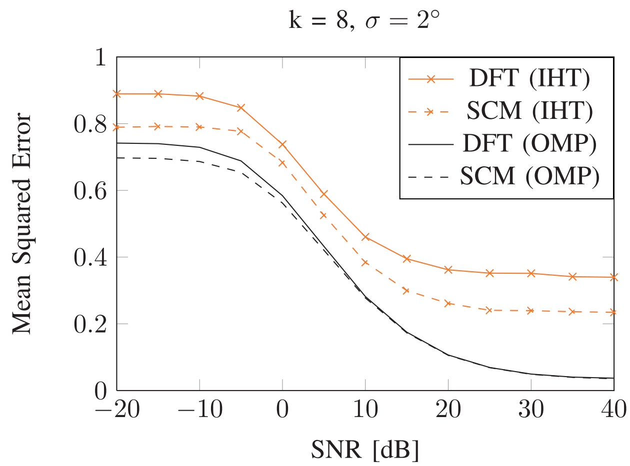

Fig. 3. Mean approximation error of order $k = 8$ for various SNR values and based on channels with 2° angular standard deviation

Let the number of antennas be $N = 64$ and let the vector h be generated according to

$$h' = \sum_{\theta\in\Theta} y_\theta a(\theta) + v$$
$$h = \frac{h'}{\|h'\|}$$

where we added noise $v \sim \mathcal{N}_{\mathbb{C}}(0, \text{SNR}^{-1}I)$ to test the robustness of the methods from deviations of the ideal Laplace model. The set Θ is obtained by first drawing the cluster center δ from a uniform distribution on $[-40°, 40°]$ and then drawing $P = 20$ angles from the Laplace distribution centered at δ and with standard deviation σ. The coefficients y_θ are obtained by drawing a phase $\varphi \sim \mathcal{U}_{[-1,1]}$ and setting $y_\theta = \exp[i\pi\varphi]/\sqrt{P}$ for each θ. The SCM dictionary D is obtained by running Alg. 1 with $\sigma = 2°, M_1 = N, M_2 = 8$, i.e., the dictionary consists of $M_1 M_2 = 512$ elements and is tailored to a standard deviation of 2°. The DFT dictionary is eight times oversampled, i.e., it consists of 512 steering vectors on the grid $\{-1, -1+2/512, -1+4/512, \ldots, 1-2/512\}$, which is uniform in the sine-space of the angle. In particular, both dictionaries have the same size.

In this extended abstract, we provide simulation results for the OMP and IHT algorithms as suboptimal estimators in (3) and both are used with either of the dictionaries. We calculate Monte Carlo estimates of the quantity

$$E\left[\|h - Dx(h, D)\|^2\right]\,,$$

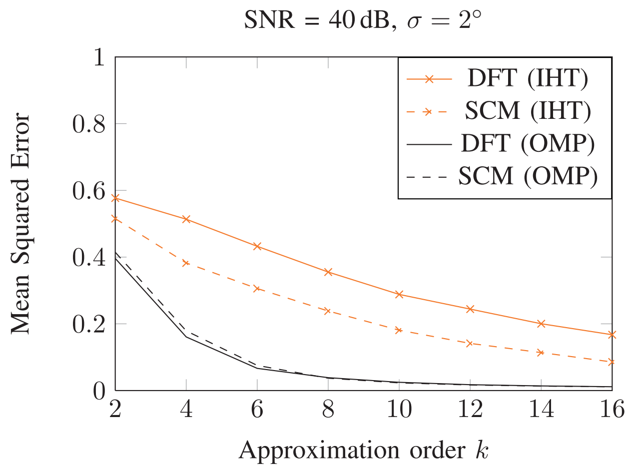

Fig. 4. Mean approximation error for various orders k at SNR=40 dB and based on channels with 2° angular standard deviation

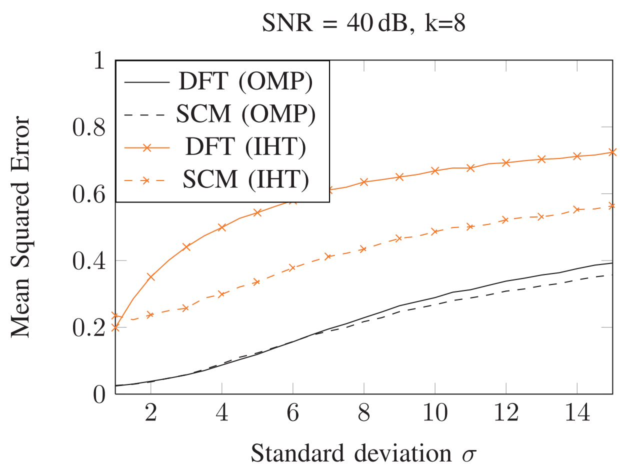

Fig. 5. Mean approximation error of order $k = 8$ at SNR=40 dB and for channels with varying angular standard deviation σ

i.e., the error between the channel and the estimate. The curves in Figs. 3–5 are based on 1000 realizations of channel vectors for various SNR values, approximation orders k, and also for different standard deviations σ to test the performance and robustness of the various choices of dictionaries.

The results show that both dictionaries show similar performance if the OMP algorithm is used, i.e., the dictionary of steering vectors seems to be a good choice for the SCM used here. In contrast, the IHT algorithm (we use a standard, non-normalized version) is sensitive to the choice of dictionary and shows better performance with the SCM dictionary.

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. Axnas, 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] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[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.

[12] 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.