Codebooks in Flag Manifolds for Limited Feedback MIMO Precoding

Renaud-Alexandre Pitaval, Ashvin Srinivasan and Olav Tirkkonen
Aalto University, Department of Communications and Networking, Espoo, Finland

***Abstract*—We consider unitary codebook design problems for various precoded MIMO scenarios, and interprete them as generalized discretization problems of flag manifolds. As a concrete example, we consider codebooks for MIMO transmission with linear receivers. In this case, the problem reduces to discretizing permutation-invariant flag manifolds. A corresponding Lloyd algorithm is given, providing low-distortion codebooks. The achievable rate of the system applying the generated codebooks is compared by simulations. We find that for a N-by-N full-rank MIMO system with $N > 2$ and a linear receiver, the gain of precoding with a small number of feedback bits is small. This differs from the behavior of low-rank transmissions, where it is known that a small number of feedback bits allows near-optimal channel adaptation.**

I. Introduction

Multiple-Input Multiple-Output (MIMO) wireless communications systems using linear precoding have been shown to achieve large capacity gains over traditional single-input-single-output (SISO) systems [1]. Full gains from precoding are achieved when the transmitter possesses perfect channel state information (CSI). In frequency-division duplex systems, the only way to acquire 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. It is usually considered more crucial to convey information about the channel direction than the channel beam gains and a unitary codebook is typically employed for this.

In point-to-point communications with maximum likelihood (ML) receiver, the performance of a unitary precoding codebook depends of the distance properties of the Grassmannian planes generated by the codebook. This has led to the well known Grassmannian codebook design [2] where the codebook is understood as a discretization of the Grassmann manifold. While Grassmann precoding has attracted much attention, other transmission scenarios or constraints may lead to designing codebooks in other spaces. The spaces of discretization to consider are quotient spaces of unitary groups, and more specifically so-called *flag manifolds*. Grassmann and Stiefel manifolds are examples of flag manifolds. Literature on flag manifolds can be found in quantum theory context, e.g. in [3]–[5].

In this paper, after describing several codebook designs from the literature where the space to discretize can be seen as a flag manifold, we concentrate more in detail on precoding for MIMO system with a linear receiver, such as a zero-forcing (ZF) or minimum mean square (MMSE) receiver. With a linear receiver, the Grassmannian precoding design used for ML receivers is not appropriate any more [6]. The spaces of interest are simple permutation-invariant flag manifolds. We describe a Lloyd algorithm for those spaces and compare the achievable rate with the generated codebooks by simulation. Of specific interest is the case when the number of streams, and the number of receive and transmit antennas are the same. In this set up, the corresponding Grassmannian collapses to a single point making Grassmannian precoding irrelevant. Accordingly, precoding does not improve the information rate with ML receiver. On the other hand, with linear receiver, gain may be obtained from precoding using the proposed flag codebook design. The simulations show that this gain is relatively small when only few feedback bits are used for more than two transmit antennas. This differs from the behavior of low-rank transmission, where it is known that a small number of feedback bits can allow near optimal channel adaptation.

This work has been supported in part by the Academy of Finland, and Tekes – the Finnish Funding Agency for Technology and Innovation

II. Flag Manifolds

The codebooks addressed have orthonormal columns, and thus consist of a number of columns from a unitary matrix.

A. Unitary Group and Stiefel Manifold

The unitary group, the space of all n-dimensional unitary matrices, is denoted by

$$\mathcal{U}_n = \left\{ \mathbf{U} \in \mathbb{C}^{n\times n} \quad | \quad \mathbf{U}^H\mathbf{U} = \mathbf{I}_n \right\}.$$

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

$$\mathcal{V}_{n,p}^{\mathbb{C}} = \left\{ \mathbf{Y} \in \mathbb{C}^{n\times p} \quad | \quad \mathbf{Y}^H\mathbf{Y} = \mathbf{I}_p \right\}. \tag{1}$$

B. Grassmann Manifold

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

$$[\mathbf{Y}] = \{\mathbf{YU} \quad | \quad \mathbf{U} \in \mathcal{U}_p\}. \tag{2}$$

where the Stiefel-matrix $\mathbf{Y}$ is an $n\times p$ matrix with orthonormal columns. The equivalence class $[\mathbf{Y}]$ thus represents a set of matrices $\mathbf{Y}$ determining the same subspace. Taking two Grassmannian points $[\mathbf{X}]$, $[\mathbf{Y}] \in \mathcal{G}_{n,p}^{\mathbb{C}}$, the chordal distance between these is defined as [7]

$$d_c([\mathbf{X}],[\mathbf{Y}]) = \frac{1}{\sqrt{2}}\|\mathbf{X}\mathbf{X}^H - \mathbf{Y}\mathbf{Y}^H\|_F \ . \tag{3}$$

This distance does not depend on the chosen representative in $[\mathbf{X}]$ and $[\mathbf{Y}]$.

C. Generalized Flag Manifolds

A *flag* in $\mathbb{C}^n$ is a sequence of nested subspaces $V_1 \subset \cdots \subset V_r \subset \mathbb{C}^n$ with $1 \le r \le n$. To every flag $(V_1, \cdots, V_r)$ we can associate the subspace decomposition $(W_1, \cdots, W_r)$ satisfying $V_i = \bigoplus_{j=1}^{i} W_j$. Given $(p_1, \ldots, p_r) \in \mathbb{N}^r$, we define[1] the flag manifold $\mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}}$ as the set of all $W_1 \oplus \cdots \oplus W_r$ in $\mathbb{C}^n$ such that $\dim W_i = p_i$.

Let $p = \sum_{i=1}^r p_i$. We represent a point on $\mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}}$ by a n-by-p Stiefel matrix $\mathbf{W}$, i.e. $\mathbf{W}^H\mathbf{W} = \mathbf{I}_p$. Under this representation, two matrices $\mathbf{W}_1, \mathbf{W}_2$ are equivalent in the sense that they represent the same flag if matrices $(\mathbf{U}_1, \ldots, \mathbf{U}_r) \in \mathcal{U}_{p_1} \times \cdots \times \mathcal{U}_{p_r}$ exist such that $\mathbf{W}_1 = \mathbf{W}_2 \operatorname{diag}(\mathbf{U}_1, \ldots, \mathbf{U}_r)$. The flag manifold can thus be expressed as the quotient space

$$\begin{aligned}\mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}} &\cong \frac{\mathcal{V}_{n,p}^{\mathbb{C}}}{\mathcal{U}_{p_1} \times \cdots \times \mathcal{U}_{p_r}} \\ &\cong \frac{\mathcal{U}_n}{\mathcal{U}_{p_1} \times \cdots \times \mathcal{U}_{p_r} \times \mathcal{U}_{n-p}},\end{aligned} \tag{4}$$

where $p = \sum_{i=1}^r p_i$. We note that $\mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}} \cong \mathcal{F}l_{n;p_1,\ldots,p_r,n-p}^{\mathbb{C}}$. The Stiefel manifold $\mathcal{V}_{n,p}^{\mathbb{C}}$ can be seen as an extreme case in (4) by setting $r = 1$ and $p_1 = 0$. For $r = 1$ and $p_1 = p$, one recovers the Grassmann manifold $\mathcal{F}l_{n;p}^{\mathbb{C}} \cong \mathcal{G}_{n,p}^{\mathbb{C}}$.

Given two flags $[\mathbf{W}_1], [\mathbf{W}_2] \in \mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}}$, represented by $\mathbf{W}_1, \mathbf{W}_2 \in \mathcal{V}_{n,p}^{\mathbb{C}}$, they can be decomposed as $\mathbf{W}_k = (\mathbf{W}_{k,1}, \mathbf{W}_{k,2}, \ldots, \mathbf{W}_{k,r})$, such that $\mathbf{W}_{k,i} \in \mathcal{V}_{n,p_i}^{\mathbb{C}}$. The notion of Grassmannian chordal distance can be generalized to flag manifolds by

$$d_f([\mathbf{W}_1],[\mathbf{W}_2]) = \sqrt{\sum_{i=1}^r d_c^2([\mathbf{W}_{1,i}],[\mathbf{W}_{2,i}])} \ . \tag{5}$$

This distance naturally arises from the canonical embedding $\mathcal{F}l_{n;p_1,\ldots,p_r}^{\mathbb{C}} \hookrightarrow \mathcal{G}_{n,p_1}^{\mathbb{C}} \times \cdots \times \mathcal{G}_{n,p_r}^{\mathbb{C}}$ and taking the chordal distance on this space.

D. Simple Flag Manifolds

Of specific interest is the case when $p_1 = \ldots = p_r = 1$, for which we use the notation

$$\mathcal{F}_{n,p}^{\mathbb{C}} \triangleq \mathcal{F}l_{n;\underbrace{1,\ldots,1}_{p}}^{\mathbb{C}} \cong \frac{\mathcal{V}_{n,p}^{\mathbb{C}}}{(\mathcal{U}_1)^p} \cong \frac{\mathcal{U}_n}{(\mathcal{U}_1)^p \times \mathcal{U}_{n-p}} \ , \tag{6}$$

and the case $p = n$, $\mathcal{F}_{n,n}^{\mathbb{C}} \cong \mathcal{U}_n/(\mathcal{U}_1)^n$.

[1]This is isomorphic to the definition where $\mathcal{F}l_{n;d_1,\ldots,d_r}^{\mathbb{C}}$ is the set of all flags $(V_1, \cdots, V_r)$ in $\mathbb{C}^n$ with $V_1 \subset \cdots \subset V_r \subset \mathbb{C}^n$ and dim $V_i = d_i$.

E. Simple Permutation-Invariant Flag Manifolds

We now add an additional equivalence relation on $\mathcal{F}_{n,p}^{\mathbb{C}}$ by considering that two representatives $\mathbf{W}_1, \mathbf{W}_2 \in \mathcal{V}_{n,p}^{\mathbb{C}}$ are equivalent under column permutation, i.e. if there exists a permuation matrix $\mathbf{P} \in \mathbb{R}^{p\times p}$ such that $\mathbf{W}_1 = \mathbf{W}_2\mathbf{P}$, then $[\mathbf{W}_1] = [\mathbf{W}_2]$. The permutation corresponds to an orientation-invariance of the elements. We denote the corresponding space by

$$\vec{\mathcal{F}}_{n,p}^{\mathbb{C}} \triangleq \mathcal{F}_{n,p}^{\mathbb{C}}/S_p \tag{7}$$

where S_p is the symmetric group whose elements are all the permutations of the p symbols. Given two elements $[\mathbf{W}_1], [\mathbf{W}_2] \in \vec{\mathcal{F}}_{n,p}^{\mathbb{C}}$, we define their distance as

$$d_p([\mathbf{W}_1],[\mathbf{W}_2]) = \min_{\mathbf{P}\in S_p} d_f([\mathbf{W}_1],[\mathbf{W}_2\mathbf{P}]). \tag{8}$$

F. The Specific Cases of $\mathcal{F}_{2,2}^{\mathbb{C}}$ and $\vec{\mathcal{F}}_{2,2}^{\mathbb{C}}$

Now consider $n = 2$. We have $\mathcal{F}_{2,2}^{\mathbb{C}} \cong \mathcal{F}_{2,1}^{\mathbb{C}} \cong \mathcal{G}_{2,1}^{\mathbb{C}}$, which further reduces to the real unit sphere $\mathcal{F}_{2,2}^{\mathbb{C}} \cong S^2$. It follows that designing codebooks in $\mathcal{F}_{2,2}^{\mathbb{C}}$ is equivalent to designing spherical codes [8].

In addition, each columm of a 2×2 unitary matrix generating a point in $\mathcal{F}_{2,2}^{\mathbb{C}}$ can be seen as two ordered antipodal points. It follows that $\vec{\mathcal{F}}_{2,2}^{\mathbb{C}}$ is the set of spherical antipodal points, or equivalently the set of lines in 3D, also known as the real Grassmannian $\mathcal{G}_{3,1}^{\mathbb{R}} \cong \vec{\mathcal{F}}_{2,2}^{\mathbb{C}}$ [7]. Codebooks in $\vec{\mathcal{F}}_{2,2}^{\mathbb{C}}$ can thus be constructed by leveraging results from known antipodal spherical codes [6].

III. Examples of Application of Flag Codebooks for Limited Feedback MIMO Precoding

A general MIMO signal model is of the form $\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{n}$. The precoding matrix $\mathbf{W} \in \mathbb{C}^{n_t\times n_s}$ is designed for channel adaptation according to the information fed back by the receiver. To acquire CSI through a limited feedback channel, a widely applied method is to use codebook-based precoding where the receiver and transmitter share a predefined codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_b}\}$. The receiver selects a codeword following some quantization rule that approximates the channel by using the codebook, $\mathrm{q}_{\mathcal{C}} : \{\mathbf{H}\} \to \{i : 1 \le i \le n_b\}$, and feeds back the index $k = \mathrm{q}_{\mathcal{C}}(\mathbf{H})$ to the transmitter. The transmitter constructs a precoding matrix based on the CSI received, $\mathbf{W} = f(k)$. Especially for point-to-point MIMO, a simplified version is to assume that the transmitter directly picks the precoding matrix from the shared codebook: $\mathbf{W} = \mathbf{C}_k$.

Furthermore, it is often assumed that $\mathcal{C}$ is a codebook of Stiefel matrices, i.e. $\mathbf{C}_i^H\mathbf{C}_i = \mathbf{I}\ \forall i$, which are used to quantize the eigendirections of the signal subspace at the receiver. In this setting, there is typically an infinity of precoding vectors leading to the same performance, and the possible precoding vectors can be grouped into equivalence classes. It follows that the set of equivalence classes of precoding matrices is a quotient space of the Stiefel manifold $\mathcal{V}_{n_t,n_s}^{\mathbb{C}}/\!\sim$, where $\sim$ is the equivalence relation defining equivalence of two precoding matrices. In order to have non-equivalent codewords, the codebook should be designed as a discretization of this space of equivalence classes of precoding matrices. A Stiefel precoding codebook is then generated by taking any representative in the equivalence classes.

Several codebook design problems for MIMO precoding can be interpreted as discretization of flag manifolds:

MIMO with ML receiver: The codebook design for this problem is well investigated and known to reduce to a discretization of the Grassmann manifold $\mathcal{G}^{\mathbb{C}}_{n_t,n_s}$ [2].

MIMO with linear receiver: With MMSE or ZF receiver, the Grassmann manifold does not correspond anymore to the set of non-equivalent precoding matrices [6], and the right space to discretize is the permutation invariant flag manifold $\vec{\mathcal{F}}^{\mathbb{C}}_{n_t,n_s}$. We describe this problem in more detail in Section V.

MIMO with bit-interleaved coded precoding: For bit-interleaved multiple beamforming, it is shown in [9] that the optimum precoder is invariant under a diagonal unitary transform. A metric and a Lloyd algorithm are described in [9], which can be seen as a Lloyd algorithm on $\mathcal{F}^{\mathbb{C}}_{n_t,n_s}$.

MIMO broadcast channel: Unitary codebook-based precoding has been well-investigated for multi-user MIMO with one-stream transmission per-user [10]–[13]. In [11], codebook design on a Riemann manifold is considered. This manifold actually is $\vec{\mathcal{F}}^{\mathbb{C}}_{n_t,n_s}$, and the metric considered is proportional to d_p. Similarly, the equivalence described in [12], [13] amounts to design a codebook in $\mathcal{F}^{\mathbb{C}}_{n_t,n_s}$. Moreover, [12] concentrates on parameterizing the codeword family for the 2×2 case; as discussed above, $\mathcal{F}^{\mathbb{C}}_{2,2} \cong \mathcal{G}^{\mathbb{C}}_{2,1}$, $\vec{\mathcal{F}}^{\mathbb{C}}_{2,2} \cong \mathcal{G}^{\mathbb{R}}_{3,1}$ and designing codebooks in those spaces reduces to designing sperical codes [8] and antipodal spherical codes [6], respectively.

Network MIMO: With multi-point and multi-user transmission, a codebook design is naturally inherited from the MIMO broadcast channel design above. Therefore, flag manifolds are of interest for unitary precoding in network MIMO.

Nested LTE codebooks: Practical codebooks in industry standards have been designed by considering additional constraints [14]. The similarity between the nested property required of the 3GPP LTE codebooks and the definition of a flag as nested subspaces is worth noticing.

IV. Lloyd Algorithm for Flag Manifold $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$

In this section, we describe a Lloyd algorithm to generate codebooks in $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$. For clarity, we drop the brackets in the notation $[\mathbf{X}] \in \vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$ and simply write $\mathbf{X} \in \vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$. It is important to keep in mind that elements in $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$ are equivalence classes of matrices.

Given a codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_N\} \subset \vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$, and any $\mathbf{V} \in \vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$, define the quantization map

$$\mathrm{q}(\mathbf{V}) = \arg\min_{1\le i\le n_b} d_p(\mathbf{V}, \mathbf{C}_i)\ . \tag{9}$$

Given a random source V on $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$, the average distortion of the codebook $\mathcal{C}$ is

$$D(\mathcal{C}) = \mathbb{E}\left[d_p^2(V, \mathbf{C}_{\mathrm{q}(V)})\right]. \tag{10}$$

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

Nearest Neighbor rule (NN): Partitioning of $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$ according to the codebook in N Voronoi cells $\{V_1, \ldots, V_N\}$ defined by

$$V_k = \{\mathbf{V} \in \vec{\mathcal{F}}^{\mathbb{C}}_{n,p} \,|\, \mathrm{q}(\mathbf{V}) = k\}. \tag{11}$$

Centroid Computation (CC): Finding the centroids of each Voronoi cell V_k given by

$$\mathbf{M}_k = \arg\min_{\mathbf{M}\in V_k} \mathbb{E}\left[d_p^2(V, \mathbf{M}) \mid V \in V_k\right]. \tag{12}$$

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

In practice, the elements of $\vec{\mathcal{F}}^{\mathbb{C}}_{n,p}$ are represented by $n\times p$ matrices in $\mathcal{V}^{\mathbb{C}}_{n,p}$. We assume a uniformly distributed source, which can be represented by a uniformly distributed random variable on $\mathcal{V}^{\mathbb{C}}_{n,p}$. The detailed steps of the algorithm with Stiefel matrices are as follows:

1) Generate an initial codebook $\mathcal{C}_0$ of size N in $\mathcal{V}^{\mathbb{C}}_{n,p}$.
2) Generate a large training set $V = \{\mathbf{V}_i\}$ of uniformly distributed matrices in $\mathcal{V}^{\mathbb{C}}_{n,p}$, and quantize these using (9).
3) Permute the training set elements acoording to their quantization: let $\vec{\mathbf{V}}_i = \mathbf{V}_i\mathbf{P}_i$ where $\mathbf{P}_i$ is the permutation matrix minimizing $d_f(\mathbf{C}_{\mathrm{q}(\mathbf{V}_i)}, \mathbf{V}_i\mathbf{P}_i)$. Define the oriented training set by $\vec{V} = \{\vec{\mathbf{V}}_i\}$.
4) NN: Partition the training set $\vec{V}$ in N cells $\{V_1, \ldots, V_N\}$, defined by $V_k = \{\vec{\mathbf{V}}_i \in \vec{V} \,|\, \mathrm{q}(\vec{\mathbf{V}}_i) = k\}$.
5) CC: To generate the centroid of V_k, we use a suboptimal approach similar to [9]. As the distance used arises from the embedding $\mathcal{F}^{\mathbb{C}}_{n,p} \hookrightarrow (\mathcal{G}^{\mathbb{C}}_{n,1})^p$, we first compute the centroid $\mathbf{M}_k$ of the image of V_k in $(\mathcal{G}^{\mathbb{C}}_{n,1})^p$ and then approximate the true centroid by the Euclidean projection $\tilde{\mathbf{M}}_k$ of $\mathbf{M}_k$ onto $\mathcal{F}^{\mathbb{C}}_{n,p}$. Concretely this consists of computing the Grassmannian centroid of each column given by
$$\mathbf{m}_{k,l} = \text{principal eigenvector of} \sum_{\vec{\mathbf{V}}_i\in V_k} \mathbf{v}_{i,l}\mathbf{v}_{i,l}^H \tag{13}$$
where $\mathbf{v}_{i,l}$ is the lth column of $\vec{\mathbf{V}}_i$, and then taking the projection [15]
$$\tilde{\mathbf{M}}_k = \mathbf{L}_k\mathbf{R}_k^H \tag{14}$$
where $\mathbf{M} = \mathbf{L}_k\Sigma_k\mathbf{R}_k^H$ is a thin singular value decomposition of $\mathbf{M}_k = [\mathbf{m}_{k,1}, \ldots, \mathbf{m}_{k,p}]$.
6) A new codebook $\mathcal{C}_1 = \{\tilde{\mathbf{M}}_1, \ldots, \tilde{\mathbf{M}}_N\}$ is generated, with distortion $D(\mathcal{C}_1)$.
7) Repeat the steps by setting $\mathcal{C}_0 = \mathcal{C}_1$ until $|D(\mathcal{C}_1) - D(\mathcal{C}_0)| < \epsilon$, for a given precision ϵ.

While not shown in this paper due to lack of space, simulations reveal that the approximate centroid computation is very close to the optimal and the algorithm converges.

V. Application to MIMO Precoding with Linear Receiver

Consider a MIMO system with n_t transmit and n_r receive antennas. After unitary precoding with $\mathbf{W} \in \mathbb{C}^{n_t\times n_s}$, a vector $\mathbf{x} \in \mathbb{C}^{n_s\times 1}$ of $n_s \le \min(n_t, n_r)$ multiplexed streams is transmitted through a fading channel $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$. The received signal is $\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{n_r \times 1}$, where $\mathbf{n} \in \mathbb{C}^{n_r \times 1}$ is the noise. We assume that the transmitted signal and the noise are Gaussian with covariances $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \gamma \mathbf{I}_{n_s}$ and $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \mathbf{I}_{n_r}$, where γ is the per-stream SNR.

A. Achievable Information Rates

Let the singular values of $\mathbf{H}$ be $\sigma_1^{1/2} \geq \ldots \geq \sigma_{n_m}^{1/2}$ with $n_m = \min(n_r, n_t)$. Without water-filling and for a given transmission rank constaint n_s, the maximum achievable rate of the system is [1], [16]

$$I_{n_s} = \sum_{k=1}^{n_s} \log_2(1+\gamma\sigma_k) \leq \log_2 \det(\mathbf{I}+\gamma\mathbf{H}\mathbf{H}^H) = I_{n_m} \quad (15)$$

where I_{n_m} is the maximum achievable rate without transmission rank constraint.

For a fixed precoding vector $\mathbf{W} \in \mathbb{C}^{n_t \times n_s}$, the achievable rate depends of the receiver type. Denote the singular values of the equivalent channel $\mathbf{H}_{\mathrm{eq}} = \mathbf{H}\mathbf{W}$ by $\lambda_1^{1/2} \geq \ldots \geq \lambda_{n_s}^{1/2}$.

ML receiver: With maximum likelihood receiver the achievable rate with precoding $\mathbf{W}$ is

$$I_{\mathrm{ml}}(\mathbf{W}) = \log_2 \det(\mathbf{I} + \gamma \mathbf{H}_{\mathrm{eq}}\mathbf{H}_{\mathrm{eq}}^H) = \sum_{k=1}^{n_s} \log_2(1+\gamma\lambda_k). \quad (16)$$

We have $I_{\mathrm{ml}}(\mathbf{W}) \leq I_{n_s}$. The latter is achievable with $\mathbf{W}_{\mathrm{opt}} = \mathbf{V}_{n_s}$ where $\mathbf{V}_{n_s} \in \mathbb{C}^{n_t \times n_s}$ is a matrix composed by the right singular vectors of $\mathbf{H}$ corresponding to its n_s-largest singular values.

Linear receiver: The receiver employs a linear receiver of the form $\mathbf{G} = \mathbf{F}\mathbf{H}_{\mathrm{eq}}^H$, where $\mathbf{F} = (\gamma\mathbf{H}_{\mathrm{eq}}^H\mathbf{H}_{\mathrm{eq}} + a\mathbf{I}_{n_s})^{-1}$. With $a = 0, 1$, we get a ZF and MMSE receiver, respectively. The corresponding rate is [17]

$$I_{\mathrm{lr}}(\mathbf{W}) = \sum_{k=1}^{n_s} \log_2(1+\gamma_k) \ , \quad (17)$$

where $\gamma_k = (\mathbf{F}_{k,k})^{-1} - a$ is the post-processing SINR of the k-th data stream. In general we have $I_{\mathrm{lr}}(\mathbf{W}) \leq I_{\mathrm{ml}}(\mathbf{W})$. In [6] it is shown that there exists a unitary matrix $\mathbf{O} \in \mathcal{U}_{n_s}$ such that $I_{\mathrm{lr}}(\mathbf{W}\mathbf{O}) = I_{\mathrm{ml}}(\mathbf{W})$ and a precoder partitioning is proposed accordingly. As for ML receiver, an optimum precoding matrix is given by $\mathbf{W}_{\mathrm{opt}} = \mathbf{V}_{n_s}$.

We notably stressed the following difference between linear and ML receivers.

Remark 1: While for $n_s = n_t = n_r$, we have $I_{\mathrm{ml}}(\mathbf{W}) = I_{n_m}$ for any $\mathbf{W} \in \mathcal{U}_{n_t}$ and thus unitary precoding does not change the transmission rate, with a linear receiver, the rate is a function of the precoding matrix, and $I_{\mathrm{lr}}(\mathbf{W}) \leq I_{n_m}$.

B. Set of Precoding Matrix Equivalence classes

Let $\sim$ be the equivalence relation declaring two precoding matrices equivalent, $\mathbf{W}_1 \sim \mathbf{W}_2$, if and only if $I(\mathbf{W}_1) = I(\mathbf{W}_2)$. It follows that the set of equivalence classes of precoding matrices is of the form $\mathcal{V}_{n_t,n_s}^{\mathbb{C}}/\sim$.

ML receiver: The information rate is invariant under any right-unitary rotation of the precoding codeword: $I_{\mathrm{ml}}(\mathbf{W}\mathbf{U}) = I_{\mathrm{ml}}(\mathbf{W})$ for any $\mathbf{U} \in \mathcal{U}_{n_s}$. The set of equivalence classes of precoding matrices is exactly the Grassmann manifold $\mathcal{G}_{n_t,n_s}^{\mathbb{C}}$.

Linear receiver: The statement above does not hold anymore. There is nevertheless a smaller equivalence class of precoding matrices. We have

Lemma 1: The information rate with linear receiver (17) is invariant under permutations and phase multiplications of columns of the precoding matrix, $I_{\mathrm{lr}}(\mathbf{W}\mathbf{D}\mathbf{P}) = I_{\mathrm{lr}}(\mathbf{W})$, for any permutation matrix $\mathbf{P} \in \mathcal{U}_{n_s}$ and any diagonal matrix $\mathbf{D} \in \mathcal{U}_{n_s}$.

Proof: $I_{\mathrm{lr}}(\mathbf{W})$ depends on the diagonal elements of $\mathbf{F} = (\gamma\mathbf{H}_{\mathrm{eq}}^H\mathbf{H}_{\mathrm{eq}} + a\mathbf{I}_{n_s})^{-1}$, where the equivalent channel is $\mathbf{H}_{\mathrm{eq}} = \mathbf{H}\mathbf{W}$. Precoding with $\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{D}\mathbf{P}$ instead, the equivalent channel covariance matrix is $\widetilde{\mathbf{H}}_{\mathrm{eq}}^H\widetilde{\mathbf{H}}_{\mathrm{eq}} = \mathbf{P}^H\mathbf{D}^H\mathbf{H}_{\mathrm{eq}}^H\mathbf{H}_{\mathrm{eq}}\mathbf{D}\mathbf{P}$ and we have $(\gamma\widetilde{\mathbf{H}}_{\mathrm{eq}}^H\widetilde{\mathbf{H}}_{\mathrm{eq}} + a\mathbf{I}_{n_s}) = \mathbf{P}^H\mathbf{D}^H(\gamma\mathbf{H}_{\mathrm{eq}}^H\mathbf{H}_{\mathrm{eq}} + a\mathbf{I}_{n_s})\mathbf{D}\mathbf{P}$ with inverse $\widetilde{\mathbf{F}} = (\gamma\widetilde{\mathbf{H}}_{\mathrm{eq}}^H\widetilde{\mathbf{H}}_{\mathrm{eq}} + a\mathbf{I}_{n_s})^{-1} = \mathbf{P}^H\mathbf{D}^H\mathbf{F}\mathbf{D}\mathbf{P}$. It is straightforward that the diagonal elements of $\widetilde{\mathbf{F}}$ and $\mathbf{F}$ are the same up to permutation, and the result follows. ■

The set of equivalence classes of precoding matrices is thus the permutation-invariant flag manifold $\vec{\mathcal{F}}_{n_t,n_s}^{\mathbb{C}}$.

C. Codebook Design

Now we assume that the channel matrix is a random variable and that the transmitter picks the precoding matrix from a codebook $\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_{n_b}\}$ following a quantization rule $\mathrm{q} : \{\mathbf{H}\} \to \{i : 1 \leq i \leq n_b\}$. Given the instantaneous transmission rate $I(\mathbf{W})$ of a precoding matrix $\mathbf{W}$, the average information rate is

$$\overline{I} = \mathbb{E}_{\mathbf{H}}\left[I(\mathbf{C}_{\mathrm{q}(\mathbf{H})})\right]. \quad (18)$$

The codebook should be designed as a discretization of the space $\mathcal{V}_{n_t,n_s}^{\mathbb{C}}/\sim$ of equivalence classes of precoding matrices. The task thus becomes:

ML receiver: discretize the Grassmann manifold $\mathcal{G}_{n_t,n_s}^{\mathbb{C}}$

Linear receiver: discretize the flag manifold $\vec{\mathcal{F}}_{n_t,n_s}^{\mathbb{C}}$.

An optimum quantization $\mathrm{q}^*(\mathbf{H}) = \arg\max_{1 \leq i \leq n_b} I(\mathbf{C}_i)$ is untractable. For Grassmannian precoding, the quantization map $\hat{\mathrm{q}}(\mathbf{V}_{n_s}) = \arg\min_{1 \leq i \leq n_b} d_c(\mathbf{V}_{n_s}, \mathbf{C}_i)$ has been considered instead, and shown to be asymptotically optimum [16]. The corresponding codebooks are thus designed to minimize the average squared distortion $\mathbb{E}[d_c^2(\mathbf{V}_{n_s}, \mathbf{C}_{\hat{\mathrm{q}}(\mathbf{V}_{n_s})})]$ [16], [18]. We extend this principle to codebook design on the flag manifold by replacing the chordal distance d_c by the distance d_p defined in (8).

D. Simulations

We illustrate by simulations the proposed design, assuming independent and identically distributed Rayleigh channels and using codebooks generated by the Lloyd algorithm described in Section IV. Fig. 1 shows the information rate for 4×2 MIMO systems with ZF receiver, rank-2 transmission and 4-bit codebooks in $\vec{\mathcal{F}}_{4,2}^{\mathbb{C}}$ and $\mathcal{G}_{4,2}^{\mathbb{C}}$. The flag manifold codebook outperforms the Grassmannian with more than 1 dB. The performance of the precoding partitioning scheme of [6] is also depicted, with a perfromance in between the Grassmannian and flag codebooks. In the partitioning, bits are split equally between Grassmannian precoding and 2-stream orthogonalization, and codebooks for both are generated by Lloyd algorithms. When using this scheme, we have selected the codeword minimizing the flag distance d_p over all combinations of Grassmannian and orthogonalization codewords.

Fig. 2 shows the information rate for a full-rank system with an equal number of transmit and receive antennas. In this scenario, as pointed out in Remark 1, precoding is irrelevant if ML receiver is used, while for a linear receiver, precoding has an effect. For 2×2 systems, it is possible to recover most of the gap between no precoding and perfect precoding with a few feedback bits. However, when the number of antennas increases, e.g. for $n_t = n_r = 4$ in Fig. 2, the gain for a small number of bits is small. In the SNR range depicted, the performance gap between no and perfect precoding is some 5dB. Precoding with 5 bits recovers only 1.5dB of this gap.

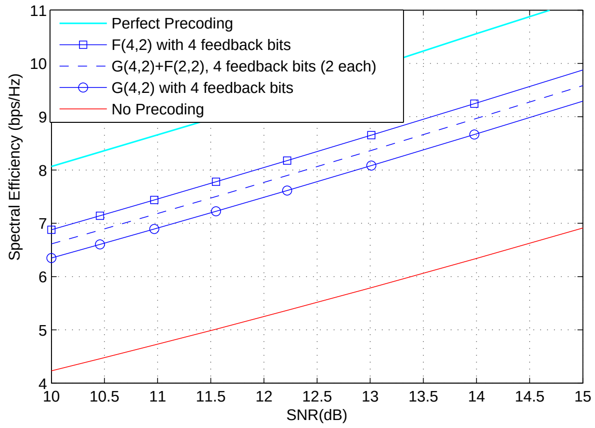

Fig. 1. Information rate of precoded MIMO systems with ZF receiver, $n_t = 4$, $n_r = n_s = 2$. Four-bit ($n_b = 16$) codebooks designed in flag $\vec{\mathcal{F}}^{\mathbb{C}}_{4,2}$ or Grassmann $\mathcal{G}^{\mathbb{C}}_{4,2}$ manifolds, and precoding partitioning [6] with 2 bits for Grassmannian and orthogonalization each.

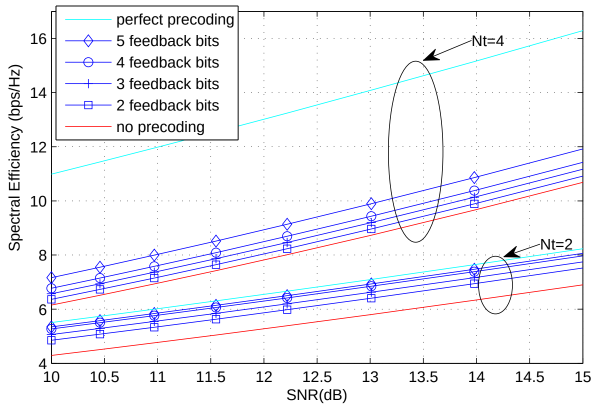

Fig. 2. Information rate of precoded MIMO systems with ZF receiver and full rank transmission, $n_t = n_r = n_s$. Codebooks with $2, 3, 4$ and 5-bits designed in flag manifold $\vec{\mathcal{F}}^{\mathbb{C}}_{n_t,n_t}$.

VI. Conclusion

Codebook-based unitary precoding is widely employed, and several codebook designs in the literature are examples of flag manifold discretization. It is thus expected that several codebook design problems of contemporary interest could be treated in a general manner as quantization problems of flag manifolds. In this paper, we concentrated on the design of unitary codebooks for MIMO systems with linear receivers. The spaces considered are certain permutation-invariant flag manifolds. We described a Lloyd algorithm to generate low-distortion codebooks in these spaces. The pertinence of the design principle was illustrated by simulations.

References

[1] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. Telecomm.*, vol. 10, pp. 585–595, 1999.

[2] 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.

[3] R. F. Picken, “The Duistermaat-Heckman integration formula on flag manifolds,” *J. of Math. Phys.*, vol. 31, no. 3, pp. 616–638, March 1990.

[4] S. Murray and C. Smann, “Quantization of flag manifolds and their supersymmetric extensions,” *Adv. Theor. Math. Phys.*, vol. 12, no. 3, pp. 641–710, 2008.

[5] K. Zyczkowski and H.-J. Sommers, “Hilbert-Schmidt volume of the set of mixed quantum states,” *J. Phys. A: Math. Gen.*, vol. 36, no. 39, p. 1011510130, 2003.

[6] H. Maattanen, K. Schober, O. Tirkkonen, and R. Wichman, “Precoder partitioning in closed-loop MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3910–3914, Aug. 2009.

[7] 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.

[8] 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.

[9] E. Sengul, H. Park, and E. Ayanoglu, “Bit-interleaved coded multiple beamforming with imperfect CSIT,” *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1505 –1513, May 2009.

[10] K. Huang, J. Andrews, and R. Heath, “Performance of orthogonal beamforming for SDMA with limited feedback,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 152 –164, Jan. 2009.

[11] I. Kim, S. Park, D. Love, and S. Kim, “Improved multiuser MIMO unitary precoding using partial channel state information and insights from the Riemannian manifold,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4014 –4023, Aug. 2009.

[12] S. Wagner, S. Sesia, and D. Slock, “On unitary beamforming for MIMO broadcast channels,” in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1 –5.

[13] ——, “Unitary beamforming under constant modulus constraint in MIMO broadcast channels,” in *Proc. 10th IEEE Workshop Signal Process. Adv. Wireless Commun.*, june 2009, pp. 151 –155.

[14] J. Lee, J.-K. Han, and J. Zhang, “MIMO technologies in 3GPP LTE and LTE-Advanced,” *EURASIP J. Wirel. Comm.*, vol. 2009, no. 1, p. 302092, 2009.

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

[16] 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.

[17] J. Heath, R.W., S. Sandhu, and A. Paulraj, “Antenna selection for spatial multiplexing systems with linear receivers,” *IEEE Commun. Lett*, vol. 5, no. 4, pp. 142 –144, April 2001.

[18] 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, March 2008.