Joint Grassmann-Stiefel Quantization for MIMO Product Codebooks

Renaud-Alexandre Pitaval and Olav Tirkkonen

***Abstract*—We consider product codebook strategy where a single small codebook is implemented at the receiver to quantize larger multi-input multi-output (MIMO) channels, e.g. aggregate channels of cooperative MIMO base stations. The present work focuses on the codebook design, codebook construction, and codeword selection under this scenario, for single- or multistream MIMO transmission. Designing point-to-point unitary precoding codebook is related to a discretization problem on the Grassmann manifold, where a Grassmannian codeword is an equivalence class of rectangular-unitary/Stiefel matrices. For practical needs, one has to choose the rectangular unitary matrix to represent each Grassmann codeword. In this paper, we choose appropriate representatives so that product codebook quantization becomes competitive with global Grassmannian quantization. For this, we propose a novel joint Grassmann-Stiefel codebook design aiming at good quantization/discretization of Grassmann and Stiefel manifolds with a single codebook. To find low-distortion codebooks, we present a vector quantizer generating a Stiefel codebook conditioned on a fixed Grassmann codebook. For this purpose, we provide an exact solution for computing centroids in the Stiefel manifold with chordal distance. Furthermore, concrete examples of analytical joint Grassmann-Stiefel packings are given. Finally, we discuss low-complexity codeword selection methods.**

***Index Terms*—Grassmann manifold, Stiefel manifold, quantization, MIMO, base station cooperation.**

I. Introduction

Multi-Input Multi-Output (MIMO) techniques are key technologies to enhance spectrum efficiency of wireless systems. Performance heavily depends on the channel state information available at the transmitter. In frequency-division duplex systems, the only way to acquire channel state information (CSI) is through a limited feedback channel. A widely applied method is to use codebook-based precoding in which the receiver selects a precoding codeword from a predefined codebook and feeds back the index to the transmitter. Since it is often more important to feed back the channel direction than the channel beam gain [1], [2], the quantization of the eigendirections of the channel is often done with a rectangular unitary code.

In point-to-point communications with maximum likelihood receiver, the performance of a unitary precoding codebook is related to its interpretation as a discretization of the Grassmann manifold [3]–[5], for both single- and multistream transmission. Codewords are rectangular unitary matrices, elements in a Stiefel manifold. However, points in the Grassmann manifold are equivalence classes of rectangular unitary matrices, where two matrices are equivalent if their columns span the same subspace. Accordingly, when creating a Grassmannian codebook, by necessity, for each codeword we must select one of the infinite number of different Stiefel representatives that generates the Grassmannian codeword. In this case, the choice of Stiefel matrices does not impact performance.

Network-level performance enhancements are expected by extending MIMO techniques to multi-point transmission [6]. To overcome the specific features of the MIMO cooperative channel, a per-cell product codebook strategy was proposed in [7]. A per-cell codebook has been advocated to be desirable in a cooperative system for flexibility and compatibility [8], [9], requiring only implementation of a single codebook for any number of cooperative BSs and user positions. With a product codebook structure, the receiver quantizes the aggregate channel matrix by concatenating online codewords from a single per-cell codebook designed offline.

Product codebook-based precoding could also naturally be used in point-to-point scenarios to construct codebooks for large antenna configurations. This flexible method that reuses point-to-point codebooks has many advantages. Only a single per-cell codebook needs to be stored for a fixed transmission rank, reducing design problems to smaller spaces which are typically easier to discretize. Large product codebooks would naturally inherit some properties of interest such as maintaining a small input alphabet or an equal-power constraint per-antenna. For example, the current LTE-A 8Tx codebook is constructed by concatenating 4Tx DFT codewords [10]. Recently, a related quantization structure was also considered for the context of interference alignment [11], where reduction of codeword selection complexity is of interest.

Since matrix concatenation is employed in the product codebook, it is not enough for the per-cell codebook to provide a good discretization of the Grassmann manifold. Instead, quantizing the full space of unitary precoders, the Stiefel manifold, has to be considered. The single implemented per-cell codebook thus needs to provide both a good Grassmann and Stiefel quantization. In this paper, we show that the representatives of a Grassmannian codebook can be chosen appropriately to provide good Stiefel quantization. This leads to a new codebook design problem, where first one has to design a Grassmannian codebook, then one chooses the representative of every codeword to efficiently quantize the Stiefel manifold. The resulting product codebooks show similar performance as global joint-cell Grassmannian codebooks.

This work was presented in part at the IEEE Vehicular Technology Conference (VTC'12 Spring). The work of Renaud-Alexandre Pitaval has been supported in part by the Academy of Finland (grant 254299), and Tekes – the Finnish Funding Agency for Technology and Innovation (grant 548/31/2011).

The authors are with the Department of Communications and Networking, Aalto University, Finland (e-mail: renaud-alexandre.pitaval, olav.tirkkonen@aalto.fi).

We investigate several criteria to design a uniform codebook over the Stiefel manifold conditioned on a Grassmann codebook: maximizing the minimum distance, or maximizing the mean distance, or minimizing an average distortion. Lloyd's algorithm is a standard method to generate low-distortion codebooks. This algorithm requires finding centroids of clusters. We derive a closed-form centroid in the Stiefel manifold equipped with the commonly used Euclidean distance. This is obtained by leveraging a theorem from [12] on centroid computation on closed surfaces embedded into Euclidean space. While applied here for the Stiefel manifold, this result can naturally be extended to other manifolds equipped with an extrinsic Euclidean/chordal metric. Next, we provide an algorithm to generate low-distortion Stiefel codebooks conditioned on predefined Grassmannian codebooks. The proposed algorithm is of Lloyd type, with some non-trivial modifications. In addition, we consider maximizing the minimum distance or the mean distance by Monte Carlo simulations. We give explicit examples of analytical codebooks having low implementation complexity: a Stiefel-improved version of the 2 transmit (Tx) antenna WCDMA Mode 1 codebook with QPSK alphabet, and the optimum 2×1 Stiefel packing conditioned on the optimum Grassmann packing of cardinality 4. The codebooks can be used in product codebooks for any even number of transmit antennas.

Finally, we consider codeword selection methods. As codebook size scales with the number of transmit antennas, the complexity of exhaustive search employed at the receiver scales exponentially [13]. To decrease codeword selection complexity, it was suggested in [14] that the receiver quantizes independently each per-cell channel. While this decreases complexity and gives more flexibility for the network, it induces a loss in performance. This loss is a consequence of dismissing the phase ambiguity between per-cell channels, as recognized in the context of single-stream beamforming in [14]–[16]. To solve this problem it was suggested that the receiver feeds back additional bits related to the phase ambiguity [14], [15], or to use a phase-sensitive algorithm [16]. We describe several selection methods for multistream multicell transmission without additional information to feed back, including a novel independent codeword selection method employing Grassmann and Stiefel distances.

This paper is organized as follows. Section II defines the pertinent mathematical spaces for the codebook design, along with useful mathematical results. In Section III, the system model with product codebook based-precoding of [7] are presented. Section IV discusses the codebook design problem. In Section V, we investigate joint Grassmann-Stiefel codebook constructions, describing Lloyd's algorithms and explicit examples of analytical codebooks. In Section VI are presented different codeword selection principles. Section VII presents simulation results, and finally Section VIII concludes.

II. Preliminaries

A. Spaces of Interest

The primary codebooks we considered are designed for transmission of n_s-streams from a n_t-antenna base station with $n_t \geq n_s$. Pertinent spaces for the codebook design are as follows.

Unitary Group: The codewords addressed have orthonormal columns, and consist thus of a number of columns from a unitary matrix. The space of all n_t-dimensional unitary matrices is denoted by the unitary group:

$$\mathcal{U}_{n_t} = \left\{ \mathbf{U} \in \mathbb{C}^{n_t \times n_t} \quad | \quad \mathbf{U}^H \mathbf{U} = \mathbf{I}_{n_t} \right\}. \tag{1}$$

Stiefel manifold: The complex Stiefel manifold $\mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ is defined as the space of orthonormal rectangular matrices (with $n_s \leq n_t$):

$$\mathcal{V}^{\mathbb{C}}_{n_t,n_s} = \left\{ \mathbf{Y} \in \mathbb{C}^{n_t \times n_s} \quad | \quad \mathbf{Y}^H \mathbf{Y} = \mathbf{I}_{n_s} \right\}. \tag{2}$$

When $n_s = 1$, the Stiefel manifold is the set of unit vectors in $\mathbb{C}^{n_t}$ which can be identified as a hypersphere in $\mathbb{R}^{2n_t}$. Otherwise, for general values of n_s, the Stiefel manifold is a subspace of a hypersphere in $\mathbb{R}^{2n_t n_s}$. The standard distance considered on the Stiefel manifold is thus

$$\begin{aligned} d_s(\mathbf{X}, \mathbf{Y}) &= \|\mathbf{X} - \mathbf{Y}\|_F &(3)\\ &= \sqrt{2n_s - 2\Re(\mathrm{Tr}[\mathbf{X}^H \mathbf{Y}])}. &(4) \end{aligned}$$

Grassmann manifold: The complex Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ is the set of all n_s-dimensional subspaces of $\mathbb{C}^{n_t}$. $\mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ can be expressed as the quotient space of the Stiefel manifold and the unitary group: $\mathcal{G}^{\mathbb{C}}_{n_t,n_s} \cong \mathcal{V}^{\mathbb{C}}_{n_t,n_s}/\mathcal{U}_{n_s}$. A point in the Grassmann manifold can thus be represented as the equivalence class of $n_t \times n_s$ orthonormal matrices whose columns span the same space:

$$[\mathbf{Y}] = \{\mathbf{YU} \quad | \quad \mathbf{U} \in \mathcal{U}_{n_s}\}, \tag{5}$$

where the Stiefel-matrix $\mathbf{Y}$ is an $n_t \times n_s$ matrix with orthonormal columns. Each column in $\mathbf{Y}$ determines a line in $\mathbb{C}^{n_t}$, so that each $\mathbf{Y}$ determines an n_s-dimensional subspace in $\mathbb{C}^{n_t}$. The equivalence class $[\mathbf{Y}]$ thus represents a set of matrices $\mathbf{Y}$ determining the same subspace. Taking two Grassmannian points $[\mathbf{X}]$, $[\mathbf{Y}] \in \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$, the representatives $\mathbf{X}$, $\mathbf{Y} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ determine two subspaces of $\mathbb{C}^{n_t}$. The chordal distance between these is defined as [17]

$$\begin{aligned} d_g([\mathbf{X}], [\mathbf{Y}]) &= \frac{1}{\sqrt{2}} \|\mathbf{X}\mathbf{X}^H - \mathbf{Y}\mathbf{Y}^H\|_F &(6)\\ &= \sqrt{n_s - \|\mathbf{X}^H \mathbf{Y}\|_F^2}. &(7) \end{aligned}$$

This distance does not depend on the representative in $[\mathbf{X}]$ and $[\mathbf{Y}]$ chosen, and is thus well-defined on Grassmann manifolds.

Stiefel and Grassmannian codebook: A code or a codebook is a finite subset of points in the considered space. Since a Grassmannian codebook is a set of equivalence classes, it may be represented by a suitable representative in each equivalence class. The obtained set of rectangular unitary matrices is inherently both a Grassmannian code and a Stiefel code. Namely, a given Stiefel codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_{cb}}\} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ can be independently treated as a discretization of the Grassmann manifold $[\mathcal{C}] = \{[\mathbf{C}_1], \ldots, [\mathbf{C}_{n_{cb}}]\} \subset \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$. The distance properties of the codebook in these two interpretations depend on the design principle. With a slight abuse of notation, we denote the minimum Grassmannian and Stiefel distances of $\mathcal{C}$ by $\delta_g(\mathcal{C})$ and $\delta_s(\mathcal{C})$, respectively.

B. Polar decomposition and Procrustes Problem

Before stating the Procrustes problem, we first recall the *thin* singular value decomposition (SVD) of a complex matrix, and then define the polar decomposition [18]. We recall that it is assumed that $n_s \leq n_t$. Analogous decompositions, for $m \times n$ matrices with $m < n$ are obtained through their Hermitian conjugate.

Thin SVD: For $\mathbf{A} \in \mathbb{C}^{n_t \times n_s}$, there exist $\mathbf{L} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$, $\mathbf{R} \in \mathcal{U}_{n_s}$ and a real diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{n_s}) \in \mathbb{R}^{n_s \times n_s}$ with $\sigma_1 \geq \ldots \geq \sigma_{n_s} \geq 0$ such that $\mathbf{A} = \mathbf{L}\Sigma\mathbf{R}^H$.

Polar decomposition: For $\mathbf{A} \in \mathbb{C}^{n_t \times n_s}$, there exist $\mathbf{U} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ and a unique Hermitian positive semi-definite matrix $\mathbf{P} \in \mathbb{C}^{n_s \times n_s}$, such that $\mathbf{A} = \mathbf{UP}$.

The polar decomposition is unique if $\mathbf{A}$ is full rank and can be easily computed from the thin SVD: given the SVD $\mathbf{A} = \mathbf{L}\Sigma\mathbf{R}^H = \mathbf{L}\mathbf{R}^H\mathbf{R}\Sigma\mathbf{R}^H$, we have $\mathbf{A} = \mathbf{UP}$ with $\mathbf{U} = \mathbf{L}\mathbf{R}^H \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ and $\mathbf{P} = \mathbf{R}\Sigma\mathbf{R}^H$ Hermitian positive semi-definite.

The Procrustes problem is to find the unitary rotation that maps a matrix to be as close as possible to another one.

Procrustes Problem: Given $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n_t \times n_s}$, find

$$\mathbf{U}_{A,B} = \arg \min_{\mathbf{U} \in \mathcal{U}_{n_s}} \|\mathbf{A} - \mathbf{BU}\|_F. \tag{8}$$

The solutions of (8) is given by the polar decomposition of $\mathbf{B}^H\mathbf{A} = \mathbf{U}_{A,B}\mathbf{P}_{A,B}$ [18], [19].

This is related to a more general problem which is to find the nearest Stiefel matrix to a complex matrix.

Nearest Stiefel Matrix: Given $\mathbf{A} \in \mathbb{C}^{n_t \times n_s}$ find

$$\mathbf{V}_A = \arg \min_{\mathbf{V} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}} \|\mathbf{A} - \mathbf{V}\|_F. \tag{9}$$

The solutions of (9) is given by the polar decomposition of $\mathbf{A} = \mathbf{V}_A\mathbf{P}_A$ [18], [19]. The two problems can be seen as equivalent when considering square matrices ($n_s = n_t$).

III. System Model

We define a $(n_{bs} \times n_t) \times n_s$ MIMO system as n_{bs} base stations (BSs) each equipped with n_t antennas transmitting cooperatively n_s-data streams. It is assumed that the BSs are able to instantaneously share the feedback information, e.g. via high speed backhauls.

A. Channel Model

When the BSs transmit to a user, the received signal is

$$\mathbf{y} = \mathbf{H}_{\mathrm{ls}}\mathbf{W}_{\mathrm{ls}}\mathbf{x} + \mathbf{n}, \tag{10}$$

where $\mathbf{y} \in \mathbb{C}^{n_r \times 1}$ is the received vector; $\mathbf{W}_{\mathrm{ls}}$ is an $n_{bs}n_t \times n_s$ aggregate precoding matrix, n_s being the number of streams; $\mathbf{x}$ is an $n_s \times 1$ vector of information symbols satisfying $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \frac{1}{n_s}\mathbf{I}_{n_s}$; $\mathbf{n} \in \mathbb{C}^{n_r \times 1}$ denote the white Gaussian noise with $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \sigma^2\mathbf{I}_{n_r}$; and

$$\mathbf{H}_{\mathrm{ls}} = [\alpha_1\mathbf{H}_1, \ldots, \alpha_{n_{bs}}\mathbf{H}_{n_{bs}}] = \mathbf{H}_{\mathrm{ss}}\mathbf{G} \tag{11}$$

is the large-scale aggregate channel matrix, concatenation of independent channels $\{\alpha_i\mathbf{H}_i\}$ from the BSs to the receiver. The aggregate small-scale channel matrix $\mathbf{H}_{\mathrm{ss}} = [\mathbf{H}_1, \ldots, \mathbf{H}_{n_{bs}}]$ accounts for small-scale flat Rayleigh fading. Its entries are assumed to be independent and identically (i.i.d.) Gaussian with unit variance. Large-scale path gains are in the matrix $\mathbf{G} = \mathrm{diag}(\alpha_1\mathbf{I}_{n_t}, \ldots, \alpha_{n_{bs}}\mathbf{I}_{n_t})$ where α_i incorporates distance-dependent path loss and shadowing from the i^{th} BS to the receiver. For $\alpha_1 = \cdots = \alpha_{n_{bs}}$, the model reduces to a classical $n_{bs}n_t \times n_s$ i.i.d. point-to-point MIMO system.

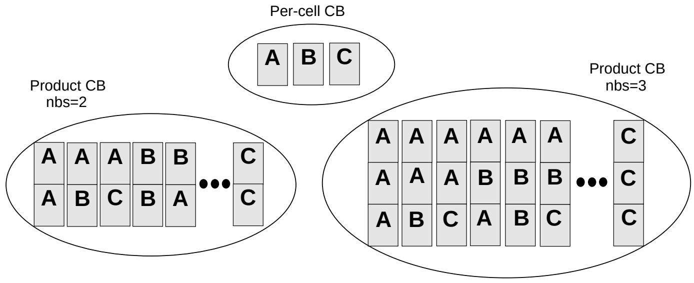


Fig. 1. Illustration of product codebook principle for $n_{bs} = 1$, 2 and 3.

We denote by $\mathbf{V}_{\mathrm{ls}}$ and $\mathbf{V}_{\mathrm{ss}} \in \mathcal{V}^{\mathbb{C}}_{n_{bs}n_t,n_s}$ the left singular vectors associated with the n_s-largest singular values of $\mathbf{H}^H_{\mathrm{ls}}$ and $\mathbf{H}^H_{\mathrm{ss}}$, respectively. The channel coefficients are assumed to be perfectly known at the receiver and unknown at the transmitters. We assume that the BSs have access to CSI only through an error-free, zero delay, and limited feedback channel.

B. Quantization with Product Codebook at the Receiver

We follow the product codebook principle of [7] for feeding back CSI. This was proposed for flexibility and compatibility in order to accommodate to the possible dynamic number of cooperating BSs and to deal with heterogeneous path loss effects. A per-cell codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_{cb}}\}$ of $(n_t \times n_s)$-Stiefel matrices is shared between the transmitters and the receiver. This codebook is independent of the number of cooperating BSs, and large-scale path loss effects. The receiver quantizes $\mathbf{V}_{\mathrm{ss}}$ with a product codebook $\mathcal{C}_{pr}$. The product codebook is a Cartesian product of the per-cell codebook: $\mathcal{C}_{\mathrm{pr}} = \frac{1}{\sqrt{n_{bs}}}\mathcal{C} \otimes \cdots \otimes \mathcal{C}$, i.e. a codeword in $\mathcal{C}_{pr}$ is a normalized concatenation of n_{bs} per-cell codewords as illustrated in Fig. 1. In [7], it was shown that $\mathbf{V}_{\mathrm{ss}}$ satisfies almost surely the structure of a product codeword for $n_t \to \infty$. Finally, the receiver feeds back the set of indexes of the codewords of $\mathcal{C}$ that form the selected product codeword.

C. Precoding at the Transmitter

The BSs reconstruct the small-scale precoding matrix $\mathbf{W}_{\mathrm{ss}} = \mathcal{Q}(\mathbf{V}_{\mathrm{ss}}) \in \mathcal{C}_{pr}$ according to the feedback bits received and corresponding to the quantization map $\mathcal{Q}$ used at the receiver. As in [7], it is assumed that the BSs know the large scale path gains of the channels contained in $\mathbf{G}$. Accordingly, the precoding matrix applied is $\mathbf{W}_{\mathrm{ls}} = \frac{\sqrt{n_t n_{bs}}}{\|\mathbf{G}\|}\mathbf{G}\mathbf{W}_{\mathrm{ss}}$ following the principle of adaptive precoding for correlated MIMO [20], [21]. The power is allocated in proportion to the large scale path gains which intuitively corresponds to the principle of a maximum ratio transmission type. The normalization guarantees that $\mathbf{W}_{\text{ls}}$ belongs to $\mathcal{V}^{\mathbb{C}}_{n_{bs}n_t,n_s}$ and so that the total transmit power is one.

The precoder, which is a function of received feedback bits, is used for channel adaptation in order to increase the information rate of the transmission. With Gaussian input and maximum-likelihood receiver, the average information rate of the system is given by

$$C = \mathbb{E}\left[\log_2 \det\left(\mathbf{I} + \rho \mathbf{W}^H_{\text{ls}} \mathbf{H}^H_{\text{ls}} \mathbf{H}_{\text{ls}} \mathbf{W}_{\text{ls}}\right)\right], \quad (12)$$

where ρ is the SNR per stream, and $\mathbf{W}_{\text{ls}}$ is constructed as described above for each channel realization and thus dependent of the random channel $\mathbf{H}_{\text{ls}}$.

IV. Per-cell Codebook Design for Product Codebooks

Considering the global $(n_{bs} \times n_t) \times n_s$-system, the performance of the product codebook is related to its interpretation as a discretization of the Grassmann manifold [2]–[5]. Optimally, the receiver would quantize the small scale channel as $\mathbf{W}_{\text{ss,opt}} = \mathbf{V}_{\text{ss}}$. This optimum precoder is not unique in the sense that any codeword that is achieved by multiplying it with an $n_s \times n_s$ unitary matrix from the right leads to the same information rate. The set of such optimum precoders can thus be seen as a point in the Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_{bs}n_t,n_s}$.

Several criteria have been investigated in the literature to design good Grassmann precoding codebooks. For an i.i.d Rayleigh fading channel $\mathbf{H}_{\text{ss}}$, the right singular vectors $\mathbf{V}_{\text{ss}}$ are uniformly distributed over the Stiefel manifold according to the Haar measure [22], and the space spanned by $\mathbf{V}_{\text{ss}}$ is uniformly distributed on the Grassmann manifold [5]. The main idea is thus to target uniform codebooks over the manifold. Designing codebooks to directly maximize (12) is untractable. Instead, good design criteria based on extremization of average distortion metrics, i.e. average squared quantization errors, have been considered [2], [21], [23]–[25]. The information rate can be approximated as a function of the distortion $D_g(\mathcal{C}_{pr}) = \mathbb{E}_{\mathbf{V}_{ss}}\left[\min_{\mathbf{W}_{ss}\in\mathcal{C}_{pr}} d^2_g(\mathbf{V}_{ss}, \mathbf{W}_{ss})\right]$ where the Grassmann chordal distance (6) is used as a quantization map [2], [25]. In [3], [4] the beamforming codebook design problem was linked to another approach, maximizing the minimum Grassmann distance of the codebook $\delta_g(\mathcal{C}_{pr}) = \min_{1\le i,j\le n_{cb}^{n_{bs}}} d_g(\mathbf{C}_{pr,i}, \mathbf{C}_{pr,j})$ with $\mathbf{C}_{pr,i}, \mathbf{C}_{pr,j} \in \mathcal{C}_{pr}$. Maximizing average pairwise distance has also been considered in [26]. From the point of view of precoding, these criteria are almost optimal and roughly equivalent. Extension to correlated channels through codebook rotation has been discussed in [10], [20], [21].

A. Space of Quantization for Per-Cell Codebook

The optimum global precoder can be written without loss of generality in terms of component codewords as $\mathbf{W}_{\text{ss,opt}} = [\mathbf{W}^H_{\text{ss,opt},1}, \dots, \mathbf{W}^H_{\text{ss,opt},n_{bs}}]^H$ where $\mathbf{W}_{\text{ss,opt},i} \in \mathbb{C}^{n_t\times n_s}$ would be an optimum quantization of the i^{th} component of $\mathbf{V}_{\text{ss}}$. While performance is invariant under joint right unitary rotation of all components, the unitary invariance does not hold for each component separately. This means that the first component $\mathbf{W}_{\text{ss,opt},1}$ may be any $n_t \times n_s$ matrix of norm less than $\sqrt{n_s}$, up to right unitary rotation. However, given the first component $\mathbf{W}_{\text{ss,opt},1}$, there is not longer a freedom to rotate the subsequent component $\mathbf{W}_{\text{ss,opt},i}$, $i > 1$. Meaning that the $\mathbf{W}_{\text{ss,opt},i}$ for $i > 1$ may be any $n_t \times n_s$ matrix of norm less than $\sqrt{n_s}$. For $n_s > 1$ it should be noted that the per-BS optimum precoder $\mathbf{W}_{\text{ss,opt},i}$ do not necessarily have orthogonal columns. The product codebook strategy considered here, however, constrains the per-cell quantizations to be unitary. The space of quantization is then the Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ for only one BS, and the Stiefel manifold $\mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ for the other $(n_{bs}-1)$-BSs.

B. Probabilistic Characterization of Per-cell Components

The uniformly distributed Stiefel matrix $\mathbf{V}_{ss} \in \mathcal{V}^{\mathbb{C}}_{n_{bs}n_t,n_s}$ is characterised by the Haar measure giving a probability density that satisfies $p(\mathbf{V}_{ss}) = p(\Phi\mathbf{V}_{ss})$ for all $\Phi \in \mathcal{U}_{n_{bs}n_t}$ [27]. The definition is naturally extended to the distribution of equivalence classes $[\mathbf{V}_{ss}]$ in the Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_{bs}n_t,n_s}$. Assuming the Haar distribution, it is possible to derive the distribution of the Stiefel part from the polar decomposition of a truncation of $\mathbf{V}_{ss}$.

Lemma 1. *Let $\mathbf{V} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ be Haar distributed, and $\mathbf{A}_{\text{sub}} \in \mathbb{C}^{n_a\times n_s}$, a submatrix of $n_a \ge n_s$ consecutive rows of $\mathbf{V}$. Given the polar decomposition, $\mathbf{A}_{\text{sub}} = \mathbf{V}_{\text{sub}}\mathbf{P}_{\text{sub}}$ where $\mathbf{V}_{\text{sub}} \in \mathcal{V}^{\mathbb{C}}_{n_a,n_s}$, the Stiefel matrix $\mathbf{V}_{\text{sub}} \in \mathcal{V}^{\mathbb{C}}_{n_a,n_s}$ is Haar distributed.*

Proof: Write $\mathbf{V} = [\mathbf{A}^H_1, \mathbf{A}^H_{\text{sub}}, \mathbf{A}^H_2]^H$ with $\mathbf{A}_i \in \mathbb{C}^{n_i\times n_s}$ and consider the block diagonal matrix $\boldsymbol{\theta} = \text{diag}(\mathbf{I}_{n_1}, \theta, \mathbf{I}_{n_2}) \in \mathcal{U}_{n_t}$ with $\theta \in \mathcal{U}_{n_a}$. As $\mathbf{V}$ is Haar distributed, we have $p(\mathbf{V}) = p(\boldsymbol{\theta}\mathbf{V})$, or equivalently in terms of the joint probability density of the parts of $\mathbf{V}$, $p(\mathbf{V}_{\text{sub}}, \mathbf{P}_{\text{sub}}, \mathbf{A}_1, \mathbf{A}_2) = p(\theta\mathbf{V}_{\text{sub}}, \mathbf{P}_{\text{sub}}, \mathbf{A}_1, \mathbf{A}_2)$. By integrating the joint density, it is a direct verification that the marginal density of $\mathbf{V}_{\text{sub}}$ satisfies $p(\mathbf{V}_{\text{sub}}) = p(\theta\mathbf{V}_{\text{sub}})$ for any $\theta \in \mathcal{U}_{n_a}$. ■

The optimum Stiefel quantization of a component of $\mathbf{V}_{\text{ss}}$ is also Haar distributed according to Lemma 1. For the first BS, $\mathcal{C}$ thus has to be a uniform Grassmannian codebook, whereas for the remaining BSs, $\mathcal{C}$ also has to be a uniform Stiefel codebook. Note that due to (5) there exists an infinity of Stiefel codebooks that represent the same Grassmannian codebook.

C. Design Criteria and Chordal Distances

The Grassmann distances of the product codebooks $\mathcal{C}_{pr}$ depend on the Stiefel representatives chosen for the per-cell codebook $\mathcal{C}$. To construct uniform per-cell codebook, we extend the standard Grassmann codebook criteria to the Stiefel manifold with chordal distance. Several non-equivalent distances on Grassmann and Stiefel manifold can be defined. The following lemma relates the Grassmann chordal distance of the product codebook to the Grassmann and Stiefel chordal distance of the per-cell codebook.

Lemma 2. *Assume the product codebook $\mathcal{C}_{\text{pr}} = \frac{1}{\sqrt{n_{bs}}}\mathcal{C} \otimes \cdots \otimes \mathcal{C}$ in $\mathcal{V}^{\mathbb{C}}_{n_{bs}n_t,n_s}$ constructed from* *per-cell codebook $\mathcal{C} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$, with $n_s < n_t$. The minimum Grassmann distance of the product codebook satisfies*

$$\delta_g^2(\mathcal{C}_{pr}) \geq \min\left\{\delta_g^2(\mathcal{C}), \frac{\delta_g^2(\mathcal{C}) + (n_{bs}-1)\delta_s^2(\mathcal{C})}{n_{bs}^2}\right\} \quad (13)$$

where $\delta_g(\mathcal{C})$ and $\delta_s(\mathcal{C})$ are the Grassmann and Stiefel minimum distance of the per-cell codebook, respectively.

Proof: There are two cases to consider. First, assume that the minimum distance is achieved with product codewords that are concatenations of single per-cell codewords. In other words, let $\mathbf{V}, \mathbf{W} \in \mathcal{C}_{pr}$ with $\mathbf{V} = \frac{1}{\sqrt{n_{bs}}}[\mathbf{V}_1^H, \mathbf{V}_1^H, \ldots, \mathbf{V}_1^H]^H$ and $\mathbf{W} = \frac{1}{\sqrt{n_{bs}}}[\mathbf{W}_1^H, \mathbf{W}_1^H, \ldots, \mathbf{W}_1^H]^H$, such that $\delta_g(\mathcal{C}_{pr}) = d_g(\mathbf{V}, \mathbf{W})$. Then it follows that

$$\begin{aligned} \delta_g^2(\mathcal{C}_{pr}) &= n_s - \|\mathbf{V}^H\mathbf{W}\|^2 = n_s - \|\mathbf{V}_1^H\mathbf{W}_1\|^2 && (14)\\ &= d_g^2(\mathbf{V}_1, \mathbf{W}_1) \geq \delta_g^2(\mathcal{C}). && (15) \end{aligned}$$

Now assume that the two closest product codewords are concatenations of different per-cell codewords. Then the minimum possible distance is satisfied for two product codewords that only differ by a single per-cell codeword, i.e. let $\mathbf{V}, \mathbf{W} \in \mathcal{C}_{pr}$ with $\mathbf{V} = \frac{1}{\sqrt{n_{bs}}}[\mathbf{V}_1^H, \ldots, \mathbf{V}_{n_{bs}}^H]^H$ and $\mathbf{W} = \frac{1}{\sqrt{n_{bs}}}[\mathbf{W}_1^H, \mathbf{V}_2^H, \ldots, \mathbf{V}_{n_{bs}}^H]^H$ such that $\delta_g(\mathcal{C}_{pr}) = d_g(\mathbf{V}, \mathbf{W})$. The minimum distance can be simplified and bounded as

$$\begin{aligned} \delta_g^2(\mathcal{C}_{pr}) &= n_s - \frac{1}{n_{bs}^2}\|(n_{bs}-1)\mathbf{I}_{n_s} + \mathbf{V}_1^H\mathbf{W}_1^H\|^2 && (16)\\ &= \frac{1}{n_{bs}^2}\left(d_g^2(\mathbf{V}_1, \mathbf{W}_1) + (n_{bs}-1)d_s^2(\mathbf{V}_1, \mathbf{W}_1)\right) \\ &\geq \frac{1}{n_{bs}^2}\left(\delta_g^2(\mathcal{C}) + (n_{bs}-1)\delta_s^2(\mathcal{C})\right). && (17) \end{aligned}$$

Combining the two cases leads to (13). ■

Intuitively from Lemma 2 one sees that the importance of the Stiefel distance increases with the number of cooperative BSs.

D. Product Codebook Performance

Using a product codebook strategy results in two sub-optimalities.
i) A per-cell codebook designed to quantize the Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ does not necessarily result in a good quantization of the Stiefel manifold $\mathcal{V}^{\mathbb{C}}_{n_t,n_s}$.
ii) A residual loss would be also expected compared to a global codebook quantizing the larger Grassmannian $\mathcal{G}^{\mathbb{C}}_{n_{bs}n_t,n_s}$, corresponding to the signal eigenspace of the receiver.

To make the performance of product codebook quantization close to optimal, we propose a novel joint Grassmann-Stiefel design of the per-cell codebook $\mathcal{C}$. An example of the achieved performance of such a design is illustrated in Fig. 2 and Fig. 3. In Fig. 2, the information rate is shown for systems with $n_{bs}n_t = 2, 4, 6, 8$ transmitter antennas and a single receiver antenna, and the information rate of systems with 4, 8, 12, 16 transmitter antennas and two receiver antennas. There is one feedback bit per transmit antenna. In Fig. 3, the information rate is shown for $n_{bs} = 2$ base stations, and 1- to 5-bit per-cell codebooks for systems with $n_t = 2$ and $n_s = 1$, and $n_t = 4$

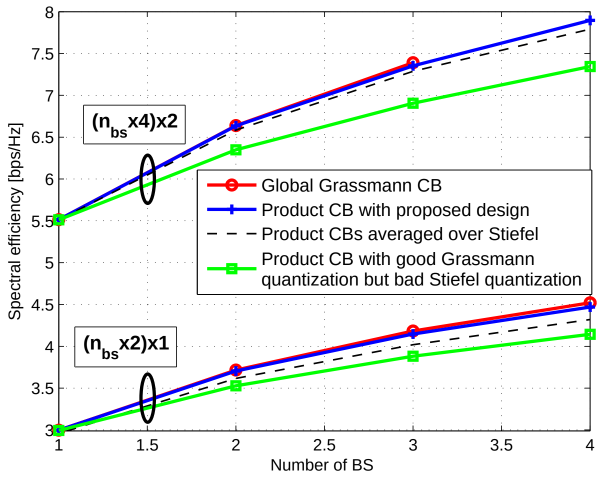


Fig. 2. Comparison of product codebooks with joint Grassmann-Stiefel design and global Grassmannian codebooks. The curves show the information rates at 10 dB SNR for 2x1, 4x1, 6x1, and 8x1 MIMO systems, as well as for 4x2, 8x2, 12x2, and 16x2 MIMO systems. Codebooks with one feedback bit per transmit antenna.

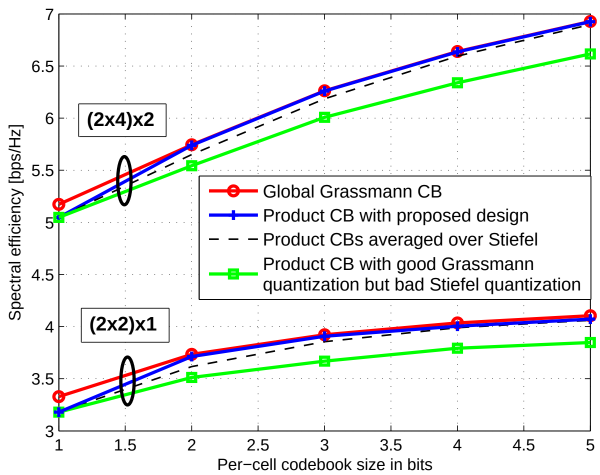


Fig. 3. Comparison of product codebooks with joint Grassmann-Stiefel design and global Grassmannian codebooks. The curves show the information rates at 10 dB SNR for 4x1 and 8x2 with 1- to 5-bit 2x1 and 4x2 per-cell codebook, respectively.

and $n_s = 2$. This corresponds to a 4x1 global system with 0.5 to 2.5 feedback bits per transmit antenna, and a 8x2 global system with 0.25 to 1.25 feedback bits per transmit antenna, respectively.

Except when the per-cell codebook size is equal to one bit, the average performance of the proposed design is close to that of a global Grassmannian codebook minimizing the average distortion D_g (constructed here via Lloyd's algorithm). The gain of the proposed design is illustrated by comparison with a product codebook based on the same per-cell Grassmannian codebook but with (putatively) worst choice of Stiefel representatives. The mean performance of the Grassmann codebook averaged over all possible Stiefel representatives is also shown. The gap between the best and worst product codebook increase with increasing number of BSs, while relatively constant with increasing feedback bits. When the codebook size grows, product codebook performance averaged over Stiefel representatives is asymptotically reaching the performance of global codebook.

V. Joint Grassmann-Stiefel Codebooks

In order to have good per-cell codebooks that can be used in product codebooks as discussed above, we propose that a codebook is constructed by first designing a Grassmannian codebook according to standard criteria such as maximizing the minimum distance or minimizing the average distortion. Then, the representative in each Grassmannian plane in the codebook is chosen to optimize a metric on the Stiefel manifold. This means that we select a good Stiefel codebook conditioned on the codebook being simultaneously a good Grassmannian codebook.

A. Problem Statement

Let us assume that the Stiefel codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_{cb}}\} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ generates the desired Grassmannian codebook $[\mathcal{C}] = \{[\mathbf{C}_1], \ldots, [\mathbf{C}_{n_{cb}}]\} \subset \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$. There exists an infinite number of possible Stiefel codebooks that generate the same Grassmann codebook; for any set of unitary rotations $\Omega = (\mathbf{U}_1, \ldots, \mathbf{U}_{n_{cb}}) \subset \mathcal{U}_{n_s}$, the Stiefel codebook $\mathcal{C}\Omega = \{\mathbf{C}_1\mathbf{U}_1, \ldots, \mathbf{C}_{n_{cb}}\mathbf{U}_{n_{cb}}\} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ generates the same Grassmannian codebook as $\mathcal{C}$, i.e. $[\mathcal{C}\Omega] = [\mathcal{C}]$. The problem is to find the best Stiefel representative of $[\mathcal{C}]$, or equivalently finding the $\Omega = (\mathbf{U}_1, \ldots, \mathbf{U}_{n_{cb}}) \subset \mathcal{U}_{n_s}$ that extremizes a given criterion. Without loss of generality, the first codeword can be fixed such that $\mathbf{U}_1 = \mathbf{I}$.

B. A Simple Example

To illustrate the joint Grassmann-Stiefel codebook design problem, we consider the toy scenario of building a real codebook of four codewords for a transmission from 3 antennas. This leads to a rare example where visualization of the proposed approach is possible. The real Grassmannian $\mathcal{G}^{\mathbb{R}}_{3,1}$ that needs to be discretized is the set of lines through the origin in the 3D Euclidean space. It can be understood as the set of antipodal points on the real unit sphere. The corresponding Stiefel manifold is the space of all 3D unit-norm vectors, and can be understood as the full sphere. A Grassmannian code is then a set of antipodal points, and choosing a representative for every Grassmannian codeword means simply choosing one of the two antipodal points on the sphere. A Stiefel-codebook, in turn, is a spherical code. The best four-codeword Grassmannian packing is found by taking the vertices of a cube – the eight vertices of the cube consist of four pairs of antipodal points, i.e. four Grassmannian lines. From this cube, there is four possible non-equivalent four-codeword spherical codes: for example by taking only points in the upper hemisphere we get a square, or by taking two points in both upper and lower hemispheres we get a tetrahedron as depicted on Fig 4. Those two alternative Stiefel codebooks generating the same Grassmannian code are given in Appendix A. The best Grassmann-Stiefel codebook is obtained by taking the vertices of the cube that form a tetrahedron. It turns out that the vertices of the tetrahedron gives actually the optimal 4-point spherical (Stiefel) codes under several criteria [26]. In this simple example, it is thus possible to have a codebook that is simultaneously an optimal Grassmannian and Stiefel packing.

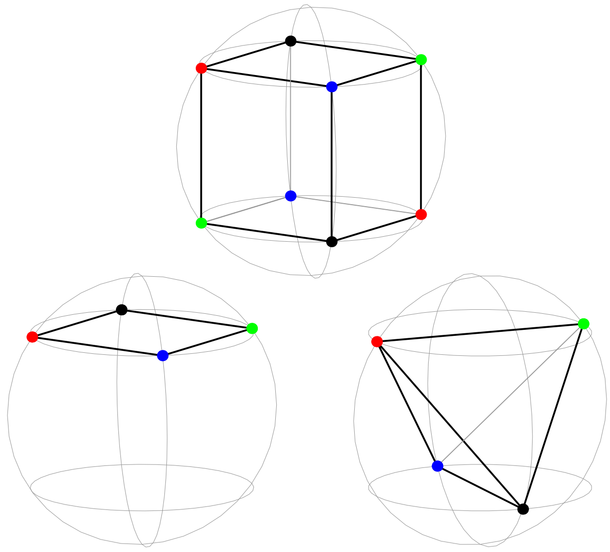


Fig. 4. Joint Grassmannian-Stiefel codebook design in toy scenario of real codebook for 3 transmit antennas. On the upper graph the optimum 2-bit Grassmannian packing in $\mathcal{G}^{\mathbb{R}}_{3,1}$, a set of 4 antipodal points forming a cube. A Grassmannian codeword may be represented by any of the two points of same color, lying on a line through the origin. On the lower part, two alternatives of 2-bit Stiefel codebooks generating the above Grassmannian codebook: a square and a tetrahedron.

C. Codebook Design Criteria

Given a codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_{cb}}\} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$, depending on the distance considered, the codebook can be treated either as a discretization of the Grassmann or the Stiefel manifold. In the following, we write the manifold of consideration by $\mathcal{M}$ associated with distance d, we have $d = d_s$ and $d = d_g$ for $\mathcal{M} = \mathcal{V}^{\mathbb{C}}_{n,p}$ and $\mathcal{M} = \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ respectively.

As discussed at the beginning of Section IV, several criteria have been investigated in the literature to design good Grassmann precoding codebook. Those criteria, defining a notion of uniformity, naturally extend for codebook design on Stiefel manifolds. In this section, we discuss the following criteria.

- Minimizing the average distortion:

$$D_{\mathcal{M}}(\mathcal{C}) = \mathbb{E}_{V\in\mathcal{M}}\left[\min_{\mathbf{C}\in\mathcal{C}} d^2(V, \mathbf{C})\right].$$

- Maximizing the minimum distance:

$$\delta_{\mathcal{M}}(\mathcal{C}) = \min_{1\le i\neq j\le n_{cb}} d(\mathbf{C}_i, \mathbf{C}_j),$$

  often referred to as packing or Tammes problem.

- Maximizing the p-mean distance:

$$M_p(\mathcal{C}) = \left(\frac{2}{n_{cb}(n_{cb}-1)} \sum_{1\le j<k\le n_{cb}} d^p(\mathbf{C}_j, \mathbf{C}_k)\right)^{1/p}.$$

This is known as Thomson problem, see references in [26]. Typical values of interest are $p = -1$, corresponding to the harmonic mean, and -2.

These are also mathematical problems of independent interest. In the following, we discuss Lloyd's algorithms generating minimum-distortion codebooks, and describe two examples of closed-form codebooks that could be conjectured (as compared to Monte Carlo simulations) to be optimum packings and Thomson configurations.

D. Low-Distortion Codebooks

1) Lloyd's Algorithm: For any $\mathbf{V} \in \mathcal{M}$, define the quantization map

$$\mathrm{q}(\mathbf{V}) = \arg \min_{1 \le i \le n_{cb}} d(\mathbf{V}, \mathbf{C}_i). \tag{18}$$

Given a random source V on $\mathcal{M}$, the average distortion of the codebook $\mathcal{C}$ on $\mathcal{M}$ is

$$D_{\mathcal{M}}(\mathcal{C}) = \mathbb{E}\left[d^2(V, \mathbf{C}_{\mathrm{q}(V)})\right] \tag{19}$$

Lloyd's algorithm aims to construct a codebook with minimum average distortion. It comprises two key steps:

Nearest Neighbor rule (NN): Partitioning of the manifold according to the codebook in n_{cb} Voronoi cells $\{\mathcal{R}_1, \ldots, \mathcal{R}_{n_{cb}}\}$ defined by

$$\mathcal{R}_k = \{\mathbf{V} \in \mathcal{M} \,|\, k = \mathrm{q}(\mathbf{V})\}. \tag{20}$$

Centroid Computation (CC): Finding the centroids of each Voronoi cell $\mathcal{R}_k$ given by

$$\mathbf{Z}_k = \arg \min_{\mathbf{Z} \in \mathcal{M}} \mathbb{E}\left[d^2(V, \mathbf{Z}) \mid V \in \mathcal{R}_k\right]. \tag{21}$$

The algorithm consists of iterating these two steps where the former codebook is replaced by the set of computed centroids.

2) Centroid Computation on Stiefel and Grassmann Manifolds: The Grassmann and Stiefel manifolds are compact manifolds without borders. Furthermore, the distance d_g and d_s that we consider arise naturally by treating the manifolds as surface embedded in an Euclidean space (more precisely a subset of an Euclidean hypersphere) and taking the canonical extrinsic distance. The exact centroid computation in the Stiefel manifold equipped with the distance d_s follows from this observation and leveraging a recent result from Euclidean geometry literature.

Lemma 3. *Given a Voronoi region $\mathcal{R}_k \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$, a random source V on $\mathcal{V}^{\mathbb{C}}_{n_t,n_s} \subset \mathbb{C}^{n_t \times n_s}$, and the polar decomposition of the center of mass $\mathbb{E}[V \mid V \in \mathcal{R}_k] = \mathbf{U}_k \mathbf{P}_k$, a centroid of $\mathcal{R}_k$ is given by*

$$\begin{aligned} \mathbf{Z}_k &= \arg \min_{\mathbf{Z} \in \mathcal{V}^{\mathbb{C}}_{n_t,n_s}} \mathbb{E}\left[d_s^2(V, \mathbf{Z}) \mid V \in \mathcal{R}_k\right] & (22) \\ &= \mathbf{U}_k & (23) \end{aligned}$$

Proof: As $(\mathcal{V}^{\mathbb{C}}_{n_t,n_s}, d_s)$ is a closed continuous surface in an Euclidean space, according to [12] centroid of a Voronoi region is obtained from the orthogonal projection onto the manifold of its center of mass in the ambient space. Orthogonal projection of any complex matrices to the Stiefel manifold is given by the polar decomposition as recall in Section II-B. ■

A closed-form solution is already known for centroids in the Grassmann manifold [28]. We remark that while the closed-form Grassmannian centroid computation is usually proven differently in the literature, it could also be derived from the same argument, as the centroid corresponds to the closest projection matrices of a given rank to the center of mass in the ambient space of Hermitian matrices. We recall this result below for consistency.

Remark 1. *Given a Voronoi region $\mathcal{R}_k \subset \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ and a random source V on $\mathcal{G}^{\mathbb{C}}_{n_t,n_s} = \mathcal{V}^{\mathbb{C}}_{n_t,n_s}/\mathcal{U}_{n_s} \subset \mathbb{C}^{n_t \times n_s}$. Under the mapping $[\mathbf{Y}] \to \mathbf{Y}\mathbf{Y}^H$, $[\mathbf{Y}] \in \mathcal{G}^{\mathbb{C}}_{n_t,n_s}$, corresponds to V an equivalent random variable VV^H in the sample space of $n_t \times n_t$ Hermitian matrices. Given the eigenvalue decomposition of the center of mass $\mathbb{E}\left[VV^H \mid V \in \mathcal{R}_k\right] = \mathbf{U}_k \Lambda_k \mathbf{U}_k^H$ with $\Lambda_k = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_t})$ such that $\lambda_1 \ge \ldots \ge \lambda_{n_t}$, a centroid of $\mathcal{R}_k$ is*

$$\begin{aligned} \mathbf{Z}_k &= \arg \min_{\mathbf{Z} \in \mathcal{G}^{\mathbb{C}}_{n_t,n_s}} \mathbb{E}\left[d_g^2(V, \mathbf{Z}) \mid V \in \mathcal{R}_k\right] & (24) \\ &= [\mathbf{U}_k \mathbf{I}_{n_t,n_s}] & (25) \end{aligned}$$

where $\mathbf{U}_k \mathbf{I}_{n_t,n_s}$ correspond to the first n_s columns of $\mathbf{U}_k$.

Note that the solution to the centroid problem may not be unique [12]. Also in practice, the source is realized by a finite training set generated according to the distribution of interest, the expectation is then approximated by an arithmetic mean.

3) Lloyd-type Algorithm for Joint Grassmann-Stiefel codebook: A Lloyd-type algorithm to generate a low-distortion Stiefel codebook conditioned on a Grassmann codebook is summarized in Algorithm 1. Non-trivial differences as compared to the conventional Lloyd's algorithm can be seen in Step 4). In each Voronoi cell $\mathcal{R}_k$ of the Stiefel manifold, the original codeword $\mathbf{C}_k$ is not replaced by the computed centroid $\mathbf{Z}_k$. Instead, the algorithm is looking for the closest codeword $\mathbf{C}_i$ to the the centroid $\mathbf{Z}_k$ using Grassmann distance rather than Stiefel distance. As a result, for each Voronoi cell $\mathcal{R}_k$, the updated codeword $\mathbf{C}_i$ is not necessarily $\mathbf{C}_k$. Then $\mathbf{C}_i$ is replaced by the Stiefel representative in $[\mathbf{C}_i]$ closest to $\mathbf{Z}_k$. Indeed, during a single iteration some codewords can be updated several times and some others not at all. This occurs in particular in the first iterations. This phenomenon is related to the non-trivial embedding of the Grassmannian into the Stiefel manifold, and is crucial for convergence.

Algorithm 1 is illustrated in Fig. 5 for a toy scenario of choosing the Stiefel representative of a real Grassmann codebook in $\mathcal{G}^{\mathbb{R}}_{2,1}$. The Stiefel manifold in this case is the unit circle $\mathcal{S}^1$, and the Grassmannian is the set of lines through the origin in 2D, or pairs of antipodal points on a circle. At Step 1) , the Stiefel representatives of the three Grassmannian lines have been given in the right half circle. In Steps 2) and 3), the algorithm generates a random source and partitions the Stiefel manifold based on a nearest neighbor rule. The Stiefel Voronoi cells corresponding to these codewords are depicted in blue, red and orange. Non-trivial differences as compared to the conventional Lloyd's algorithm can be seen in Step 4) where the algorithm sequentially computes a centroid and

Algorithm 1 Lloyd-type algorithm on Stiefel manifold conditioned on a Grassmannian codebook.

1) **Initialization:** Take Stiefel codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_{cb}}\} \subset \mathcal{V}^{\mathbb{C}}_{n_t,n_s}$, a representative of the desired Grassmannian codebook.
2) **Source:** Generate random source V in $\mathcal{V}^{\mathbb{C}}_{n_t,n_s}$.
3) **NN:** Partition $\mathcal{V}^{\mathbb{C}}_{n_t,n_s}$ in Voronoi cells $\{\mathcal{R}_1, \ldots, \mathcal{R}_{n_{cb}}\}$ according to (18) and (20) with Stiefel distance d_s.
4) For all k perform the following
   a) **Centroid:** Compute Stiefel centroid $\mathbf{Z}_k$ according to Lemma 3.
      i) *Center of mass*: $\mathbf{M}_k = \mathbb{E}\left[V \mid V \in \mathcal{R}_k\right]$.
      ii) *Polar decomposition:* $\mathbf{M}_k = \mathbf{Z}_k \mathbf{P}_k$.
   b) Find the Grassmannian plane from $\mathcal{C}$ closest to $\mathbf{Z}_k$:
   $$i = \arg \min_{1 \le l \le n_{cb}} d_g(\mathbf{C}_l, \mathbf{Z}_k). \tag{26}$$
   c) **Procrutes problem:** Find rotation between the centroid and the Stiefel matrix generating the closest Grassmannian plane $\mathbf{C}_i$:
   $$\mathbf{R} = \arg \min_{\mathbf{U} \in \mathcal{U}_{n_s}} d_s(\mathbf{Z}_k, \mathbf{C}_i \mathbf{U}). \tag{27}$$
   A solution is given by the polar decomposition of $\mathbf{C}_i^H \mathbf{Z}_k$ as discussed in Section II-B.
   d) **Update:** Replace codeword i
   $$\mathbf{C}_i \leftarrow \mathbf{C}_i \mathbf{R} \tag{28}$$
5) Loop back to Step 2) until convergence.

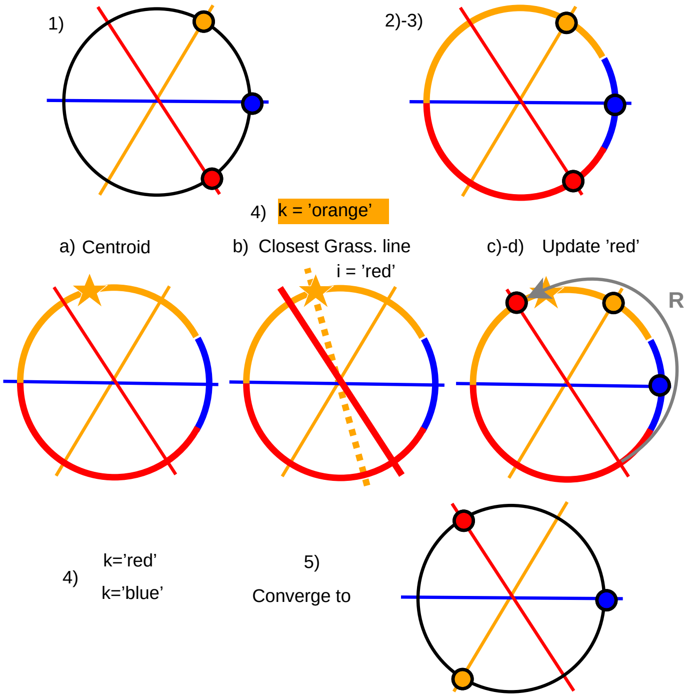


Fig. 5. Illustration of Algorithm 1 with a codebook in real Grassmannian $\mathcal{G}^{\mathbb{R}}_{2,1}$ and real Stiefel manifold $\mathcal{V}^{\mathbb{R}}_{2,1} \cong \mathcal{S}^1$.

updates a codeword. The centroid of the orange Voronoi cell is depicted in a). It happens to be closer to the red Grassmannian line than to the orange one as depicted in b). Thus in c)-d) we update the Stiefel representative of the red line to this centroid, not the representative of the orange line. Next, if we consider the centroid of the red Voronoi region, we update the representative of the orange line, whereas the centroid of the blue region leads to the representative of the blue line being fixed. As a consequence, we have found the optimum three-element Grassmann-Stiefel packing in 5).

In Algorithm 1, the respective computation steps on Grassmann and Stiefel manifolds are clearly delineated. Note that computation of the centroid $\mathbf{Z}_k$ is unnecessary, and comparison can be directly done w.r.t. the center of mass: Equation (26) is equivalent to $i = \arg\min_{1 \le l \le n_{cb}} \|\mathbf{C}_l \mathbf{C}_l^H - \mathbf{M}_k \mathbf{M}_k^H\|_F$, and Equation (27) to $\mathbf{R} = \arg\min_{\mathbf{U} \in \mathcal{U}_{n_s}} \|\mathbf{M}_k - \mathbf{C}_i \mathbf{U}\|$. Alternatively, the two exhaustive searches of Equations (26) and (27) could be done jointly as $[i, \mathbf{R}] = \arg\min_{l, \mathbf{U}} \|\mathbf{M}_k - \mathbf{C}_l \mathbf{U}\|$.

Simulation results show that the proposed algorithm converges well as shown in Fig. 6 wherein the random source is taken uniformly distributed. In Fig. 7, we compare the average Grassmann and Stiefel distortions of codebooks generated in three ways. As benchmarks, we generate Grassmann codebooks and Stiefel codebooks with their respective standard Lloyd's algorithms. These are compared to codebooks generated by cascading two numerical searches — first Lloyd's algorithm is used to generate a minimum distortion Grassmannian codebook, then Algorithm 1 is used to generate a Stiefel codebook conditioned on the found Grassmannian codebook. The codebooks generated by the cascade of algorithms have thus been numerically optimized to have both low Grassmann and low Stiefel distortions. Random sources were simulated with $2 \cdot 10^4$ trials; convergence was assessed when distortion was fluctuating below a $5 \cdot 10^{-4}$ precision threshold, or if a maximum of 30 iterations was reached. As it can be seen from Fig. 7, conditioning the Stiefel codebook on a Grassmannian one leads only to a minor loss in the Stiefel distortion except when the cardinality is two. On the other hand, with standard Grassmann and Stiefel Lloyd's algorithms, one does not have control on the distortion in the other space. The distortion in the other space takes random values so we have plotted an average over several codebooks. The resulting average distortions are worse than the one obtained from the cascade of algorithms. In the special case $n_{cb} = 2$, the average Stiefel distortion is not improved from Algorithm 1, the optimum Grassmannian codebook is generated by orthogonal Stiefel matrices, fixing the Stiefel distance. On the other hand, an optimum 2-point Stiefel codebook corresponds to two antipodal points belonging to the same Grassmannian plane.

E. Maximum Distances Codebooks by Monte-Carlo Method, and Brute-Force Search on Restricted Alphabet

In the previous section, we optimized the properties of codebooks as quantizations on the manifolds. Here, we look at the distance properties of codebooks as point sets on the respective manifolds. Pertinent measures to optimize would be maximizing the minimum distance or the average distance.

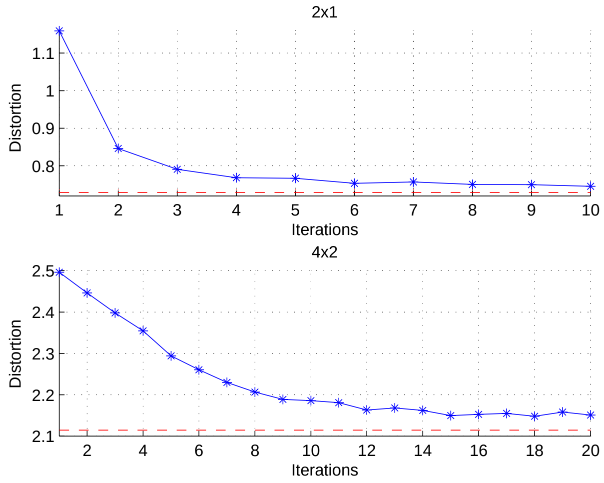


Fig. 6. Average Stiefel distortion versus iterations from the proposed Lloyd-type algorithm for Stiefel codebook conditioned on Grassmann codebook. Upper graph: $\mathcal{V}^{\mathbb{C}}_{2,1}$, conditioned on the Square Codebook in Table I. Lower graph: $\mathcal{V}^{\mathbb{C}}_{4,2}$, conditioned on C-Codebook in [29]. Red dashed lines give the obtained distortion with Lloyd's algorithm with non-constrained Stiefel codebook.

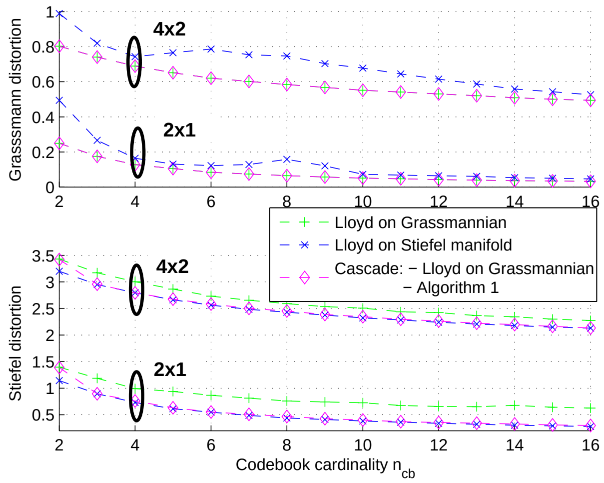


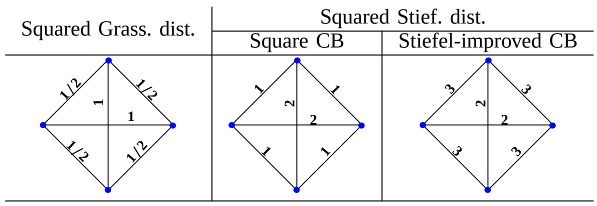
Fig. 7. Average distortions for Stiefel codebook obtained by a cascade of Lloyd's algorithms, the first generating Grassmannian codebooks, the second, Algorithm 1 generating a Stiefel codebooks conditioned on top of the Grassmannian codebooks. Comparison is done with average distortions of Grassmann and Stiefel codebooks from standard Lloyd's algorithm.

We consider Monte Carlo simulations for maximization of the minimum and/or average Stiefel distance of a Grassmannian codebook. The search is performed by right-multiplication of the Stiefel codeword by Haar-distributed unitary matrices.

Unlike Lloyd's algorithm, the search can be easily restricted to additional constraints of interest. For example to decrease the implementation complexity and storage at the receiver, codebooks with a small finite number of entries have been considered [26], [29]–[32]. In [26], it was found that maximizing the average distance may be a slightly better approach for precoding codebook than maximum minimum distance. Here, with a finite alphabet constraint, several codebooks may lead to the same maximum minimum distance; choosing the one maximizing the average distance tends to have slightly

TABLE I
2-BIT *Square Codebook* FOR 2TX ANTENNA AND MODIFIED CODEBOOK MAXIMIZING STIEFEL DISTANCES.

Square CB	$\frac{1}{\sqrt{2}}\left\{\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}\begin{bmatrix}1\\i\end{bmatrix}\begin{bmatrix}1\\-i\end{bmatrix}\right\}$
Stiefel-improved CB	$\frac{1}{\sqrt{2}}\left\{\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}\begin{bmatrix}-1\\-i\end{bmatrix}\begin{bmatrix}-1\\i\end{bmatrix}\right\}$



better performance.

In Table I, we give a Stiefel-improved version of the 2-bit *Square Codebook* [26], which is related to the Mode 1 codebook of WCDMA [33] and the LTE codebook for 2-transmit antennas [32]. The modified version is obtained by only changing the sign of the third and last codeword. The squared Stiefel distances between the codewords of the proposed codebook are either 2 or 3, while for the original codebook they were either 1 or 2. This codebook has been found by brute-force search over a QPSK alphabet for the three last codewords. Furthermore, this is putatively the best Stiefel codebook conditioned the Square Codebook. Monte Carlo simulations over all possible phases suggest that it maximizes the $p-$mean Stiefel distance for $p = 1, 2, -1$ and -2.

In Appendix B, we give another example of codebook having small number of different entries. This codebook is an optimum Grassmann Thomson (p-mean distance) and packing (minimum distance) configuration, the *Tetrahedron Codebook*, with putatively optimum Stiefel-conditioned representatives for Thomson and packing problem.

VI. Codeword Selection

For maximizing the information rate, the optimal method for codeword selection is clearly to maximize the quantity inside the expectation of (12) over all possible product codewords. For single-rank transmission, this is equivalent to minimizing the Grassmannian distance between the product codeword and the right eigenspace of the channel. For higher-rank transmission, minimizing the Grassmann distance provides in practice almost optimal performance, and asymptotical optimality with increasing codebook size [2].

Optimum quantization of the $n_{bs}n_t \times n_s$ channel eigenspace requires exhaustive search over the product codebook of cardinality $n_{cb}^{n_{bs}}$. This leads to exponential complexity w.r.t the number of BSs $\mathcal{O}\left(n_{cb}^{n_{bs}}\right)$. Complexity can be reduced from $\mathcal{O}\left(n_{cb}^{n_{bs}}\right)$ to $\mathcal{O}\left(n_{bs}n_{cb}\right)$ by selecting per-cell components rather than selecting jointly the product codeword [7], [14]. In [7], a lower-complexity selection algorithm trading performance against complexity is proposed, based on two successive exhaustive searches of size $n_{bs}n_{cb}$ and $k^{n_{bs}}$, where k is the cardinality of preselected per-cell sub-codebooks.

Independent and serial selection are proposed for single-stream beamforming in [14], [16].

Here, we discuss five different codeword selection principles for multi-stream product codebooks. Two joint selection methods of complexity $\mathcal{O}\left(n_{cb}^{n_{bs}}\right)$ are considered, followed by three selections of complexity $\mathcal{O}\left(n_{bs}n_{cb}\right)$. For non-unitary matrices the distances d_s and d_g are defined in (3), (6) respectively.

a) Joint Codeword Selection: In joint selection [7], the product codebook codeword minimizing

$$\mathbf{W}_{\mathrm{ss}} = \mathcal{Q}_{\mathrm{js}}(\mathbf{V}_{\mathrm{ss}}) = \arg\min_{\mathbf{C}_{\mathrm{pr}}\in\mathcal{C}_{\mathrm{pr}}} d_g(\mathbf{C}_{\mathrm{pr}}, \mathbf{V}_{\mathrm{ss}}) \quad (29)$$

is selected.

b) Joint Codeword Selection with Transformed Codebook: Joint selection can be improved in case of path loss imbalance [14] by borrowing the idea of transformed codebook for spatially correlated channel [20]:

$$\mathbf{W}_{\mathrm{ss}} = \mathcal{Q}_{\mathrm{js/tr}}(\mathbf{V}_{\mathrm{ls}}) = \arg\min_{\mathbf{C}_{\mathrm{pr}}\in\mathcal{C}_{\mathrm{pr}}} d_g(\mathbf{G}\mathbf{C}_{\mathrm{pr}}, \mathbf{V}_{\mathrm{ls}}). \quad (30)$$

The two joint selection methods provide same performance for $\mathbf{G} \propto \mathbf{I}$.

c) Independent Grassmann Codeword Selection: As an alternative, each single cell channel component could be quantized independently [14]:

$$\mathbf{W}_{\mathrm{ss},k} = \mathcal{Q}_{\mathrm{ind}}(\mathbf{V}_{\mathrm{ss},k}) = \arg\min_{\mathbf{C}\in\mathcal{C}} d_g(\mathbf{C}, \mathbf{V}_{\mathrm{ss},k}). \quad (31)$$

This method leads to a loss of performance as it does not take into account the phase ambiguity between the components of the optimum precoding vector as recognized in [14], or the more general unitary matrix ambiguity.

d) Independent Grassmann-Stiefel Codeword Selection: In order to quantize the per-cell channel components independently and efficiently, the unitary matrix ambiguity between the different channels should be taken into account. We suggest that first the strongest channel (with the largest α_i) is quantized using the Grassmannian distance

$$\mathbf{W}_{\mathrm{ss},1} = \mathcal{Q}_{\mathrm{ind}}(\mathbf{V}_{\mathrm{ss},1}) = \arg\min_{\mathbf{C}\in\mathcal{C}} d_g(\mathbf{C}, \mathbf{V}_{\mathrm{ss},1}) \ . \quad (32)$$

The unitary rotation not seen by this Grassmannian codeword selection can be found by performing the polar decomposition[1] $\mathbf{V}_{\mathrm{ss},1}^H\mathbf{W}_{\mathrm{ss},1} = \mathbf{R}\mathbf{P}$, where $\mathbf{R} \in \mathcal{U}_{n_s}$, and $\mathbf{P}$ is a positive-semidefinite Hermitian matrix. The channels from the other BSs, with the rotation $\mathbf{R}$ taken into account, are then quantized using the Stiefel distance:

$$\mathbf{W}_{\mathrm{ss},k} = \mathcal{Q}_{\mathrm{stief}}(\mathbf{V}_{\mathrm{ss},k}) = \arg\min_{\mathbf{C}\in\mathcal{C}} d_s(\mathbf{C}, \mathbf{V}_{\mathrm{ss},k}\mathbf{R}). \quad (33)$$

To clarify the proposed codeword selection, one can write the product codeword in terms of Stiefel component codewords $\mathbf{C}_{\mathrm{pr}} = n_{bs}^{-1/2}[\mathbf{C}_1^H, \ldots, \mathbf{C}_{n_{bs}}^H]^H$, where $\mathbf{C}_k \in \mathcal{C}$. Given the polar decomposition of $\mathbf{V}_{\mathrm{ss},1}^H\mathbf{C}_1 = \mathbf{R}\mathbf{P}$, define the polar decompositions $(\mathbf{V}_{\mathrm{ss},k}\mathbf{R})^H\mathbf{C}_k = \mathbf{U}_k\mathbf{P}_k$ for any $k = 2, \ldots, n_{bs}$. Joint codeword selection is done by minimizing the Grassmannian chordal distance $d_g(\mathbf{C}_{\mathrm{pr}}, \mathbf{V}_{\mathrm{ss}})$, equivalent to maximizing $\|\mathbf{V}_{\mathrm{ss}}^H\mathbf{C}_{\mathrm{pr}}\|_F$ over the possible $\mathbf{C}_{\mathrm{pr}} \in \mathcal{C}_{\mathrm{pr}}$. In terms of component codewords, we have to maximize

$$\|\sum_{k=1}^{n_{bs}}\mathbf{V}_{\mathrm{ss},k}^H\mathbf{C}_k\|_F = \|\mathbf{R}\mathbf{P} + \sum_{k=2}^{n_{bs}}\mathbf{V}_{\mathrm{ss},k}^H\mathbf{C}_k\|_F \quad (34)$$

$$= \|\mathbf{P} + \sum_{k=2}^{n_{bs}}(\mathbf{V}_{\mathrm{ss},k}\mathbf{R})^H\mathbf{C}_k\|_F = \|\mathbf{P} + \sum_{k=2}^{n_{bs}}\mathbf{U}_k\mathbf{P}_k\|_F. \quad (35)$$

By minimizing the Grassmann chordal distance between the first per-cell channel component and the per-cell codeword $d_g(\mathbf{C}_1, \mathbf{V}_{\mathrm{ss},1})$, this corresponds to maximizing $\|\mathbf{V}_{\mathrm{ss},1}^H\mathbf{C}_1\|_F = \|\mathbf{P}\|_F$, i.e. the norm of the first element of the sum of (35). Intuitively, to maximize the norm of the sum of complex matrices (35), we have to do two things. First, we have to maximize each $\|\mathbf{P}_k\|_F$, then we have to make each $\mathbf{U}_k\mathbf{P}_k$ collinear with $\mathbf{P}$, i.e. get the $\mathbf{U}_k$ as close as possible to the identity $\mathbf{I}$. This is what (33) is producing by minimizing the Stiefel distance $d_s(\mathbf{C}_k, \mathbf{V}_{\mathrm{ss},k}\mathbf{R})$, which corresponds to maximizing the inner product $\langle\mathbf{C}_k, \mathbf{V}_{\mathrm{ss},k}\mathbf{R}\rangle = \Re(\mathrm{Tr}[(\mathbf{V}_{\mathrm{ss},k}\mathbf{R})^H\mathbf{C}_k]) = \Re(\mathrm{Tr}[\mathbf{U}_k\mathbf{P}_k])$. As a consequence of Von Neumann's trace inequality [34], this is bounded above by $\langle\mathbf{C}_k, \mathbf{V}_{\mathrm{ss},k}\mathbf{R}\rangle \leq \mathrm{Tr}[\mathbf{P}_k]$, with equality if and only if $\mathbf{U}_k = \mathbf{I}$.

e) Serial Codeword Selection: This method borrows the main idea of serial selection from [16], adapted here to perform codeword selection with a transformed codebook in a sequential manner. Without loss of generality, assume that the channels are sorted so that $\alpha_1 \geq \ldots \geq \alpha_{n_{bs}}$. The strongest channel (with the largest α_i) is first quantized as in (32). Then the per-cell components are selected sequentially. If the first $(k-1)$ per-cell codewords have been selected, the k^{th} codeword is

$$\mathbf{W}_{\mathrm{ss},k} = \arg\min_{\mathbf{C}\in\mathcal{C}} d_g(\mathbf{C}_{1\to k}, \mathbf{V}_{\mathrm{ls},1\to k}) \quad (36)$$

where $\mathbf{C}_{1\to k} = \left[\alpha_1\mathbf{W}_{\mathrm{ss},1}^H, \ldots, \alpha_{k-1}\mathbf{W}_{\mathrm{ss},k-1}^H, \alpha_k\mathbf{C}^H\right]^H$ is a concatenation of the previous chosen per-cell codewords with the k^{th} trial codeword taking into account the large-scale channel components, and $\mathbf{V}_{\mathrm{ls},1\to k} = \left[\mathbf{I}_{kn_t}\mathbf{0}_{kn_t,(n_{bs}-k)n_t}\right]\mathbf{V}_{\mathrm{ls}}$ is the $kn_t \times n_s$ upper sub-matrix of $\mathbf{V}_{\mathrm{ls}}$.

VII. Simulations

In this section, the codebook designs and codeword selections discussed are compared through numerical evaluation of the spectral efficiency (12).

A. Comparison of Codebook Criteria

In Fig. 8, we compare the spectral efficiency with 2 cooperative BSs using joint Grassmann-Stiefel codebooks constructions described in Section V: low-distortion codebooks using Lloyd-type Algorithm 1, and max-min distance codebooks from Monte Carlo simulations.

The original per-cell Grassmannian codebooks are constrained-alphabet constructions available from the literature. Codewords are selected using joint selection. We consider equal large scale path loss for each channel which corresponds to the scenario where the MS is at the cell edge, or a scenario where a product codebook is used

[1]Or equivalently the left and right polar decompositions of the square matrix $\mathbf{W}_{\mathrm{ss},1}^H\mathbf{V}_{\mathrm{ss},1} = \mathbf{P}\mathbf{R}^H = \mathbf{R}^H\mathbf{P}'$.

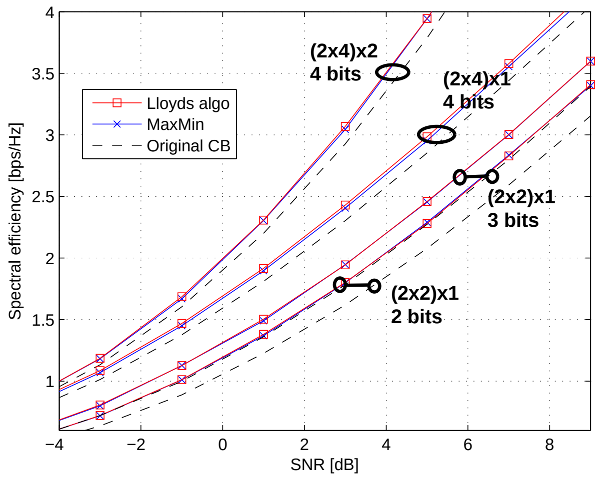


Fig. 8. Spectral efficiency of 4×1, 8×1 and 8×2 systems using 2×1, 4×1 and 4×2 codebook, respectively. This corresponds to two cooperative BSs at cell edge. Original per-cell Grassmann codebooks are low-complexity codebooks from [26], [29], [32], improved codebooks are according to Section V.

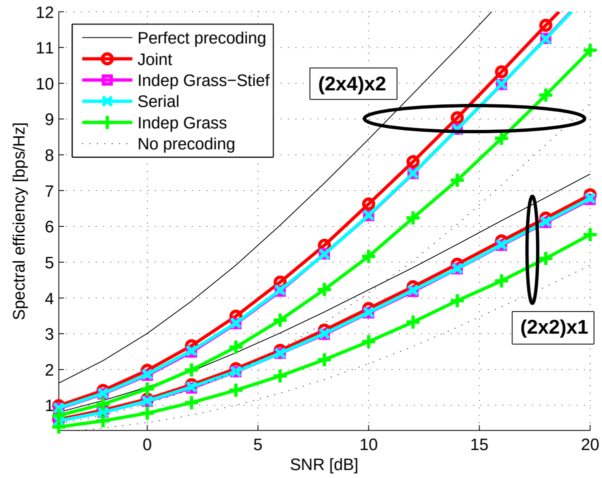


Fig. 9. Performance of codewords selection methods. Spectral efficiency of 4×1 and 8×2 systems using 2×1 and 4×2 codebooks respectively, correponding to two cooperative BS serving a user at cell edge. Codebooks with one feedback bit per transmit antenna.

for quantizing point-to-point arrays twice larger than the component codebook.

Simulations are done for a 2-transmit and 1-receive antenna system with the 2-bit Square Codebook given in Table I, and the 3-bit Square Antiprism Codebook from [26]; for a 4-transmit and 1-receive antenna system with the 4-bit codebook from [32]; and for a 4-transmit and 2-receive antenna system with the 4-bit C-Codebook from [29].

The figure shows that the max-min distance and low-distortion criteria lead to a similar performance. Maximum average distance codebooks from Monte Carlo simulations also give similar performance. These have been left out from the figure for clarity. The marginal gain between an original codebook and its Stiefel-improved version depends on which Stiefel representative is initially used for the Grassmann codebook. The spectral efficiency of the 2-bit improved codebook reaches the performance of the non-improved 3-bit codebook, giving a gain of 1 bit per BS. For 2-bit component codebooks, the best product codebook performance can be reached using the improved versions with finite alphabet shown in Table I and Appendix B. Even if the Square Codebook is not an optimal Grassmannian codebook, its Stiefel-improved version reaches the same performance than as the Stiefel-improved version of the Grassmann-optimal Tetrahedron Codebook given in Appendix B. The performance of Tetrahedron Codebook can be found in [35] and in Fig. 2.

B. Comparison Between Selection Methods

Figure 9 depicts the spectral efficiency of the proposed schemes for 2 cooperative BSs with 2 Tx antennas and the 2-bit Stiefel-improved Square Codebook of Table I; and with 4 Tx antennas serving a 2 Rx antenna user with 4-bit Stiefel-improved version by Algorithm 1 of the C-Codebook from [29]. Also here, we consider equal large scale path loss for each channel.

Using Stiefel distance consequently improves performance of independent selection. Serial selection offers slightly better performance than independent Grassmannian-Stiefel selection. Both independent Grassmannian-Stiefel selection and serial selection perform close to joint selection. In [35], simulations showing similar trends for 2 and 3 cooperative BSs with 2-bit Tetrahedron Codebook [26], and for 2 BSs with 3-bit Square Antiprism Codebook [26] can be found. Stiefel-improved codebooks lead to better performance compared to original codebooks for all codeword selection methods except independent selection with Grassmann distance.

C. Comparison for Large Scale Path Gain Imbalance

Figure 10 depicts the variation of performance when the large scale path gain for the first and the second BS are different. The lower curves represent two BSs with 2 Tx antenna each, and the upper curves represent two 4-antenna BSs. Same codebooks as in Fig. 9 are used. The graph can be interpreted as the performance depending of the position of the user, from the center of the cell to the cell edge.

The graph shows that the performance gap between the joint selection and independent/serial selection first reduces and then is reversed when the large scale path gain imbalance grows. With imbalance, joint selection is not optimal anymore as it quantizes the channel components with equal weight. Transforming the codebook according to (30) mitigates this problem, and gives the overall best selection method: matching the performance of independent selections for large imbalance and joint selection for no imbalance. When the user is at the center of the first cell $\alpha_2/\alpha_1 \approx 0$, independent selection outperforms joint selection without codebook transformation, agreeing with the results of [14]. Intuitively, in the limiting case $\alpha_2 = 0$ when there is no cooperation and only a single BS is transmitting, quantizing only the first channel is enough, on the other hand joint transmission balances this quantization with equal weight by a quantization of a zero-gain second channel. Also, at this point, the performance of

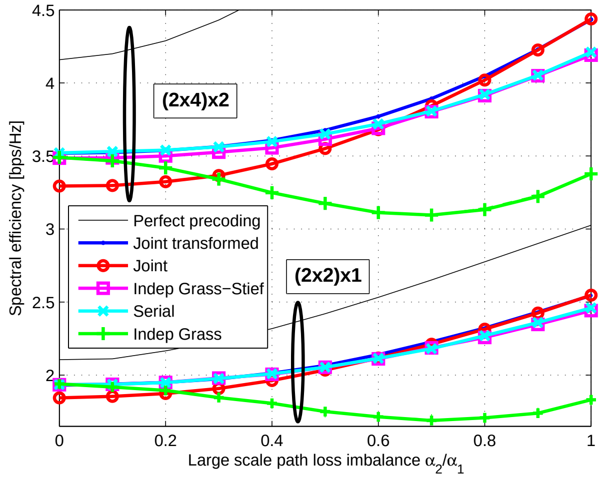


Fig. 10. Spectral efficiency of 4×1 and 8×2 systems using 2×1 and 4×2 codebooks respectively, as a function of large scale path gains imbalance. The strongest channel is fixed at a SNR of 6 dB. Codebooks with one feedback bit per transmit antenna.

the independent selection methods merge which is justified since the signal of the second BS vanishes.

VIII. Conclusion

We have considered product codebook quantization where codewords from a single small point-to-point codebook are concatenated to quantize larger MIMO channels, e.g. channels from cooperative BSs. We have proposed a joint Grassmann-Stiefel codebook design to remove the performance gap between product codebook quantization and global Grassmannian quantization. We have investigated methods to construct good Stiefel codebooks conditioned on Grassmannian codebooks. A Lloyd-type algorithm on Stiefel manifold conditioned on a given Grassmannian codebook is proposed, as well as some closed-form examples of joint Grassmann-Stiefel codebooks. For the Lloyd-type algorithm, a closed-form solution to the centroid problem on the Stiefel manifold was provided. Finally, we discussed low-complexity codeword selection methods showing good performance.

Appendix

A. Illustrative Codebooks of Fig. 4

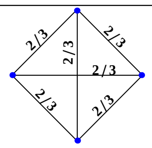
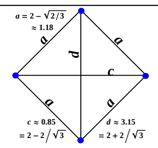
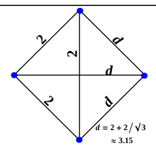
The following table gives the two alternative real Stiefel codebooks in $\mathcal{V}^{\mathbb{R}}_{3,1} \cong S^2$, both generating the same optimum 2-bit Grassmannian packing in $\mathcal{G}^{\mathbb{R}}_{3,1}$, which is a set of 4 antipodal points forming a cube. The second alternative gives a tetrahedron configuration which is an optimum packing in the Stiefel manifold $\mathcal{V}^{\mathbb{R}}_{3,1} \cong S^2$.

Square	$\frac{1}{\sqrt{3}}$	$\left\{ \begin{bmatrix}1\\1\\1\end{bmatrix} \begin{bmatrix}1\\-1\\1\end{bmatrix} \begin{bmatrix}-1\\1\\1\end{bmatrix} \begin{bmatrix}-1\\-1\\1\end{bmatrix} \right\}$
Tetrahedron	$\frac{1}{\sqrt{3}}$	$\left\{ \begin{bmatrix}1\\1\\1\end{bmatrix} \begin{bmatrix}-1\\1\\-1\end{bmatrix} \begin{bmatrix}1\\-1\\-1\end{bmatrix} \begin{bmatrix}-1\\-1\\1\end{bmatrix} \right\}$

B. Tetrahedron Codebook in $\mathcal{G}^{\mathbb{C}}_{2,1}$

In the table below the *Tetrahedron Codebook* from [26] and an Stiefel-improved version of it are presented. The Tetrahedron Codebook is an optimum Grassmannian packing, and also an optimum Thomson configuration for several values of p. Comparison to Monte Carlo simulations suggests that the proposed Stiefel version below is the best Stiefel codebook conditioned on the Tetrahedron Codebook according to many metrics. It maximizes the Stiefel minimum distance and p-mean distance for several values of p, e.g. $p = -1, -2, 1, 2$. Nevertheless, this is not an optimum unconditioned Stiefel codebook. The optimum Stiefel codebook is a simplex with squared distance $\frac{8}{3} \approx 2.67$, while the conditioned codebook achieves a squared minimum Stiefel distance of 2. The codebook is written in term of the following constants: $\alpha_\pm = \sqrt{\frac{1}{6}(3 \pm \sqrt{3})}$ and $\gamma_\pm = e^{\pm i \frac{\pi}{4}} = \frac{1 \pm i}{\sqrt{2}}$.

Codebook in [26]	$\left\{ \begin{bmatrix}\alpha_+\\\alpha_-\end{bmatrix} \begin{bmatrix}\alpha_+\\-\alpha_-\end{bmatrix} \begin{bmatrix}\alpha_-\\i\alpha_+\end{bmatrix} \begin{bmatrix}\alpha_-\\-i\alpha_+\end{bmatrix} \right\}$
Grass-Stief CB	$\left\{ \begin{bmatrix}\alpha_+\\\alpha_-\end{bmatrix} \begin{bmatrix}-\alpha_+\\\alpha_-\end{bmatrix} \begin{bmatrix}\gamma_+\alpha_-\\-\gamma_-\alpha_+\end{bmatrix} \begin{bmatrix}\gamma_-\alpha_-\\-\gamma_+\alpha_+\end{bmatrix} \right\}$

<table>
<tr><th rowspan="2">Squared Grass. dist.</th><th colspan="2">Squared Stief. dist.</th></tr>
<tr><th>CB in [26]</th><th>Grass-Stief CB</th></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

References

[1] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341 –1365, Oct. 2008.

[2] W. Dai, Y. Liu, B. Rider, and V. Lau, "On the information rate of MIMO systems with finite rate channel state feedback using beamforming and power on/off strategy," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5032–5047, Nov. 2009.

[3] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[4] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[5] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[6] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wirel. Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[7] Y. Cheng, V. Lau, and Y. Long, "A scalable limited feedback design for network MIMO using per-cell product codebook," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3093–3099, October 2010.

[8] C. Yang, S. Han, X. Hou, and A. Molisch, "How do we design CoMP to achieve its promised potential?" *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 67–74, Feb. 2013.

[9] F. Yuan and C. Yang, "Bit allocation between per-cell codebook and phase ambiguity quantization for limited feedback coordinated multi-point transmission systems," *IEEE Trans. on Commun.*, vol. 60, no. 9, pp. 2546–2559, Sep. 2012.

[10] T. Shuang, T. Koivisto, H.-L. Maattanen, K. Pietikainen, T. Roman, and M. Enescu, "Design and evaluation of LTE-Advanced double codebook," in *Proc. IEEE Veh. Technol. Conf.*, May 2011, pp. 1–5.

[11] Y. Zhang and R. S. Cheng, "On the design of interference alignment scheme for multi-user MIMO with limited feedback," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 1–5.

[12] Q. Du, M. D. Gunzburger, and L. Ju, "Constrained centroidal voronoi tessellations for surfaces," *SIAM J. Sci. Comput.*, vol. 24, no. 5, pp. 1488–1506, 2003.

[13] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: a practical limited feedback technique for massive MIMO systems," *submitted to IEEE Trans. Commun. available at http://arxiv.org/abs/1305.4976*, May 2013.

[14] D. Su, X. Hou, and C. Yang, "Quantization based on per-cell codebook in cooperative multi-cell systems," in *Proc. IEEE Wireless Commun. & Net. Conf.*, Mar. 2011, pp. 1753–1758.

[15] F. Yuan and C. Yang, "Phase ambiguity quantization for per-cell codebook based limited feedback coordinated multi-point transmission systems," in *Proc. IEEE Veh. Technol. Conf.*, May 2011, pp. 1–5.

[16] X. Hou and C. Yang, "Codebook design and selection for multi-cell cooperative transmission limited feedback systems," in *Proc. IEEE Veh. Technol. Conf.*, May 2011, pp. 1–5.

[17] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian space," *Exp. Math*, vol. 5, pp. 139–159, 1996.

[18] N. J. Higham, "Matrix nearness problems and applications," in *Applications of Matrix Theory, Oxford University Press*, 1989, p. 127.

[19] K. Fan and A. J. Hoffman, "Some metric inequalities in the space of matrices," *Proc. Amer. Math. Soc.*, vol. 6, no. 1, pp. 111–116, Feb. 1955.

[20] D. Love and J. Heath, R.W., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, Mar. 2006.

[21] P. Xia and G. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[22] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, Jun. 1964.

[23] J. Roh and B. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[24] ——, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[25] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1108–1123, Mar. 2008.

[26] R.-A. Pitaval, H.-L. Maattanen, K. Schober, O. Tirkkonen, and R. Wichman, "Beamforming codebooks for two transmit antenna systems based on optimum Grassmannian packings," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6591–6602, Oct. 2011.

[27] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Statist.*, vol. 25, pp. 40–75, 1954.

[28] B. Mondal, S. Dutta, and R. Heath, "Quantization on the Grassmann manifold," *IEEE Trans. Signal Proc.*, vol. 55, no. 8, pp. 4208–4216, Aug. 2007.

[29] R.-A. Pitaval, O. Tirkkonen, and S. Blostein, "Low complexity MIMO precoding codebooks from orthoplex packings," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–5.

[30] T. Inoue and R. Heath, "Kerdock codes for limited feedback precoded MIMO systems," *Trans. Signal Process.*, vol. 57, no. 9, pp. 3711–3716, Sep. 2009.

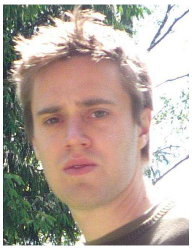
[31] B. Mondal, T. Thomas, and M. Harrison, "Rank-independent codebook design from a quaternary alphabet," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2007, pp. 297–301.

[32] P. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-Advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, no. 1, p. 302092, 2009.

[33] R. Wichman and A. Hottinen, "Transmit diversity in the WCDMA system," *Int. J. Wireless Inf. Networks*, vol. 6, pp. 171–180, 1999.

[34] D. Rhea, "The case of equality in the Von Neumann trace inequality," preprint (2011), available at http://www.drhea.net/wp-content/uploads/2011/01/vonNeumann.pdf.

[35] R.-A. Pitaval and O. Tirkkonen, "Incorporating Stiefel geometry in codebook design and selection for improved base station cooperation," in *Proc. IEEE Veh. Technol. Conf.*, May 2012, pp. 1–5.



Renaud–Alexandre Pitaval received the "Diplôme d'Ingénieur" in electrical engineering from the Grenoble Institute of Technology in 2008 and the M.Sc. degree in communications engineering from Helsinki University of Technology in 2009. From 2009 to 2010, He was a research assistant at Queen's University, Canada, and spent a research visit at Imperial College London in spring 2013. He is currently completing his D.Sc. at the Department of Communications and Networking, Aalto University.

Olav Tirkkonen is associate professor in communication theory at the Department of Communications and Networking in Aalto University, Finland, where he has held a faculty position since August 2006. He received his M.Sc. and Ph.D. degrees in theoretical physics from Helsinki University of Technology in 1990 and 1994, respectively. Between 1994 and 1999 he held post-doctoral positions at the University of British Columbia, Vancouver, Canada, and the Nordic Institute for Theoretical Physics, Copenhagen, Denmark. From 1999 to 2010 he was with Nokia Research Center (NRC), Helsinki, Finland, most recently acting as Research Fellow. He has published some 150 papers, and is coauthor of the book "Multiantenna transceiver techniques for 3G and beyond". His current research interests are in multiantenna techniques, cognitive and heterogeneous cellular systems, and self-organization of wireless networks.