# Riemannian Coding for Covariance Interpolation in Massive MIMO Frequency Division Duplex Systems


Alexis Decurninge, Maxime Guillaud
Huawei Technologies, Co. Ltd.
France Research Center
Mathematical and Algorithmic Sciences Lab
{alexis.decurninge,maxime.guillaud}@huawei.com

Dirk Slock
EURECOM
Mobile Communications Department
dirk.slock@eurecom.fr



***Abstract*—In the context of multi-user Massive MIMO frequency division duplex (FDD) systems, the acquisition of channel state information cannot benefit from channel reciprocity. However, it is generally expected that covariance information about the downlink channel must be estimated and fed back by the user equipment (UE). As an alternative, it was also proposed to infer the downlink covariance based on the observed uplink covariance and a stored dictionary of uplink/downlink covariance matrices. This inference was performed through an interpolation in the Riemannian space of Hermitian positive definite matrices. We propose to rewrite the interpolation step as a Riemannian coding problematic. In this framework, we estimate the decomposition of the observed uplink matrix in the dictionary of uplink matrices and recover the corresponding downlink matrix assuming that its decomposition in the dictionary of downlink matrices is the same. Moreover, since this space is of large dimension in the Massive MIMO setting, it is expected that these decompositions will be sparse. We then propose new criteria based on this further constraint.**


## I. Introduction

Accurate and up-to-date channel state information (CSI) is a critical parameter in the operation of Massive MIMO [1], [2]. However, for FDD systems, reciprocity does not hold between uplink and downlink channels. Obtaining CSI then requires over-the-air feedback. In order to lower the amount of needed feedback in the Massive MIMO setting, several authors have proposed to use channel covariance knowledge [3]–[5]. Exploiting relations between uplink and downlink covariances was considered e.g. in [6] and [7] but these methods rely on specific geometric assumptions. It was proposed in [8] to deduce a new downlink covariance from an estimated uplink covariance and a dictionary of pairs of known uplink/downlink covariance matrices that has been stored in a first training phase. This dictionary exploitation was performed on the space of $N \times N$ Hermitian positive definite matrices $\mathcal{S}_N^+$ via various interpolation criteria.

The exposed problematic shares similarities with coding techniques used for classification popular in the machine learning literature. The coding formulation is as follows: given a dictionary of codewords, a query is associated to one or multiple dictionary elements with different weights. These weights, or codes, act as the new representation for the query and serve as input to a classifier. Several coding methods have been proposed in the literature in both Euclidean [9] and Riemannian [10]–[12] settings.
In the covariance interpolation setting, the new query will consist in a new uplink covariance matrix. We will then use the codes computed for the uplink matrix to deduce the downlink covariance matrix by using a "decoding" process. Thus, by doing this, we suppose that the codes of new uplink/downlink covariance matrices are the same in their respective dictionaries of uplink and downlink matrices.

Moreover, in the massive MIMO setting, $N$ is supposed to be large. The considered space of covariance matrices $\mathcal{S}_N^+$ will then be high-dimensional. Since the considered covariance matrices represent real physical scenes and lie in this high-dimensional space, it is expected that they could be decomposed in a sparse basis. This assumption could be explained through two contradictory insights:

- The mapping relating an uplink covariance matrix to its corresponding downlink covariance can reasonably be assumed to be continuous, then it is natural to consider only the nearest elements of any new uplink matrix to locally interpolate the corresponding downlink covariance. Note that this insight does not rely on the space being high-dimensional.
- On the other hand, we may consider that the covariance matrix is the superposition of a small number of phenomena which are different by nature and the sparse decomposition would aim to recover this small number of features.

Some authors, taking into account the first point, add neighborhood constraints [12]. However, we will not have such constraints, favoring a more agnostic approach. Furthermore, we assume that $K$ satisfies $K \leq N^2$, i.e. the size of the dictionary is smaller than the dimension of the considered space. This constraint comes from storage constraints: the base station (BS) will be able to store a moderate number of large covariance matrices.

The aim of this paper is to introduce interpolation criteria in the spirit of Riemannian coding adapted for interpolation purposes and study these criteria through their sparsity and performance properties.

The paper is organized as follows. Section II briefly recalls the model introducing uplink and downlink covariance matrices. Section III is devoted to the interpolation criteria. Finally, we will present an illustrative simulated scenario in Section IV and compare the performances of the proposed criteria.

## II. System Description and Model

For simplicity, we assume in the following single-antenna UEs. Consider that during the downlink transmission phase, a BS uses $N_T$ antennas to transmit, while during the uplink phase, the BS uses $N_R$ antennas to receive signals (when no confusion is possible, we will use the notation $N$ to indifferently designate $N_T$ or $N_R$). Let us further assume that the coefficients of the downlink (respectively uplink) channel $\mathbf{h}$ (resp. $\mathbf{h}^{\mathrm{UL}}$) are correlated, and that they can be written as (see e.g. [4])

$$\mathbf{h} = \mathbf{R}^{1/2}\mathbf{w} \quad \text{and} \quad \mathbf{h}^{\mathrm{UL}} = \left(\mathbf{R}^{\mathrm{UL}}\right)^{1/2}\mathbf{w}^{\mathrm{UL}} \tag{1}$$

where

- $\mathbf{R}$ is the $N_T \times N_T$ BS-side covariance matrix of the downlink channel and $\mathbf{R}^{\mathrm{UL}}$ is the $N_R \times N_R$ uplink covariance matrix.
- $\mathbf{w}$ and $\mathbf{w}^{\mathrm{UL}}$ respectively are an $N_T$-dimensional vector and an $N_R$-dimensional vector with independent, unit variance coefficients capturing the fast fading.

If $\mathbf{R}$ is known and rank limited (equal to $N_{RL}$), it is shown in [4], [5], [13] that the length of the training sequence can be reduced from $N_T$ to $N_{RL}$, thereby reducing the amount of pilot symbols required to estimate $\mathbf{h}$. The knowledge of $\mathbf{R}$ is therefore crucial in the Massive MIMO setting. However, the BS has only access to $\mathbf{R}^{\mathrm{UL}}$.

Let us assume that during a training phase, $K$ pairs of downlink/uplink covariance matrices $(\mathbf{R}_1, \mathbf{R}_1^{\mathrm{UL}}), \dots, (\mathbf{R}_K, \mathbf{R}_K^{\mathrm{UL}})$ have been collected (see [8]). The problem is then to estimate a new $\mathbf{R}$ from an observed $\mathbf{R}^{\mathrm{UL}}$ on the basis of this dictionary.

## III. Downlink Covariance Estimation Through Interpolation

We will consider in the following a general Riemannian distance $\mathrm{d}(\cdot,\cdot)$ in $\mathcal{S}_N^+$. Let us mention for example the Affine Invariant distance $\mathrm{d}_{AI}(\mathbf{X},\mathbf{Y}) = \|\log(\mathbf{X}^{\frac{1}{2}}\mathbf{Y}^{-1}\mathbf{X}^{\frac{1}{2}})\|_{\mathrm{F}}$ and the Log-Euclidean $\mathrm{d}_{LE}(\mathbf{X},\mathbf{Y}) = \|\log(\mathbf{X}) - \log(\mathbf{Y})\|_{\mathrm{F}}$ where $\exp$ and $\log$ refer to matrix exponential and logarithm. In the following, the $\exp_{\mathbf{X}}$ and $\log_{\mathbf{X}}$ operators refer to the Riemannian exponential and logarithm maps (applied to a certain point $\mathbf{X} \in \mathcal{S}_N^+$) corresponding to one of these distances (see e.g. [14] for further details).

The coding/decoding method evoked for covariance interpolation in Section I can be seen as a "mirror" approach:

- **Coding**: "decompose" the new uplink matrix in the basis composed by the dictionary of stored uplink covariance matrices through weights computed for each codeword.
- **Decoding**: compute the interpolated downlink covariance matrix by doing the inverse operation with the associated downlink codewords.

We then assume that the decomposition of a pair of uplink/downlink covariance matrices are the same in the two related dictionaries.

### A. Decoder as a weighted barycenter

Similar to [8], the interpolated downlink matrices will be interpreted as a weighted barycenter of the downlink covariance matrices in which weights (denoted by $\mathbf{w} = [w_1, ..., w_K]^T$) are computed according to the situation of $\mathbf{R}^{\mathrm{UL}}$ in the space of uplink covariance matrices:

$$\hat{\mathbf{R}} = \arg\min_{\mathbf{Y}\in\mathcal{S}_{N_T}^+(\mathbb{C})} \sum_{k=1}^{K} w_k d(\mathbf{R}_k, \mathbf{Y})^2 \tag{2}$$

We furthermore choose to constrain the weights to be such that $\sum_{k=1}^{K} w_k = 1$ in order to be compliant with the Euclidean framework. Note however that we do not constrain the weights to be nonnegative. The existence and uniqueness of such a minimum is still an open problem for an arbitrary distance $d$. Sander proves a partial result in [15] restraining the existence domain of the weights in a ball but he conjectures that the general result should be true since the sectional curvatures of the space of positive definite matrices are negative.

### B. Encoder: choice of the interpolation weights

The weights $\mathbf{w} = [w_1, \dots, w_K]^T$ are the "codes" of the observed uplink covariance matrix $\mathbf{R}^{\mathrm{UL}}$ in the basis constituted by the uplink dictionary elements.
In [8], we considered the following criterion:

$$\hat{\mathbf{w}} = \arg\min_{\substack{\sum_{k=1}^{K} w_k = 1 \\ w_k \geq 0}} \left\| \sum_{k=1}^{K} w_k \log_{\mathbf{R}^{\mathrm{UL}}}(\mathbf{R}_k^{\mathrm{UL}}) \right\|_{\mathrm{F}}. \tag{3}$$

This intrinsic criterion has been considered with different variants (considering other constraints or adding regularization terms) in [10], [12] to perform Riemannian coding. The choice of positive weights restrain the decoder to consider barycenters in the convex hull of the downlink covariance codewords. This constraint may seem arbitrary. A first proposition is then to allow to define negative weights considering as a consequence a form of extrapolation in the decoding process.

*1) A way to introduce extrapolation:* A criterion allowing a compromise between a kind of sparsity and extrapolation would be to consider a $\ell_1$ regularization like in [11]

$$\hat{\mathbf{w}} = \arg\min_{\substack{\sum_{k=1}^{K} w_k = 1 \\ \sum_{k=1}^{K} |w_k| \leq 1+\eta}} \left\| \sum_{k=1}^{K} w_k \log_{\mathbf{R}^{\mathrm{UL}}}(\mathbf{R}_k^{\mathrm{UL}}) \right\|_{\mathrm{F}}. \tag{4}$$

where $\eta$ is a real positive parameter. Let us explain the two chosen constraints. The first one, namely $\sum_{k=1}^{K} w_k = 1$ is there to avoid that $\hat{\mathbf{w}} = 0$ is a solution.
The second one permits a form of extrapolation while imposing a form of sparsity. If $\eta = 0$, the $\ell_1$ constraint in eq. (4) is equivalent to impose that every weight is positive reducing eq. (4) to be equal to eq. (3). Choosing a positive but small $\eta$ however means that we constrain the approximation to be $\eta$-close to the convex hull but not strictly inside, thereby achieving a compromise between weights sparsity and dictionary accuracy.

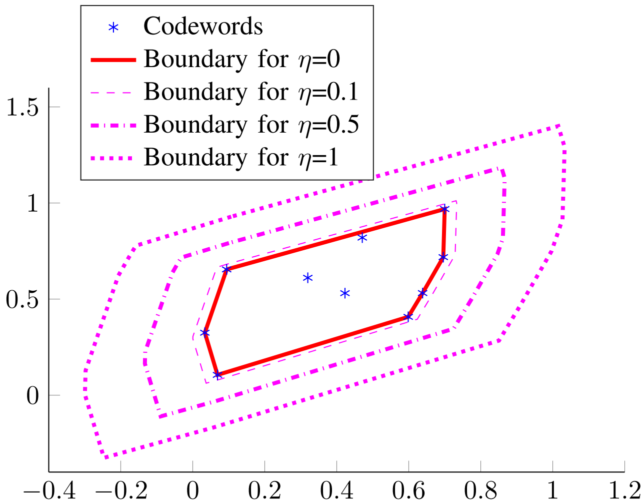


Fig. 1: Boundary of the feasible set from eq. (4) for a toy dictionary in an Euclidean framework of dimension 2

In order to give insights on the role of $\eta$, we may look at the Euclidean version of eq. (4), namely,

$$\hat{\mathbf{w}} = \arg \min_{\substack{\sum_{k=1}^K w_k = 1 \\ \sum_{k=1}^K |w_k| \leq 1+\eta}} \left\| \left( \sum_{k=1}^K w_k \mathbf{R}_k^{\mathrm{UL}} \right) - \mathbf{R}^{\mathrm{UL}} \right\|_{\mathrm{F}} .$$

We represent in Fig. 1 the boundaries of the set of "approximating elements" $\left(\sum_{k=1}^K w_k \mathbf{R}_k^{\mathrm{UL}}\right)$ in a space of dimension 2 with $\mathbf{w}$ satisfying $\sum_{k=1}^K w_k = 1$ and $\sum_{k=1}^K |w_k| \leq 1 + \eta$. From a numerical point of view, let $\mathbf{M}$ be the $N^2 \times K$ matrix defined by

$$\mathbf{M} = \left[\mathrm{vec}\left(\log_{\mathbf{R}^{\mathrm{UL}}}(\mathbf{R}_1^{\mathrm{UL}})\right), \ldots, \mathrm{vec}\left(\log_{\mathbf{R}^{\mathrm{UL}}}(\mathbf{R}_K^{\mathrm{UL}})\right)\right]$$

with vec the vectorization operator. Then, if we denote $\mathbf{A} = \mathrm{Re}(\mathbf{M}^H \mathbf{M})$, eq. (4) is equivalent to

$$\hat{\mathbf{w}} = \arg \min_{\substack{\sum_{k=1}^K w_k = 1 \\ \|\mathbf{w}\|_1 \leq 1+\eta}} \mathbf{w}^T \mathbf{A} \mathbf{w}. \tag{5}$$

Quadratic programming can then be used to solve (5). Since $K \leq N^2$, $\mathbf{A}$ is a full rank matrix with probability one and $\mathbf{w} \mapsto \mathbf{w}^T \mathbf{A} \mathbf{w}$ is then a strictly convex function. Thus, the minimization problem is well defined and $\hat{\mathbf{w}}$ is unique.

*2) Intrinsic sparsity criterion:* A way to introduce more sparsity in the choice of weights is to consider weights in a 1-norm ball around $\mathbf{w}^*$ corresponding to the nearest neighbor solution. With probability one, it holds $\mathbf{w}^* = [0, \ldots, 0, 1, 0, \ldots, 0]^T$. Noting $\mathbb{1} = [1, \ldots, 1]^T$, we would like to solve

$$\hat{\mathbf{w}} = \arg \min_{\substack{\|\mathbf{w}-\mathbf{w}^*\|_1 \leq \eta \\ \mathbf{w}^T \mathbb{1} = 1}} \mathbf{w}^T A \mathbf{w}. \tag{6}$$

The parameter $\eta$ tunes the sparsity of the solution. If $\eta = 0$, the solution $\hat{\mathbf{w}}$ corresponds to the nearest neighbor solution. Conversely, considering $\eta = +\infty$ is equivalent to remove the constraint.

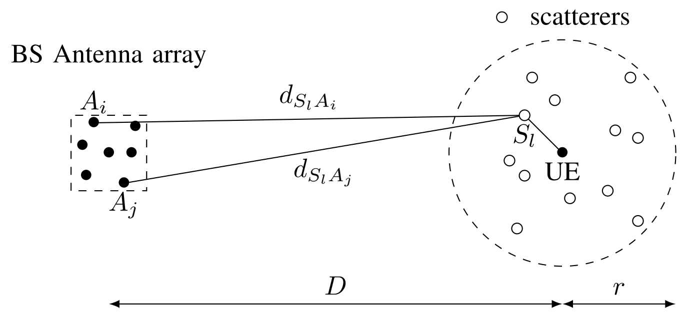


Fig. 2: Illustration of a ray tracing model: the UE antenna is assumed to be surrounded by scatterers randomly located in a ball of radius $r$ (black and white bullets respectively represent antennas and scatterers).

## IV. Performance Evaluation

We now compare simulated performance results corresponding to the introduced interpolation methods for a scenario based on the ring model described in [8] that we briefly recall.

### A. Simulated scenario

We consider an array of $N_R = N_T = 32$ antennas communicating with a single user at distance $D$ uniformly distributed in the interval $[100, 900]$ meters. In order to capture the limited angular spread under which the UE is seen at the BS, the channel is generated using a modification of the ring model of [16], each single-antenna UE being surrounded by scatterers uniformly distributed in a ball of radius $r = 50$ meters containing $N_S = 1000$ scatter points (see Figure 2). The random position of the UE antenna is taken uniformly in the area at a distance between 100m and 900m around the base station. Moreover, the uplink and downlink channels operate at different frequencies (downlink at 1.8 GHz, uplink at 2.8 GHz).
A ray tracing model with a simple quadratic pathloss is assumed, whereby the covariance $\mathbf{R} = (R_{ij})_{1 \leq i,j \leq N}$ of the channel at wavelength $\lambda$ is then modeled by [16]

$$R_{ij} = \frac{P}{D^2 N_S} \sum_{l=1}^{N_S} e^{2i\frac{\pi}{\lambda}(d_{S_l A_i} - d_{S_l A_j})} + P_N \delta_{ij}, \tag{7}$$

where $P$ is the received power at the user side and $P_N$ the power of thermal noise, $d_{S_l A_i}$ denotes the distance between the $l$-th scatterer and the $i$-th BS antenna, and with $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$. For each position of the UE, the uplink/downlink covariance matrices are obtained through the sample covariance computed from $L = 1000$ realizations of the channel $\mathbf{h}$. A realization of the channel is a realization of the random vector $\mathbf{h} = \mathbf{R}^{\frac{1}{2}} \mathbf{w}$ with $\mathbf{w} \sim \mathcal{CN}(0, I_N)$. Note that $\mathbf{R}$ depends on the UE location according to (7).

We will compare the following criteria using the log-Euclidean distance exploring extreme cases for each criterion:

- interpolation with positive weights (eq. (4) with $\eta = 0$),
- extrapolation criterion (eq. (4)) with $\eta = 1$,
- a strong sparse criterion (eq. (6)) with $\eta = 0.2$,

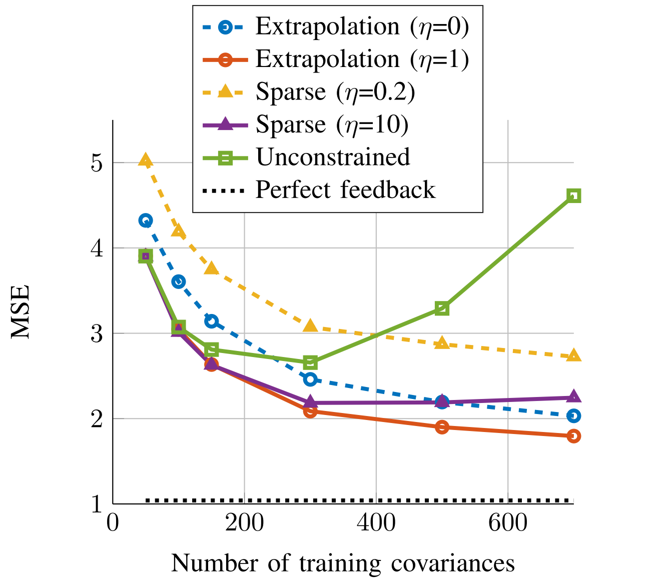


Fig. 3: MSE for different interpolation criteria vs dictionary size

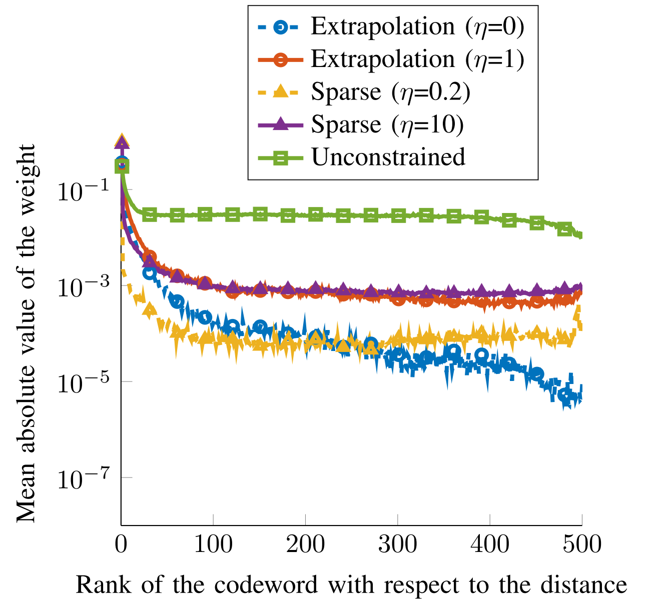


Fig. 4: Mean absolute value of interpolating weights vs. the rank of the codeword sorted with respect to distance to uplink covariance matrix ($K = 500$)

- a weak sparse criterion (eq. (6)) with $\eta = 10$,
- the unconstrained solution ($\eta = +\infty$ for both criteria (4) and (6)),
- the perfect feedback case as a lower bound.

The MSE used for comparison is computed as the mean error between the true covariance matrix $\mathbf{R}$ and the estimated ones $\hat{\mathbf{R}}_n$ where $n$ is the index over $N_{MC}$ Monte Carlo runs

$$\text{MSE} = \frac{1}{N_{MC}} \sum_{n=1}^{N_{MC}} \left\| \log(\mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}_n^{-1} \mathbf{R}^{\frac{1}{2}}) \right\|_{\text{F}}.$$

### B. Performances results

Figure 3 represents the Riemannian interpolation error for $K$ between 50 and 700 (lower than the dimension of the considered space equal to 1024). We also illustrate the sparsity properties in the following way: let $\hat{\mathbf{w}} = [\hat{w}_1, ..., \hat{w}_K]$ be the estimated weights for any criteria associated to the uplink matrix $\mathbf{R}^{\text{UL}}$. We compute the permutation $\sigma$ illustrating the rank of each codeword, i.e. such that

$$d(\mathbf{R}^{\text{UL}}, \mathbf{R}^{\text{UL}}_{\sigma(1)}) < \cdots < d(\mathbf{R}^{\text{UL}}, \mathbf{R}^{\text{UL}}_{\sigma(K)}).$$

We then indicate in Fig. 4 for each $1 \leq k \leq K$ the absolute value $|\hat{w}_{\sigma(k)}|$ averaged over Monte Carlo runs. We deduce from these figures the following insights:

- All criteria favor close codewords. This argues against introducing further neighbor constraint since it is already naturally taken into account.
- When few codewords are available, it is less efficient to consider any constraint. We will then use all available information in the dictionary. However, we see in Fig. 3 that the error of the unconstrained criterion increases with the size of the dictionary as soon as $K > 200$ because the sparse structure of our simulated model is not taken into account.
- The difference between sparse and extrapolation criteria lies in the distribution of the weights amongst the dictionary (see Fig. 4). The sparse criterion with $\eta = 0.2$ leads to consider less weights with high modulus (i.e. closer to $10^{-1}$). The consequence of that behavior is a higher MSE. On the other hand, the criteria that lower too much the impact of far codewords are less efficient.
- The extrapolation citerion with $\eta = 0.1$ cumulating the two aforementioned properties is then the best criterion.

## V. Conclusion

We proposed a study of different criteria for covariance interpolation through Riemannian coding. We considered the regime where the number of covariance matrices in the dictionary is smaller than the dimension of their space. These criteria introduce a parameter tuning the sparsity of the “codes”. It appears that there exists an optimal parameter lowering the approximation error in that regime corresponding to the true degree of sparsity of the simulated model. A perspective for further research would be to chaeck if dictionary learning techniques (see e.g. [9]) could lower the size of dictionary without degrading the approximation error by computing “optimal” codewords from an initial dictionary.

## References

[1] T. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] J. Choi, D. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

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

[4] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing – the large-scale array regime," *IEEE Trans. on Inform. Th.*, vol. 59, no. 10, 2013.

[5] B. Tomasi and M. Guillaud, "Pilot length optimization for spatially correlated multi-user mimo channel estimation," *Proc. Asilomar Conference on Signals, Systems and Computers*, 2015.

[6] B. Hochwald and T. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 642–653, Mar. 2001.

[7] Y. Han, J. Ni, and G. Du, "The potential approaches to achieve channel reciprocity in FDD system with frequency correction algorithms," in *Proc. Communications and Networking in China Conferece (CHINACOM)*, Beijing, China, Aug. 2010.

[8] A. Decurninge, M. Guillaud, and D. Slock, "Channel covariance estimation in massive mimo frequency division duplex systems," *IEEE Global Telecommunications Conference (GLOBECOM)*, 2015.

[9] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," *Advances in Neural Information Processing Systems*, vol. 21, pp. 1033–1040, 2009.

[10] M. Harandi and M. Salzmann, "Riemannian coding and dictionary learning: Kernels to the rescue," 2015, pp. 3926–3935.

[11] Y. Xie, J. Ho, and B. Vemuri, "On a nonlinear generalization of sparse coding and dictionary learning," vol. 28, 2013, pp. 1480–1488.

[12] M. Yin, S. Xie, Y. Guo, J. Gao, and Y. Zhang, "Neighborhood preserved sparse representation for robust classification on symmetric positive definite matrices," *arXiv:1601.07336v1*, 2016.

[13] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, 2004.

[14] M. P. do Carmo, *Riemannian geometry*. Birkhauser, 1992, vol. 1st Edition.

[15] O. Sander, "Geodesic finite elements of higher order," *IMA Journal of Numerical Analysis*, 2015.

[16] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Communications*, vol. 48, no. 3, 2000.