GRASSMANNIAN PREDICTIVE CODING FOR LIMITED FEEDBACK MULTIUSER MIMO SYSTEMS

Takao Inoue
National Instruments
11500 N. Mopac Expwy,
Austin, TX 78759 USA.
Email: takao@ieee.org

Robert W. Heath, Jr. *
The University of Texas at Austin
1 University Station C0803,
Austin, TX 78712 USA.
Email: rheath@ece.utexas.edu

ABSTRACT
Grassmannian beamforming is an efficient way to quantize channel state information in multiple-input multiple-output wireless systems. Unfortunately, multiuser systems require larger codebooks since the quantization error creates residual interference that limits the sum rate performance. To reduce the feedback requirements in multiuser systems, we propose Grassmannian predictive coding to exploit temporal channel correlation. The proposed algorithm exploits the differential geometric structure of the Grassmann manifold. The difference between points, prediction, and quantization are defined using the tangent space of the Grassmann manifold. We show that with practical feedback rates, a significant sum rate improvement can be obtained as a function of the channel correlation.

Index Terms— array signal processing, feedback communication, MIMO systems, prediction methods.

1. INTRODUCTION
Grassmannian beamforming is a limited feedback technique used in multiple antenna wireless systems (see [1] and the references therein). The key idea is that a beamforming vector is quantized by selecting a codeword from a specially designed codebook representing points on the Grassmann manifold. Grassmannian beamforming was first applied to multiple-input multiple-output (MIMO) wireless systems. Later it was extended to multiuser multiple input multiple output (MU-MIMO) systems. Unfortunately, MU-MIMO requires larger codebook sizes, and thus higher feedback overhead, motivating the development of more advanced Grassmannian coding methods [2].

The foundation of Grassmannian beamforming and limited feedback in general is memoryless vector quantization [3]. A traditional approach for reducing quantization overhead is to employ predictive vector quantization (PVQ), which exploits temporal correlation in the source. Conventional PVQ has been applied to encode channel vectors in MU-MIMO systems [4, 5, 6]. It was shown that conventional PVQ can improve achievable sum rates if sufficient training can be performed to obtain a codebook suitable for a given temporal channel correlation. A more structured approach for predictive coding of normalized channel vectors in multiple-input single-output systems was proposed [7]. They used the complex Householder transform to successively decompose the prediction error. It was shown that the unit delay predictor is optimal in their framework.

The codebook was designed using Lloyd algorithm for a given long term correlation statistics. A differential feedback approach using a rotation codebook has been proposed for spatial multiplexing system [8]. They also require long term correlation statistics to design suitable codebook and exploit the structure of the Riemannian manifold. Unfortunately, the codebooks are specific to the given long term statistics and it is not clear what the impact is when there is mismatch in the codebook. A simpler technique to exploit the temporal correlation and reduce feedback requirements for multiuser system was proposed in [9]. Using the differential geometric structure of the Grassmann manifold, quantized and an infrequently sent feedback error was used to extrapolate the future CSI thereby reducing the amount of feedback.

In this paper, we propose a Grassmannian predictive coding (GPC) algorithm for compressing correlated time series on the Grassmann manifold. Predictive quantization in the Grassmann manifold is challenging due to its non-Euclidean structure where the usual linear operations do not apply. This makes it difficult to perform basic tasks like computing an error signal, predicting the future value of a signal, and quantizing the prediction error. To solve this problem, we exploit the differential geometric structure of the Grassmann manifold to define a notion of one step prediction as well as tangential representations of the error signal. The tangential prediction error is encoded using a reasonable codebook size. We apply the proposed algorithm to reduce feedback requirements in a limited feedback MU-MIMO system. Our numerical results show that the proposed algorithm provides significant improvement in the accuracy of feedback CSI, leading to sum rates close to the case with perfect CSI under mild temporal correlation.

Notation: We use lower case bold letters, e.g., v, to denote vectors and upper case bold letters, e.g., H, to denote matrices. We use T, †, and H to denote the transposition, Hermitian transpose, and pseudo inverse, respectively. The n-th column entry of a matrix A is denoted by [A]n. Expectation is denoted E[·].

2. SYSTEM OVERVIEW
In this section we review the MU-MIMO system model. While we focus on MU-MIMO for performance evaluation, the proposed predictive quantization algorithm can be applied to single user MIMO and multicell MIMO applications of Grassmannian beamforming.

2.1. Discrete Time System Model
Consider a limited feedback zero forcing MU-MIMO system with $N_t$ transmit antennas at the base station and $U = N_t$ mobile users.
each equipped with single receive antenna [2]. To isolate the impact of channel direction limited feedback, we assume that $U$ users are already scheduled from a user pool and do not consider the problem of user scheduling and power allocation. Let $s_u[k]$, $v_u[k]$, and $h_u[k]$ be the complex baseband transmit symbol, $N_t \times 1$ unit norm beamforming vector, and $N_t \times 1$ channel vector for $u$-th user at time index $k$, respectively. We assume that the transmit vector, 
$$
\mathbf{s} = [s_1, \ldots, s_U]^T,
$$
satisfies the total transmit power constraint $\mathbb{E}[|\mathbf{s}|^2] \leq P$. The received signal at the $u$-th user may be written in discrete-time (assuming perfect synchronization) as

$$
y_u[k] = \mathbf{h}_u^*[k]\mathbf{v}_u[k]s_u[k] + \mathbf{h}_u^*[k]\sum_{n=1, n \neq u}^U \mathbf{v}_n[k]s_n[k] + n_u[k] \tag{1}
$$

where $n_u[k]$ is an independent and identically distributed (i.i.d.) zero mean complex Gaussian noise with unit variance for user $u$. Let $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_U]^*$ be the $U \times N_t$ composite channel matrix. With perfect CSI, the interference can be perfectly eliminated by choosing the unit norm beamforming vector $\mathbf{v}_u$ as the normalized columns of pseudo inverse composite channel matrix, i.e., $\mathbf{v}_u = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$. With the assumption of a perfect channel estimate $\hat{\mathbf{h}}_u[k]$ at the receiver, the channel direction is $\mathbf{g}_u(k) = \hat{\mathbf{h}}_u[k]/\|\hat{\mathbf{h}}_u[k]\|$. The quantized channel direction $\mathbf{g}_u(k)$ is fed back to the transmitter via a delay and error free limited feedback channel. We use feedback rate to mean the number of feedback bits required at each update. In this regime, the transmitter collects the quantized channel vectors $\mathbf{g}_u(k)$ to form the composite channel matrix $\tilde{\mathbf{H}}(k) = [\mathbf{g}_1^H(k) \cdots \mathbf{g}_U^H(k)]$. The beamforming vector for each user $u$ is computed as $\tilde{\mathbf{v}}_u[k] = \tilde{\mathbf{H}}^H_u[k]/\|\tilde{\mathbf{H}}^H_u[k]\|$.  

### 2.2. Throughput Performance

The first term in (1) is the desired signal for the $u$-th user and the second summation term is the interference signal, created because of quantization error. For the purpose of analysis, the signal to interference plus noise ratio (SINR) for the $u$-th user can be written

$$
\text{SINR}_u = \frac{\frac{P}{T} \|\mathbf{h}_u^*\mathbf{v}_u\|^2}{1 + \sum_{n \neq u} \frac{P}{T} \|\mathbf{h}_u^*\mathbf{v}_n\|^2}, \tag{2}
$$

If the transmit signal $s_u$ is assumed to be Gaussian, the achievable throughput for user $u$ is given by

$$
\mathcal{R}_u = \log_2(1 + \text{SINR}_u) \tag{3}
$$

and the sum rate as $\mathcal{R} = \sum_{u=1}^U \mathcal{R}_u$. In this paper, we will use the sum rate with various feedback rate as one of the performance measures for our proposed algorithm. In the limited feedback regime, the SINR can be rewritten using the channel direction as

$$
\text{SINR}_u = \frac{\frac{P}{T} \|\mathbf{h}_u\|^2 \|\mathbf{g}_u\| \|\mathbf{v}_u\|^2}{1 + \sum_{n \neq u} \frac{P}{T} \|\mathbf{h}_u\|^2 \|\mathbf{g}_u\| \|\mathbf{v}_n\|^2}. \tag{4}
$$

We may observe that due to the absolute value around $\mathbf{g}_u^H \mathbf{v}_u$, the SINR is independent of arbitrary unitary rotations of the channel direction, i.e., $|\mathbf{g}_u^H \mathbf{v}_u|^2 = |\mathbf{g}_u^H \mathbf{v}_u e^{j\theta}|^2$ for $\theta \in (0, 2\pi)$. Therefore, we may correspond the space of channel direction to the Grassmannian manifold. Thus, our problem is to feedback channel directions on the Grassmann manifold from each user $u$, and use the collected CSI at the transmitter for zero forcing.

### 3. GRASSMANNIAN PREDICTIVE CODING

In this section, we provide an overview of the geometry of Grassmann manifold, outline the proposed Grassmannian predictive coding for limited feedback MU-MIMO systems, and briefly discuss the design of codebooks for tangential prediction error.

#### 3.1. Preliminaries

Under the sum rate performance measure (3) and (2), the channel direction vectors can be identified as living on the Grassmann manifold $G_{N_t,1}$; a set of one dimensional subspaces in $N_t$ dimensional Euclidean space [10]. The Grassmann manifold has non-Euclidean geometry [10]. Thus we begin by defining the basic notion of distance. The most commonly used distance metric on $G_{N_t,1}$ is the chordal distance given by the sine of the subspace angle between two points, $\mathbf{x}, \mathbf{y} \in G(N_t,1)$, as

$$
d(x, y) = \sin(\theta) = \sqrt{1 - \|\mathbf{x}^H \mathbf{y}\|^2}. \tag{5}
$$

Using the chordal distance metric, we define the correlation of two sequences $\{\mathbf{x}[k]\}_{k \in \mathbb{N}}, \{\mathbf{y}[n]\}_{n \in \mathbb{N}} \in G_{N_t,1}$ by $\mathcal{C}_x(y(n)) = \mathbb{E}_k[d(\mathbf{x}[k], \mathbf{y}[k + n])]$ which can be interpreted as the mean chordal distance between two sequences on the Grassmann manifold.

Based on the smooth manifold structure of the Grassmannian, it is possible to relate two points $\mathbf{x}[k], \mathbf{x}[k+1] \in G_{N_t,1}$ by considering the tangent vector emanating from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$. Let $\rho = \mathbf{x}[k][\mathbf{x}[k+1]]$ denote the inner product which is related to the instantaneous subspace angle as $d = d(\mathbf{x}[k], \mathbf{x}[k+1]) = \sqrt{1 - \|\rho\|^2}$.

**Lemma 1 (Tangent)** If $\mathbf{x}[k], \mathbf{x}[k+1] \in G_{N_t,1}$, then the tangent vector emanating from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$ is

$$(\mathbf{e}[k+1] = \frac{1}{\text{tan}^{-1}(d/|\rho|)} \cdot \frac{\mathbf{x}[k+1] - \mathbf{x}[k]}{|\mathbf{x}[k+1] - \mathbf{x}[k]|} = \mathbf{e}[k+1]/|\mathbf{e}[k+1]| \tag{5}
$$

such that $||\mathbf{e}[k+1]||_2 = \text{tan}^{-1}(d/|\rho|)$ is the arc length between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$ and $\mathbf{e}[k+1] = (\mathbf{x}[k+1]/|\mathbf{x}[k+1]|)/(d/|\rho|)$ is the unit tangent direction vector.

Due to space limitations, all the theorems are provided without proof (please see [11] for details). For notational brevity, we denote the tangent operation by $\mathbf{e}[k+1] = L(\mathbf{x}[k], \mathbf{x}[k+1])$. The tangent is conveniently expressed as the product of magnitude component and the normalized directional component. We note that the error tangent vector is, by construction, orthogonal to the vector at the base.

Given a tangent vector $\mathbf{e}[k+1]$ with respect to $\mathbf{x}[k]$, the tangent can be mapped onto the manifold using a one parameter map.

**Lemma 2 (Geodesic)** If $\mathbf{x}[k], \mathbf{x}[k+1] \in G_{N_t,1}$ and $\mathbf{e}[k+1]$ is the tangent vector emanating from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$, then the geodesic path between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$ is

$$
G(\mathbf{x}[k], \mathbf{e}[k+1], t) = \mathbf{x}[k] \cos(|\mathbf{e}[k+1]| |2t| + \mathbf{e}[k+1] \sin(|\mathbf{e}[k+1]| |2t|) \tag{6}
$$

for $t \in [0, 1]$ such that $G(\mathbf{x}[k], \mathbf{e}[k+1], 0) = \mathbf{x}[k]$ and $G(\mathbf{x}[k], \mathbf{e}[k+1], 1) = \mathbf{x}[k+1]$.

The parameter $t$ can be used to map to points on the manifold along the geodesic, i.e., shortest distance path, between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$. A tangent expression with respect to $\mathbf{x}[k+1]$ such that it follows the path between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$ is needed to make a one step prediction. This is accomplished by the parallel transport.
Lemma 3 (Parallel Transport) Let \( x[k], x[k + 1] \in G_{n,1} \) and \( e[k + 1] \) be the tangent vector emanating from \( x[k] \) to \( x[k + 1] \). Then, the parallel transported tangent vector emanating from \( x[k + 1] \) along the geodesic direction \( e[k + 1] \) is \( e[k + 2] = \tan^{-1} \left( \frac{d}{\rho} \right) \left( [x[k + 1] - x[k]] \right) / d \). 

The parallel transport transforms the tangent vector onto another tangent space connected by the geodesic. The next theorem shows the one step prediction while preserving the direction and distance.

Theorem 4 Let \( x[k], x[k - 1] \in G_{n,1} \). The one step predicted vector \( \hat{x}[k + 1] \in G_{n,1} \) along the geodesic direction from \( x[k] \) to \( x[k + 1] \) is \( \hat{x}[k + 1] = [\hat{\rho}x[k] + \rho^* x[k] - x[k - 1]] \) such that 
\[
d(x[k], \hat{x}[k + 1]) = d(x[k - 1], x[k]).
\]

For notational brevity, we denote the prediction operation by a map 
\( P : G_{n,1} \times G_{n,1} \rightarrow G_{n,1} \) which takes current and previous estimate vectors and outputs the predicted vector, i.e., \( \hat{x}[k + 1] = P(x[k - 1], x[k]) \). The surprising outcome of Theorem 4 is that the predicted vector \( \hat{x}[k + 1] \) can be easily computed by the knowledge of \( x[k - 1] \) and \( x[k] \) using linear operations.

Algorithm 1 GPC encoder algorithm
Input: \( x[k] \)
1: Initialize \( \hat{x}[1] \) and \( \hat{x}[0] \)
2: for all \( k = 1, 2, \ldots \) do
3: \( e[k] = L(\hat{x}[k], x[k]) \)
4: \( i[k] = Q(e[k]) \)
5: \( \hat{x}[k] = G(\hat{x}[k], c_i[k], 1) \)
6: \( \hat{x}[k + 1] = P(\hat{x}[k - 1], \hat{x}[k]) \)
7: end for
Output: \( i[k] \)

Algorithm 2 GPC decoder algorithm
Input: \( i[k] \)
1: Initialize \( \hat{x}[1] \) and \( \hat{x}[0] \)
2: for all \( k = 1, 2, \ldots \) do
3: \( c_i[k] = Q^{-1}(i[k]) \)
4: \( \hat{x}[k] = G^{-1}(\hat{x}[k], c_i[k], 1) \)
5: \( \hat{x}[k + 1] = P(\hat{x}[k - 1], \hat{x}[k]) \)
6: end for
Output: \( \hat{x}[k] \)

3.2. Proposed Algorithm
Let \( \{x[k]\} \in G_{n,1} \) be the input sequence, i.e., normalized channel vectors, with time index \( k \). The proposed algorithm closely follows that of classical PVQ with specialized manifold operations [3].

The main idea of predictive coding is to quantize the tangential error \( e[k] \) between the predicted vector \( \hat{x}[k] \) and the current observed vector \( x[k] \). Then, the quantized tangent error is used to construct the estimate \( \hat{x}[k] \) of the current observed vector. The current and previous estimated vectors, \( \hat{x}[k] \) and \( \hat{x}[k - 1] \), are used to compute the predict vector \( \hat{x}[k + 1] \).

The pseudo code for the encoder at the receiver side is shown in Algorithm 1. At time \( k \), using (5), the error tangent vector \( e[k] \) emanating from \( x[k] \) to \( x[k] \) is computed. If \( C = \{c_i\}_{i=1}^{N_C} \) is the size of \( N_C = 2^N \) codebook of error tangent vectors, the error tangent vector is quantized as
\[
i[k] = \arg \min_{i \in \{1, 2, \ldots, N_C\}} d(G(\hat{x}[k], c_i, 1), x[k]).
\]

The codeword that yields the geodesic map with shortest distance to the observed vector \( x[k] \) is selected. For notational brevity, we denote the quantization step by \( Q : C^2 \rightarrow \mathbb{N} \) that takes the error tangent vector and outputs the codeword index, i.e., \( i[k] = Q(e[k]) \). In the conventional PVQ design, one of the strategy for codebook design is to employ open-loop approach followed by closed-loop approach to refine the codebook [3]. We employ the same approach in this paper. The open loop approach uses the prior vectors from training sequence to perform the prediction \( \hat{x}[k] = P(x[k - 1], x[k]) \) instead of on the estimated vectors \( x[k - 1] \) and \( x[k] \). The error tangent vector is computed using (5). Then the Lloyd iteration is used to obtain the open-loop codebook. The closed-loop approach uses the open-loop codebook with GPC on the training sequence to obtain another set of error vectors. Again, the Lloyd iteration is performed on the closed-loop error tangent vectors to obtain the final codebook.

Unfortunately, it is difficult to show the Lloyd iteration optimality of the open-loop and closed-loop approaches due to the feedback structure of the GPC but these approaches have been known to provide good results in the PVQ literature [3]. Thus, in this paper, we employ the open-loop and closed-loop approach to obtain the error tangent vector codebook.

The indices are transmitted to the decoder via a finite rate communication channel. Continuing at the encoder, the knowledge is available at the decoder. The received indices are decoded in \( Q^{-1} \) to recover \( c_i[k] \). The predicted vector \( \hat{x}[k] \) is mapped to the estimated vector using the codewords as in (8). The decoder output is \( \hat{x}[k] \) which is used for zero forcing.

Using the estimates, prediction is also performed to obtain \( \hat{x}[k + 1] \) for the next time period. Note that for the first iteration of the decoder, the knowledge of \( \hat{x}[k] \), or equivalently \( \hat{x}[k - 1] \) and \( \hat{x}[k - 2] \), is needed. Initializing the vectors with the encoder is important because if \( \hat{x}[k] \) is different from the encoder, the received error tangent direction and magnitude no longer represents the error tangent vector. For example, initialization can be accomplished through single-shot quantization with a Grassmannian beaming vector [11].

4. SIMULATION RESULTS
In this section, we illustrate the performance of our proposed algorithm for \( N_t = U = 4 \). We assume temporally correlated vector channel modelled according to a first order autoregressive model (or Gauss-Markov model) with correlation coefficient \( \alpha = J_0(2\pi f_D T_s) \) where \( J_0 \) is Bessel function of zeroth order, \( f_D \) is the Doppler frequency, and \( T_s \) is the sampling interval [6]. The product \( f_D T_s \) is called the normalized Doppler frequency. Temporally correlated channel is generated according to \( h_n[k] = \alpha h_n[k - 1] + \sqrt{1 - \alpha^2} z[k] \) where \( z[k] \) is an i.i.d. zero-mean, spatially white complex Gaussian random vector. Each user’s channel is assumed to have the same correlation.

Fig. 1 shows the sample chordal distance between \( \hat{x}[k] \) and \( x[k] \) using the proposed algorithm and the chordal distance between the
Comparison of Chordal Distance Error Over Time

Fig. 1. Chordal distance comparison over time between memoryless quantizer using 9-bit Grassmannian codebook and the proposed algorithm using 9-bit tangent codebook.

quantized vector and the observed vector for memoryless quantization using Grassmannian codebook with the same feedback rate of 9-bits. Normalized Doppler frequency of $f_D T_s = 0.001$ is used. Substantial improvement in the quantization accuracy compared with memoryless technique is obtained.

Fig. 2 illustrates the achievable sum rate estimate obtained with i.i.d. Gaussian channel with perfect CSI at the transmitter, 9-bit random codebook memoryless limited feedback approach, and the proposed limited feedback using 9-bit tangent codebook for various normalized Doppler frequency $f_D T_s$. Contrary to the memoryless strategy, the proposed algorithm provides significant sum rate gain. In fact, for $f_D T_s = 0.001$, the system starts to become interference limited above SNR of 20dB illustrating the superior CSI accuracy when the channel is highly correlated. Furthermore, each user is equipped with the same codebooks which eliminates the need to store multiple codebooks at the transmitter, thus reducing the overhead for practical applications.

5. CONCLUSION

In this paper, we proposed a new predictive coding algorithm on the Grassmann manifold for application to limited feedback MU-MIMO systems. We exploited the differential geometric structure of the Grassmann manifold to compute a tangential error. We also proposed an efficient technique to quantize the error. We showed that the proposed feedback strategy improves the accuracy of the CSI at the transmitter, with comparable feedback overhead, thus providing higher sum rates compared to prior methods. Future work will also consider delay in the feedback channel.

6. REFERENCES