Reduced Complexity Recursive Grassmannian Quantization

Stefan Schwarz, *Senior Member, IEEE*, and Markus Rupp, *Fellow, IEEE*

***Abstract*—We propose a novel recursive multi-stage approach to Grassmannian quantization. Compared to the commonly employed single-stage quantization, our method has the advantage of significantly decreasing the number of codebook searches required for quantization and, thus, reducing the complexity. On the downside, the multi-stage approach causes a slight rate-distortion degradation compared to single-stage quantization. We analyze the rate-distortion performance of the proposed recursive quantization approach, considering random vector quantization within the individual stages. We furthermore propose a bit-allocation optimization amongst the stages of the quantizer, given a constraint on the total number of quantization bits.**

***Index Terms*—Grassmannian quantization, CSI feedback, random vector quantization.**

I. Introduction

Grassmannian quantization deals with quantization of points on a Grassmann manifold $\mathcal{G}(n,m)$, i.e., with quantization of m-dimensional subspaces of an n-dimensional real- or complex-valued Euclidean space. Grassmannian quantization has a long-standing and successful history in the context of channel state information (CSI) feedback in wireless communication systems to support adaptive multiple-input multiple-output (MIMO) beamforming and precoding [1], [2].

It is well known that maximally spaced subspace packings are optimal for memoryless quantization of uniformly distributed points on the Grassmannian [3]. Such packings are difficult to obtain in general; yet, suboptimal codebooks with good performance can efficiently be constructed [4]. Unfortunately, the required quantization codebook size to achieve a certain quantization distortion grows exponentially with the dimensions n and m of the Grassmann manifold [5]. Hence, for large-dimensional manifolds the storage requirements and the computational complexity of the codebook search are practically often not feasible.

Research has therefore focused on exploiting correlation properties of the source to reduce the required codebook size. Efficient differential/predictive manifold quantizers, which exploit temporal correlation of the source, are proposed in [6]–[9] and correlated codebook constructions for quantization of non-isotropically distributed source samples are available in [10]–[13]. Furthermore, a number of computationally efficient Grassmannian quantization approaches have been developed. In [14], the authors propose a trellis coded quantization approach that achieves a performance close to random vector quantization (RVQ) and allows for efficient quantization by means of the Viterbi algorithm. Recently, a cube-split quantizer has been proposed in [15], which supports computationally efficient quantization on the Grassmannian of lines. Cube-split is essentially a scalar quantizer that transforms the quantization variables on-the-fly to allow for scalar quantization on the unit interval, thereby avoiding storage of quantization codebooks. However, it requires storage of a look-up-table for the cumulative distribution function (CDF) of Gaussian random variables, with a resolution that grows exponentially with the number of quantization bits; thus, the claimed "zero storage requirement" is only partially true. Furthermore, the method is restricted to quantization of one-dimensional subspaces.

Contribution: In this paper, we propose a general approach for recursive decomposition of a large Grassmannian quantization problem into a series of smaller Grassmannian quantization problems, which can be solved with reduced computational complexity. This multi-stage recursive Grassmannian quantization approach is less efficient than a single-stage quantizer in terms of rate-distortion performance; however, its computational demands are orders of magnitude lower. We specifically employ RVQ within the individual stages of the quantizer, since this allows for an analytic performance characterization. Nevertheless, any other computationally and/or rate-distortion efficient Grassmannian quantization codebook construction can (and in practice should) be utilized within the individual quantization stages. E.g., if we decompose the problem into a series of one-dimensional Grassmannian quantization problems, we can employ the cube-split quantizer within each stage. This work is an extension of our dual-stage quantizer proposed in [16] to multi-stage recursive quantization.

Notation: The Grassmann manifold of m-dimensional subspaces of the complex-valued n-dimensional Euclidean space is $\mathcal{G}(n,m)$. The trace of matrix $\mathbf{A}$ is $\operatorname{tr}(\mathbf{A})$, the conjugate-transpose is $\mathbf{A}^{\mathsf{H}}$ and the Frobenius norm is $\|\mathbf{A}\|$. The m-dimensional subspace spanned by the orthogonal basis $\mathbf{U} \in \mathbb{C}^{n\times m}$, $\mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_m, m \leq n$ is $\operatorname{span}(\mathbf{U})$, where $\mathbf{I}_m$ is an $m \times m$ identity matrix. The expected value of random variable x is $\mathbb{E}(x)$. The operation $a_{\min} = \arg\min_{a\in\mathcal{A}} f(a)$ determines the minimizer $a_{\min}$ of the function $f(a)$ over the set $\mathcal{A}$.

II. Grassmannian Quantization

We consider quantization of points that are uniformly distributed on the complex-valued Grassmann manifold $\mathcal{G}(n,m)$. We employ orthogonal bases $\mathbf{U} \in \mathbb{C}^{n\times m}, \mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_m, m \leq$

Manuscript received December 16, 2019; revised January 20, 2020; accepted January 23, 2020. Date of publication January 27, 2020; date of current version February 27, 2020. This work was supported in part by the Austrian Federal Ministry for Digital and Economic Affairs and in part by the National Foundation for Research, Technology and Development. Stefan Schwarz leads the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shiwen He. *(Corresponding author: Stefan Schwarz.)*

The authors are with the Institute of Telecommunications, TU Wien, 1040 Vienna, Austria (e-mail: sschwarz@nt.tuwien.ac.at; rupp@nt.tuwien.ac.at).

Digital Object Identifier 10.1109/LSP.2020.2969841

n to represent points on the Grassmannian. Thus, the Grassmannian $\mathcal{G}(n,m)$ can be defined as the following set of subspaces spanned by orthogonal bases of size $n \times m$

$$\mathcal{G}(n,m) = \left\{\text{span}\left(\mathbf{U}\right) | \mathbf{U} \in \mathbb{C}^{n\times m}, \mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_m\right\}. \quad (1)$$

To quantize the m-dimensional subspace $\text{span}(\mathbf{U}) \in \mathcal{G}(n,m)$ onto a d-dimensional subspace, we utilize a quantization codebook $\mathcal{Q}_d^{(n)}$ consisting of $D = 2^b$ orthogonal bases of size $n \times d$, where $m \leq d \leq n$,

$$\mathcal{Q}_d^{(n)} = \left\{\mathbf{Q}_\ell \in \mathbb{C}^{n\times d} | \mathbf{Q}_\ell^{\mathsf{H}}\mathbf{Q}_\ell = \mathbf{I}_d, \ell \in \{1,\ldots,D\}\right\}. \quad (2)$$

Here, b denotes the number of quantization bits. Specifically, we consider minimal chordal distance quantization

$$\mathbf{W} = \underset{\mathbf{Q}_\ell \in \mathcal{Q}_d^{(n)}}{\arg\min}\ \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U},\mathbf{Q}_\ell\right), \quad (3a)$$

$$\mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U},\mathbf{Q}_\ell\right) = m - \mathrm{tr}\left(\mathbf{U}^{\mathsf{H}}\mathbf{Q}_\ell\mathbf{Q}_\ell^{\mathsf{H}}\mathbf{U}\right), \quad (3b)$$

to determine the subspace $\text{span}(\mathbf{W}), \mathbf{W} \in \mathcal{Q}_d^{(n)}$, with the largest overlap with $\text{span}(\mathbf{U})$. This quantization metric naturally arises in the context of CSI feedback for MIMO wireless communications [1], [2], [17].

We do not consider optimization of the codebook $\mathcal{Q}_d^{(n)}$ in this work, but rather assume that it is randomly constructed with codebook entries following an isotropic distribution, such that $\text{span}(\mathbf{Q}_\ell), \mathbf{Q}_\ell \in \mathcal{Q}_d^{(n)}$, is uniformly distributed on $\mathcal{G}(n,d)$; that is, we consider RVQ.

III. Single-Stage Quantization

In the following, we consider the important case that the subspace dimensions of the source sample $\mathbf{U}$ and its quantized representation $\mathbf{W}$ are equal. In single-stage quantization, this is achieved by solving the quantization problem (3) with a codebook $\mathcal{Q}_m^{(n)}$ consisting of $d = m$-dimensional orthogonal bases. We denote the corresponding optimal single-stage solution as $\hat{\mathbf{U}}_{\mathrm{SS}} = \mathbf{W} \in \mathcal{Q}_m^{(n)}$.

The quantization performance of single-stage quantization of uniformly distributed source samples $\mathbf{U}$ has been well-investigated in [5]. In [5, Theorem 2], the authors provide upper- and lower-bounds on the average chordal distance quantization distortion. The upper bound specifically corresponds to the performance of RVQ. These bounds are tight for sufficiently large codebook size D; however, for smaller codebook sizes they can be relatively loose depending on the dimensions of the considered Grassmannian $\mathcal{G}(n,m)$.

A better bound for RVQ with small codebook sizes is obtained by recognizing from (3) that the chordal distance quantization error is dictated by $\max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \mathrm{tr}(\mathbf{U}^{\mathsf{H}}\mathbf{Q}_\ell\mathbf{Q}_\ell^{\mathsf{H}}\mathbf{U})$. This term can be bounded as

$$\max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \mathrm{tr}\left(\mathbf{U}^{\mathsf{H}}\mathbf{Q}_\ell\mathbf{Q}_\ell^{\mathsf{H}}\mathbf{U}\right) = \max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \sum_{k=1}^{m} \left\|\mathbf{U}^{\mathsf{H}}\mathbf{q}_{k,\ell}\right\|^2 \leq \sum_{k=1}^{m} \max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \left\|\mathbf{U}^{\mathsf{H}}\mathbf{q}_{k,\ell}\right\|^2, \quad (4)$$

where $\mathbf{q}_{k,\ell}$ denotes the k-th column of $\mathbf{Q}_\ell$. Equality in (4) is achieved if the dimension $m = 1$ or the codebook size $D = 1$.

Each term $\|\mathbf{U}^{\mathsf{H}}\mathbf{q}_{k,\ell}\|^2$ in (4) follows a beta distribution with shape parameters $n-m$ and m, $\beta(n-m,m)$ [18]. Thus, the distribution of $\max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \|\mathbf{U}^{\mathsf{H}}\mathbf{q}_{k,\ell}\|^2$ is determined by the maximum order statistics of D independent beta random variables. Utilizing the recurrence relation of [19], we can calculate the expected value $\mu_{n,m}^{(D)} = \mathbb{E}(\max_{\mathbf{Q}_\ell \in \mathcal{Q}_m^{(n)}} \|\mathbf{U}^{\mathsf{H}}\mathbf{q}_{k,\ell}\|^2)$. Together with Eqs. (3) and (4), $\mu_{n,m}^{(D)}$ immediately provides a lower bound on the average chordal distance quantization error

$$\bar{\mathrm{d}}_{\mathrm{c,SS}}^2 = \mathbb{E}\left(\mathrm{d}_{\mathrm{c}}^2\left(\mathbf{U},\hat{\mathbf{U}}_{\mathrm{SS}}\right)\right) \geq m\left(1-\mu_{n,m}^{(D)}\right). \quad (5)$$

For small codebook size D, this bound is tighter than the bound of [5]; for $m = 1$ or $D = 1$ the bound is even achieved with equality. However, for large codebook sizes, calculating the recurrence relation of [19] becomes prohibitively complex; we then apply the asymptotically tight RVQ bound of [5].

IV. Multi-Stage Recursive Quantization

Single-stage quantization may require a very large codebook size to achieve a sufficiently low quantization error, which implies that solving (3) can be prohibitively complex. To reduce this complexity, we propose to recursively split the quantization problem into R subproblems/stages, each employing a much smaller quantization codebook size $D_i = 2^{b_i} \ll D$. We thereby ensure that the total number of quantization bits is the same $\sum_{i=1}^{R} b_i = b$, while the total quantization search complexity is significantly reduced $\sum_i D_i = \sum_i 2^{b_i} \ll 2^b = D$.

Specifically, the recursive multi-stage quantized representation $\hat{\mathbf{U}}_{\mathrm{MS}} \in \mathbb{C}^{n\times m}$ of the source sample $\mathbf{U} \in \mathbb{C}^{n\times m}$ is

$$\hat{\mathbf{U}}_{\mathrm{MS}} = \prod_{i=1}^{R} \mathbf{W}_i,\ \mathbf{W}_i \in \mathbb{C}^{d_{i-1}\times d_i}, \quad (6)$$

$$d_0 = n,\ d_{i-1} > d_i,\ d_R = m. \quad (7)$$

Here, d_i is the subspace dimension of the i-th recursion. Starting with $i = 1$, each $\mathbf{W}_i$ is recursively obtained by solving a reduced size quantization problem. Specifically, defining $\mathbf{B}_0 = \mathbf{U}$, the quantization subproblem of the i-th recursion is

$$\mathbf{W}_i = \underset{\mathbf{Q}_\ell \in \mathcal{Q}_{d_i}^{(d_{i-1})}}{\arg\min}\ \mathrm{d}_{\mathrm{c}}^2\left(\mathbf{B}_{i-1},\mathbf{Q}_\ell\right), \quad (8)$$

where $\mathcal{Q}_{d_i}^{(d_{i-1})}$ is constructed as in (2) and has size $D_i = 2^{b_i}$.

After calculating $\mathbf{W}_i$, matrix $\mathbf{B}_i$ is obtained by applying the subspace quantization based combining (SQBC) principle [20]

$$\mathbf{B}_i = \mathbf{W}_i^{\mathsf{H}}\mathbf{B}_{i-1}\left(\mathbf{B}_{i-1}^{\mathsf{H}}\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\mathbf{B}_{i-1}\right)^{-1/2} \in \mathbb{C}^{d_i\times m}. \quad (9)$$

The m-dimensional orthogonal SQBC matrix $\mathbf{B}_i$ is constructed such that $\mathrm{d}_{\mathrm{c}}^2(\mathbf{B}_{i-1},\mathbf{W}_i) = \mathrm{d}_{\mathrm{c}}^2(\mathbf{B}_{i-1},\mathbf{W}_i\mathbf{B}_i)$. Hence, it corresponds to the projection of $\mathbf{B}_{i-1}$ onto the subspace spanned by $\mathbf{W}_i$. After calculating $\mathbf{B}_i$, the quantizer proceeds to the next recursion/stage $i+1$, until $i = R$ and $d_R = m$.

A. Quantization via the Orthogonal Complement

When the dimension steps $\Delta_i = d_{i-1} - d_i$ are small, calculating (8) may require large matrix multiplications. We can then reduce complexity by employing a quantization codebook for the orthogonal complements of the matrices in $\mathcal{Q}_{d_i}^{(d_{i-1})}$

$$\bar{\mathcal{Q}}_{\Delta_i}^{(d_{i-1})} = \left\{\bar{\mathbf{Q}}_\ell \in \mathbb{C}^{d_{i-1}\times\Delta_i} | \bar{\mathbf{Q}}_\ell^{\mathsf{H}}\bar{\mathbf{Q}}_\ell = \mathbf{I}_{\Delta_i}, \bar{\mathbf{Q}}_\ell^{\mathsf{H}}\mathbf{Q}_\ell = \mathbf{0}, \mathbf{Q}_\ell \in \mathcal{Q}_{d_i}^{(d_{i-1})}, \ell \in \{1,\ldots,D_i\}\right\}. \quad (10)$$

The quantization step (8) is in this case replaced by

$$\bar{\mathbf{W}}_i = \underset{\bar{\mathbf{Q}}_\ell \in \bar{\mathcal{Q}}_{\Delta_i}^{(d_{i-1})}}{\mathrm{argmin}}\ \mathrm{tr}\left(\bar{\mathbf{Q}}_\ell^{\mathsf{H}}\mathbf{B}_{i-1}\mathbf{B}_{i-1}^{\mathsf{H}}\bar{\mathbf{Q}}_\ell\right). \quad (11)$$

Given $\bar{\mathbf{W}}_i$, $\mathbf{W}_i$ is calculated by finding a d_i-dimensional basis for the orthogonal complement of $\bar{\mathbf{W}}_i$. Hence, rather than determining the d_i-dimensional subspace $\mathrm{span}(\mathbf{W}_i)$ that has maximum overlap with $\mathrm{span}(\mathbf{B}_{i-1})$ directly by solving (8), we first find the Δ_i-dimensional subspace $\mathrm{span}(\bar{\mathbf{W}}_i)$ that has minimum overlap with $\mathrm{span}(\mathbf{B}_{i-1})$ and then calculate its orthogonal complement $\mathrm{span}(\mathbf{W}_i)$. This approach is less complex if $\Delta_i < d_i$. It is especially suitable when $\Delta_i = 1$, as then the individual subproblems (11) reduce to one-dimensional Grassmannian quantization problems, for which many efficient codebook constructions are available, e.g., [14], [15].

B. Quantization Performance

The average chordal distance quantization error $\bar{\mathrm{d}}^2_{\mathrm{c,MS}} = \mathbb{E}(\mathrm{d}^2_{\mathrm{c}}(\mathbf{U}, \hat{\mathbf{U}}_{\mathrm{MS}}))$ is governed by the following theorem:

Theorem 1 (Multi-Stage Quantization Error): The normalized average chordal distance quantization error for recursive multi-stage quantization of points $\mathrm{span}(\mathbf{U})$ uniformly distributed on the complex-valued Grassmannian $\mathcal{G}(n, m)$ is

$$\frac{1}{m}\,\bar{\mathrm{d}}^2_{\mathrm{c,MS}} = 1 - \prod_{i=1}^{R}\left(1 - \frac{1}{m}\,\bar{\mathrm{d}}^2_{\mathrm{c,i}}\right), \tag{12}$$

where $\bar{\mathrm{d}}^2_{\mathrm{c,i}}$ denotes the average chordal distance of the i-th quantization stage. Proof: see Appendix A.

When employing RVQ, the average distortions $\bar{\mathrm{d}}^2_{\mathrm{c,i}}$ of the individual stages are governed by the rate-distortion bounds provided in [5, Theorem 2]. For sufficiently large codebook sizes $D_i = 2^{b_i}$, we have

$$\bar{\mathrm{d}}^2_{\mathrm{c,i}} = k_{d_{i-1},m,d_i} D_i^{-\frac{1}{m(d_{i-1}-d_i)}}, \tag{13}$$

where the constants $k_{n,p,q}$ are defined in [5].[1]

C. Optimal Bit-Allocation

Given a total number of quantization bits b, the performance achieved by the multi-stage quantizer depends on the allocation of bits b_i amongst the quantization stages. With the result of Theorem 1 and the average distortion achieved by RVQ according to (13), we formulate an optimization problem to minimize the overall distortion

$$\max_{b_i} \sum_{i=1}^{R} \log\left(1 - \frac{1}{m} k_{d_{i-1},m,d_i} 2^{-\frac{b_i}{m(d_{i-1}-d_i)}}\right),$$

$$\text{subject to:}\quad b_i \geq 0, \forall i \in \{1, \ldots, R\}, \quad \sum_{i=1}^{R} b_i = b. \tag{14}$$

Here, we converted the product in (12) to a sum by applying the logarithm. Notice that these logarithmic terms are concave in the bit-allocation and, hence, their sum is also concave. This convex maximization problem can therefore be handled by state-of-the-art optimization tools, such as, CVX [21].

When the dimension steps $\Delta_i = d_{i-1} - d_i$ are equal for all stages of the quantizer $\Delta_i = \Delta, \forall i$, the optimal bit-allocation can be determined from the following theorem:

Theorem 2 (Optimal Bit-Allocation for Equal Dimension Steps): If all dimension steps Δ_i of the recursive multi-stage quantizer are equal to $\Delta, \forall i$, the optimal bit-allocation of RVQ for sufficiently large codebook sizes, such that [5, Theorem 2] is applicable, is calculated as

$$b_i = m\Delta\left(\log_2\left(k_{d_{i-1},m,d_i}\right) - \log_2\left(\bar{\mathrm{d}}^2_{\mathrm{c}}\right)\right), \tag{15}$$

$$\log_2\left(\bar{\mathrm{d}}^2_{\mathrm{c}}\right) = \frac{m\Delta\sum_{i=1}^{R}\log_2\left(k_{d_{i-1},m,d_i}\right) - b}{\sum_{i=1}^{R} m\Delta}, \tag{16}$$

where m is the subspace dimension of the Grassmannian source $\mathcal{G}(n, m)$ and b is the total number of quantization bits.

With this bit allocation, every quantization stage achieves the same average distortion of $\bar{\mathrm{d}}^2_{\mathrm{c}}$. Proof: see Appendix B.

Notice, for small number of bits b, Eq. (15) may provide negative bit-allocations to some early stages $j < i_{\min}$ of the quantizer, which does not make sense. If this is the case, we fix the bit-allocation of these stages to $b_j = 0, \forall j < i_{\min}$ and solve for the optimal bit-allocation of the remaining stages by summing in (16) only from $i_{\min}$ to R.

[1]We combine the multiple separate constants of [5] into the single constant $k_{n,p,q}$; n is the dimension of the embedding space, p is the dimension of the source subspace and q is the dimension of the codebook entries.

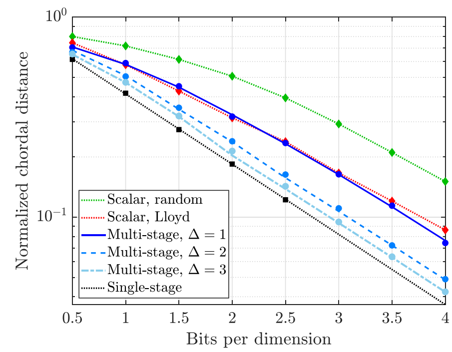

Fig. 1. Quantization distortion versus number of quantization bits on $\mathcal{G}(7, 1)$.

V. Simulations

We first investigate the rate-distortion performance of our quantizer for $\mathcal{G}(7, 1)$, considering the three cases $\Delta_i = \Delta \in \{1, 2, 3\}$ corresponding to $R \in \{6, 3, 2\}$. The results are given in terms of normalized chordal distance $\frac{1}{m}\,\mathbb{E}(d_c^2)$ versus bits per dimension $\frac{b}{n\,m}$. We compare to single-stage quantization and to scalar quantization of the individual entries of $\mathbf{U}$, employing random scalar quantization (i.e., a quantization codebook where the entries are randomly generated from the same distribution as the entries of $\mathbf{U}$, following the same principle as RVQ) and a Lloyd-optimized scalar codebook. The results are shown in Fig. 1. The marked points correspond to simulated values and the lines correspond to the theoretical distortion of RVQ, employing Theorem 1 and the approach discussed in Section III to estimate the RVQ performance. As expected, single-stage RVQ performs best, with multi-stage RVQ following closely, depending on Δ and, hence, the number of quantization stages. The Lloyd-optimized scalar quantizer performs similar to multi-stage RVQ with $\Delta = 1$. However, if we would use an optimized codebook instead of RVQ, the multi-stage approach would also perform better.

The benefit of multi-stage quantization is shown in Fig. 2, where we plot the total number of codebook searches required by single-stage and multi-stage quantization. In the multi-stage approach, the total number of bits b is split amongst the stages and since $2^b \gg \sum_i 2^{b_i}$, the search complexity is significantly reduced, especially when the number of stages R is large.

In our second simulation, we consider quantization on $\mathcal{G}(16, 2)$. The rate-distortion results are shown in Fig. 3. Notice,

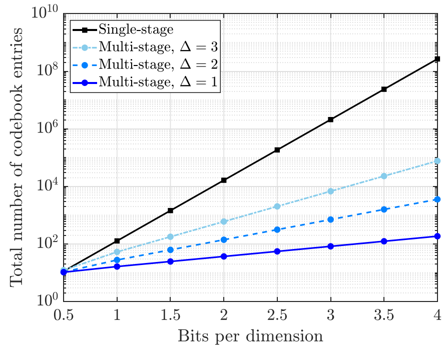

Fig. 2. Codebook entries versus number of quantization bits on $\mathcal{G}(7, 1)$.

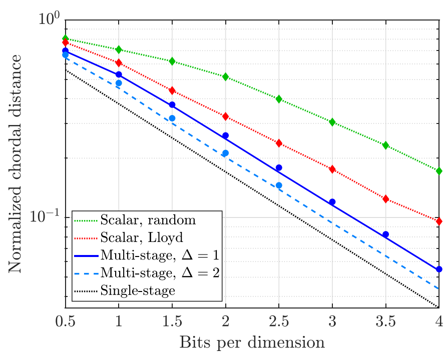

Fig. 3. Quantization distortion versus number of quantization bits on $\mathcal{G}(16, 2)$.

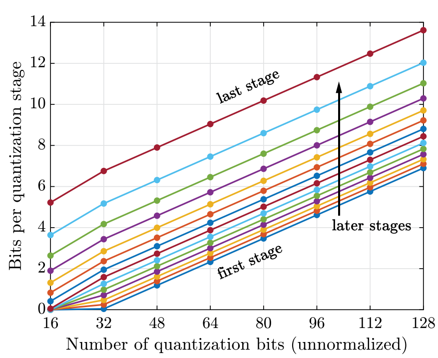

Fig. 4. Distribution of quantization bits b_i over quantization stages versus number of quantization bits b (unnormalized) on $\mathcal{G}(16, 2)$ for $\Delta_i = \Delta = 1$.

single-stage quantization requires already for 0.5 bits/dimension a codebook size of $2^{16} = 65\,536$, which is hardly feasible in practice. We therefore were not able to simulate the performance of single-stage quantization and only show the theoretical estimate. On the other hand, the total codebook size of the multi-stage quantizer for $\Delta = 1$ scales from 72 entries for 0.5 bits/dimension to 23 408 entries for 4 bits/dimension.

In Fig. 4, we investigate how the optimal bit-allocation according to Theorem 2 distributes the total number of bits b amongst quantization stages. We observe that later stages of the quantizer get more bits assigned. If we want to further limit the search complexity of the multi-stage approach, we could additionally consider an upper bound on the number of bits per stage in the optimization (15) to obtain a more equal bit-allocation amongst stages; however, this would also imply a degradation of the rate-distortion performance.

VI. Conclusion

Recursive multi-stage Grassmannian quantization can drastically reduce the quantization complexity for the cost of a moderate loss in rate-distortion performance. In particular, applying a dimension step of $\Delta = 1$ and quantization via the orthogonal complement is promising for multi-dimensional Grassmannians $\mathcal{G}(n, m)$ with $m > 1$, since it reduces the multi-dimensional single-stage Grassmannian quantization problem to a series of one-dimensional quantization problems and, thus, allows for the application of computationally and rate-distortion efficient one-dimensional Grassmannian codebooks.

Appendix

A. *Proof of Theorem 1*

Theorem 1 imposes the condition that $\operatorname{span}(\mathbf{U})$ is uniformly distributed on $\mathcal{G}(n, m)$, which implies that the orthogonal basis $\mathbf{U}$ is isotropically distributed. To prove Theorem 1, the central observation is that all the subspaces $\operatorname{span}(\mathbf{B}_i)$ are thus also uniformly distributed on their respective Grassmannians. This follows as the projection of $\mathbf{B}_0 = \mathbf{U}$ onto $\mathbf{W}_1$ is isotropically distributed within the subspace $\operatorname{span}(\mathbf{W}_1)$ and, hence, $\mathbf{B}_1$ is isotropically distributed; from this it follows that the projection of $\mathbf{B}_1$ onto $\mathbf{W}_2$ is isotropically distributed within $\operatorname{span}(\mathbf{W}_2)$ and, hence, $\mathbf{B}_2$ is isotropically distributed; and so on.

Now consider the normalized chordal distance quantization error of the recursive quantizer

$$\frac{1}{m}\bar{\mathrm{d}}^2_{\mathrm{c,MS}} = 1 - \frac{1}{m}\mathbb{E}\left(\operatorname{tr}\left(\mathbf{W}_R^{\mathsf{H}} \dots \mathbf{W}_1^{\mathsf{H}} \mathbf{B}_0 \mathbf{B}_0^{\mathsf{H}} \mathbf{W}_1 \dots \mathbf{W}_R\right)\right).$$

By the SQBC construction (9) we have

$$\mathbf{W}_1^{\mathsf{H}} \mathbf{B}_0 \mathbf{B}_0^{\mathsf{H}} \mathbf{W}_1 = \mathbf{B}_1 \left(\mathbf{B}_0^{\mathsf{H}} \mathbf{W}_1 \mathbf{W}_1^{\mathsf{H}} \mathbf{B}_0\right) \mathbf{B}_1^{\mathsf{H}},$$

and by isotropy of $\mathbf{B}_0$ we conclude

$$\mathbb{E}\left(\mathbf{B}_0^{\mathsf{H}} \mathbf{W}_1 \mathbf{W}_1^{\mathsf{H}} \mathbf{B}_0\right) = \left(1 - \frac{1}{m}\,\bar{\mathrm{d}}^2_{\mathrm{c},1}\right)\mathbf{I}_m.$$

Inserting these results into the equation of the normalized chordal distance quantization error and recursing down until $\mathbf{W}_R$ in a similar way, we obtain the result of Theorem 1.

B. *Proof of Theorem 2*

To prove Theorem 2, it is simpler to consider the equivalent problem of maximizing $\prod_{i=1}^{R}(1 - \frac{1}{m}\,\bar{\mathrm{d}}^2_{\mathrm{c,i}})$ rather than maximization of the logarithmic sum as stated in (14). Let us ignore the constraint $b_i \geq 0$ for now and calculate the derivative of the Lagrangian of the corresponding optimization problem. From this we arrive at the optimality conditions

$$\prod_{j \neq i}\left(1 - \frac{1}{m}\,\bar{\mathrm{d}}^2_{\mathrm{c,j}}\right)\frac{\bar{\mathrm{d}}^2_{\mathrm{c,i}}}{\Delta_i} = \lambda, \forall i.$$

Since $\Delta_i = \Delta, \forall i$, the optimal solution is to allocate the bits such that all stages achieve equal distortion $\bar{\mathrm{d}}^2_{\mathrm{c,i}} = \bar{\mathrm{d}}^2_{\mathrm{c}}, \forall i$. From this condition and the constraint $\sum_{i=1}^{R} b_i = b$ we can calculate the Lagrange multiplier λ and $\bar{\mathrm{d}}^2_{\mathrm{c}}$, and finally obtain Theorem 2.

Acknowledgment

S. Schwarz leads the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion.

References

[1] D. Love, R. Heath, Jr., and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, Oct. 2003.

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

[3] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: Packings in Grassmannian spaces,” *Exp. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[4] I. S. Dhillon, R. Heath, Jr., T. Strohmer, and J. A. Tropp, “Constructing packings in Grassmannian manifolds via alternating projection,” *Exp. Math.*, vol. 17, no. 1, pp. 9–35, 2008.

[5] 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.

[6] D. Sacristan-Murga, M. Payaro, and A. Pascual-Iserte, “Transceiver design framework for multiuser MIMO-OFDM broadcast systems with channel Gram matrix feedback,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1774–1787, May 2012.

[7] O. El Ayach and R. Heath, Jr.,“Grassmannian differential limited feedback for interference alignment,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6481–6494, Dec. 2012.

[8] S. Schwarz, R. Heath, Jr., and M. Rupp, “Adaptive quantization on a Grassmann-manifold for limited feedback beamforming systems,” *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4450–4462, Sep. 2013.

[9] S. Schwarz, R. Heath, Jr., and M. Rupp, “Adaptive quantization on the Grassmann-manifold for limited feedback multi-user MIMO systems,” in *Proc. 38th Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5021–5025.

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

[11] P. Xia and G. B. 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.

[12] V. Raghavan, R. Heath, and A. Sayeed M., “Systematic codebook designs for quantized beamforming in correlated MIMO channels,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1298–1310, Sep. 2007.

[13] S. Schwarz and M. Rupp, “Evaluation of distributed multi-user MIMO-OFDM with limited feedback,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6081–6094, Aug. 2014.

[14] J. Choi, Z. Chance, D. Love, and U. Madhow, “Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[15] A. Decurninge and M. Guillaud, “Cube-split: Structured quantizers on the Grassmannian of lines,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2017, pp. 1–6.

[16] S. Schwarz, M. Rupp, and S. Wesemann, “Grassmannian product codebooks for limited feedback massive MIMO with two-tier precoding,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1119–1135, Sep. 2019.

[17] N. Ravindran and N. Jindal, “Limited feedback-based block diagonalization for the MIMO broadcast channel,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[18] J. Nielsen, “The distribution of volume reductions induced by isotropic random projections,” *Adv. Appl. Probability*, vol. 31, no. 4, pp. 985–994, Dec. 1999.

[19] P. Y. Thomas and P. Samuel, “Recurrence relations for the moments of order statistics from a beta distribution,” *Statist. Papers*, vol. 49, no. 1, pp. 139–146, Mar. 2008.

[20] S. Schwarz and M. Rupp, “Subspace quantization based combining for limited feedback block-diagonalization,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5868–5879, Nov. 2013.

[21] M. Grant and S. Boyd, “CVX: MATLAB software for disciplined convex programming, version 2.1,” Mar. 2014. [Online]. Available: http://cvxr.com/cvx