Predictive Quantization and Joint Time-Frequency interpolation technique for MIMO-OFDM precoding

Agrim Gupta*, Kumar Appaiah*, Rahul Vaze†

*Department of Electrical Engineering, Indian Institute of Technology Bombay

† School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai

Abstract

Precoding transmissions in wireless MIMO systems is essential to enable optimal utilization of the spatial degrees of freedom. However, communicating the precoding matrices to the transmitter from the receiver is challenging, owing to large feedback requirements. Past work has shown that predictive quantization in time, as well as interpolation over frequency can be used for effective feedback of the precoders. Moreover, interpolation strategies for MIMO systems where the number of antennas at the transmitter (N_T) is different from the number of antennas at the receiver (N_R), required quantization of redundant information, thereby increasing feedback overheads. Our key contributions cover two aspects. First, we propose an efficient interpolation strategy for systems with $N_T \neq N_R$. Building upon this, we propose a predictive quantization as well as a joint time-frequency interpolation strategy for such precoding matrices. The key insight we use is that local tangent spaces in the underlying manifold structure of the precoder matrices permit effective combination of both temporal and frequency domain information for more accurate precoder reconstruction. Simulations reveal that we obtain a significant improvement in achievable rate as well as BER reduction when compared to existing strategies.

I. Introduction

The use of orthogonal frequency division multiplexing (OFDM) in multi-antenna ($N_T \times N_R$) wireless systems allows the treatment of a frequency-selective channel as a frequency-flat multiple-input multiple-output (MIMO) channel on a per subcarrier basis. This enables low-complexity

receiver implementations, particularly for equalization. To maximally utilize the benefit of having MIMO antennas, the receiver needs to feed back channel state information (CSI) to the transmitter, which then uses it for precoding (matrix transformation of the input signal). Precoding enables efficient allocation of power across transmit antennas by waterfilling, and efficient signaling to increase rate while reducing bit-error-rate (BER) [1].

Modern wireless OFDM systems support thousands of subcarriers, thus the CSI requirement for precoding is rather large. However, only a certain number of subcarriers are reserved for CSI transmission in practice. Also, the receiver has to quantize these fed back matrices using just a few bits, due to the limited feedback constraint. In the MIMO case, since these precoding matrices are high dimensional entities, quantizing them with few bits incurs substantial quantization error. The key challenge at the transmitter is to both estimate the precoding matrices at the non-fed back subcarriers, and to counter the substantial quantization error that arises due to quantization with just a few bits. This necessitates exploiting the correlations across frequency for interpolation algorithms [2] to estimate precoding matrices at subcarriers in which the receiver did not provide feedback. To address the quantization error issue, the transmitter requires the receiver to utilize temporal correlations and perform predictive quantization, instead of just independently quantizing the matrices in each time frame. This enables lowering of quantization error with time and improves the performance of the MIMO system [3, 4].

One of the key challenges in designing the interpolation and predictive quantization algorithms is that the precoding matrices do not form a vector space, which renders linear algorithms for interpolation and predictive quantization ineffective. However, these matrices have an underlying manifold geometry [3–5], which can be exploited to design the required algorithms to interpolate and predict. A traditional method of obtaining the precoding matrix is to take the SVD of the channel matrix at the receiver and feed back the resulting quantized matrices which represent the directions and magnitudes of the eigenmodes, to the transmitter. Depending on the degrees of freedom that the transmitter has in allocating power to its N_T transmit antennas, the optimal precoders to maximize the achievable rate reside either on the Grassmannian manifold or the Stiefel Manifold [3]. When the transmitter treats each channel eigenmode in the same manner, viz. doesn't discriminate between the strongest and weakest channel eigenmodes, the optimal precoders reside on the Grassmannian manifold, since the required information for the receiver to feed back to the transmitter constitutes of the subspace formed by directions of the channel

eigenmodes. However, the transmitter can allocate the transmit power among its N_T transmit antennas via waterfilling to prioritize the benefits offered by stronger the eigenmodes. The receiver, in this case, needs to feed back the directions of the individual channel eigenmodes as well, and not only the subspace that they form. Hence, in the waterfilling case, the optimal precoders lie on the Stiefel manifold. In the waterfilling case, when $N_T = N_R$, the optimal precoders form unitary matrices and reside on the Unitary manifold. To enable waterfilling, the receiver also needs to feed back the magnitudes of individual channel eigenmodes, which form a vector space and allow for linear algorithms to interpolate and predict. In this paper, we work with the case in which transmitter has the degree of freedom in allocating transmit power via waterfilling. The optimal precoders, obtained via SVD of the channel matrix reside on the Stiefel Manifold, and we capture the temporal, frequency correlations by utilizing the underlying manifold structure.

Exploiting the manifold structure with local linear operations in the tangent space allows for algorithms to interpolate precoders at those subcarriers where feedback is not available at the transmitter. This also allows the receiver to improve upon the quantization error using predictive quantization methods. The most popular interpolation strategy for unitary precoders is the Geodesic scheme [2], which interpolates across two fed back unitary precoding matrices by joining them with the shortest path between them, and splitting the path into equal sections to obtain precoders at subcarriers where feedback is unavailable. Although not very explicit, this scheme also uses the linearity of the local tangent space. The role of local linear tangent spaces is more explicit in [3, 4], both of which first utilize the tangent spaces to make predictions for the new precoders at both the transmitter and the receiver, and then capture the channel innovation by quantizing the tangent space near the obtained prediction at the receiver. The reciever then feeds back the quantized direction in the tangent space to the transmitter so that the transmitter obtains a refined estimate for the new precoder from the predicted precoder.

Several past approaches have largely exploited frequency or temporal correlation in isolation for interpolation and prediction [3, 4, 6–11]. For the case where the precoding matrix is not square ($N_T \neq N_R$), temporal correlation has been used in [3] and an interpolation algorithm has been suggested in [5]. A non-manifold based approach to simultaneously capture frequency and temporal correlation is presented in [12]. Past work that has considered utilizing the manifold structure for interpolation algorithms has largely been concerned with Unitary manifolds [2, 8],

where the algorithms are implemented over the full $N_T \times N_T$ unitary space. When $N_T > N_R$, the effective dimensions are given by $N_T \times N_R$ matrices residing in the Stiefel manifold. Therefore, algorithms for limited feedback that operate over the complete unitary space are redundant. Stiefel manifold can be used to work with the effective dimensions of the optimal precoder matrices, thereby reducing the CSI overhead, although computing geodesics over the Stiefel manifold is challenging. In recent work, a Lloyd type codebook scheme has been suggested for direct quantization of matrices residing in the Stiefel manifold [13], and it has been shown that closed form unique quasi-geodesics can be computed over the Stiefel manifold [14]. We exploit both of these recent results to reduce the overhead requirement by quantizing at the receiver and interpolating at the transmitter, both while working with the effective dimensions of the Stiefel manifold.

In addition to working with the Stiefel manifold quasi-geodesics, we also focus on jointly exploiting the temporal and the frequency correlation for efficient limited CSI feedback. To the best of our knowledge, prior work has considered quantization over the Stiefel manifold only in the presence of temporal correlations [3], or in the frequency domain [5], but not jointly. Exploiting temporal and frequency correlation presents novel challenges, and we propose a predictive quantization and interpolation strategy that can be implemented over the Stiefel manifold, that builds upon the ideas presented in [3, 9]. The novel idea here is that the local tangent space, in addition to offering linearity for easier mathematical modeling, also allows for effective combination of both temporal and frequency domain information, which is extracted via past precoding matrices fed back to the transmitter. This allows for much better utilization of the limited feedback obtained in the past, for more effective quantization and interpolation, thereby significantly improving the system performance.

We provide extensive simulations to illustrate the benefits of our approach in terms of improving the BER, the achievable rate and channel estimation error (measured via the chordal distance). Our simulations for various mobility conditions as well as different wireless channel profiles reveal that, using 6 bits per fed back precoder for the 4×2 MIMO channel, we are able to reduce the E_b/N_0 requirement by around ~5 dB for the same BER, when compared to the earlier temporal prediction based approaches. In addition, the achievable rate we obtain using our limited feedback scheme is close to 95% of the rate achievable using perfect precoder knowledge at the every subcarrier at the transmitter.

The rest of the paper is organized as follows. Section II describes the system model and SVD based precoding schemes. Section III elaborates on interpolation directly on the Stiefel Manifold via the Cayley lifting map. We present a frequency hopping strategy, predictive quantization technique and joint time-frequency interpolation scheme for precoders on the Stiefel manifold in Section IV. Section V presents the simulation results and finally Section VI concludes.

II. System Model

We consider a point-to-point MIMO-OFDM wireless system that has N_T transmit antennas and N_R receive antennas. The available bandwidth is divided into N subcarriers such that each subcarrier has a nearly flat frequency response, as is commonly the case in the OFDM systems. The transmitter communicates an $N_s \times 1$ data vector, where $N_s \leq \min(N_T, N_R)$, and the $N_T \times N_s$ precoding matrix maps the $N_s \times 1$ data vector onto the $N_T \times 1$ transmit vector emanating out of the transmitter. In this discussion, we assume that $N_T > N_R$ and $N_s = N_R$. Keeping notations consistent with [3], the $N_R \times 1$ data stream received is denoted by

$$\mathbf{y}_{i,t} = \mathbf{H}_{i,t}^H \tilde{\mathbf{U}}_{i,t} \mathbf{x}_{i,t} + \mathbf{w}_{i,t}$$

where $\tilde{\mathbf{U}}_{i,t} \in \mathbb{C}^{N_T \times N_R}$ denotes the precoding matrix used by the transmitter, which is a function of quantized channel information fed back by the receiver, $\mathbf{y}_{i,t} \in \mathbb{C}^{N_R \times 1}$ is the received data stream, $\mathbf{x}_{i,t} \in \mathbb{C}^{N_R \times 1}$ denotes the transmitted signal at the i-th subcarrier ($i \in [0, \ldots N-1]$), of the t-th OFDM frame, $\mathbf{H}_{i,t} \in \mathbb{C}^{N_T \times N_R}$ denotes the MIMO channel matrix and $\mathbf{w}_{i,t}$ denotes the i.i.d. complex Gaussian noise with $\mathbf{w}_{i,t} \sim \mathcal{N}_{\mathbb{C}}(0, N_0 \mathbf{I}_{N_R})$ (N_0 is the noise variance). Using a MMSE decoder and the received $\mathbf{y}_{i,t}$, one obtains an estimate of $\mathbf{x}_{i,t}$, denoted by $\hat{\mathbf{x}}_{i,t}$. The system block diagram is presented in in Fig. 1.

We assume that the channel matrices $\mathbf{H}_{i,t}$ are estimated at the receiver exactly. If the exact channel matrices $\mathbf{H}_{i,t}$ are somehow made available at the transmitter, the optimal policy at the transmitter is to use the right singular vectors of $\mathbf{H}_{i,t}^H$ as the precoding matrices $\tilde{\mathbf{U}}_{i,t}$ to maximise the achievable rate. In particular, when we take compact SVD of $\mathbf{H}_{i,t}$, we get $\mathbf{H}_{i,t} = \mathbf{U}_{i,t} \Sigma_{i,t} \mathbf{V}_{i,t}$, and we have $\tilde{\mathbf{U}}_{i,t} = \mathbf{U}_{i,t}$. Since $\mathbf{U}_{i,t}$ is a lower dimensional entity than $\mathbf{H}_{i,t}$, it makes sense for the receiver to quantize and feed back $\mathbf{U}_{i,t}$ instead of $\mathbf{H}_{i,t}$, without losing out on the feedback benefits. Notice that matrices $\mathbf{U}_{i,t}$ reside on the Stiefel manifold $\mathrm{St}(N_T, N_R)$, since the columns of $\mathbf{U}_{i,t}$ form a set of N_R orthogonal vectors in N_T dimensions. Given practical limitations, only

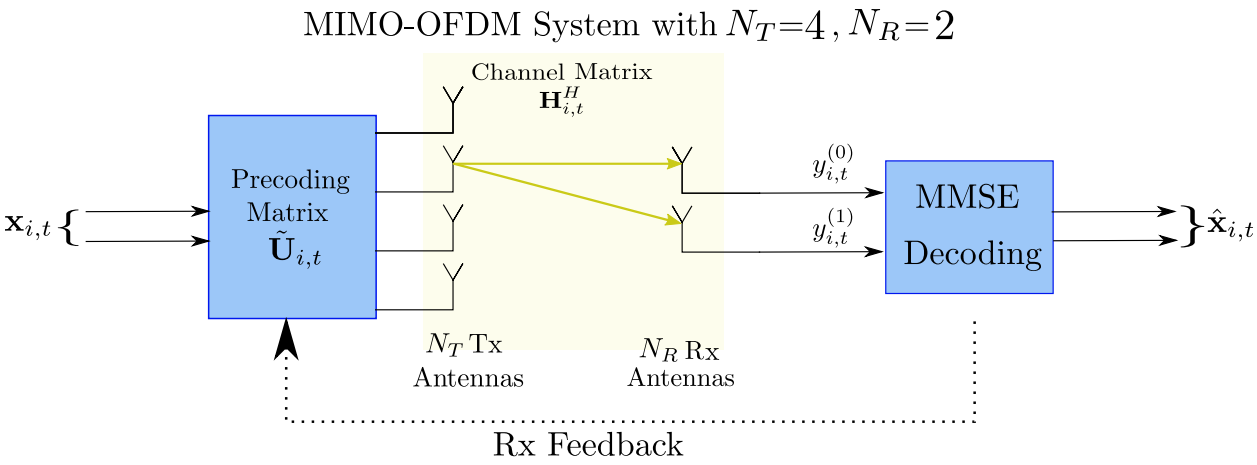


Fig. 1: Sample point-to-point MIMO-OFDM setting that we consider, with unequal number of transmit and receive antennas. The system shown is a 4×2 system, where the precoding matrix $\tilde{\mathbf{U}}_{i,t}$ is a 4×2 with $\tilde{\mathbf{U}}_{i,t}^H \tilde{\mathbf{U}}_{i,t} = \mathbf{I}_2$. Channel estimates at the receiver are used to feed back $\tilde{\mathbf{U}}_{i,t}$ to the transmitter.

limited feedback is available from the receiver, and the objective is to find 'reasonable' estimate of $\mathbf{U}_{i,t}$ at the transmitter, given by $\tilde{\mathbf{U}}_{i,t}$ using the available feedback bits.

Before proceeding, we discuss the geometry of the Stiefel manifold. The (compact) Stiefel manifold, $\mathrm{St}(m, n)$ represents set of all ordered m orthogonal vectors in the vector space $\mathbb{C}^n$ [3]:

$$\mathrm{St}(m, n) = \{\mathbf{U} \in \mathbb{C}^{n \times m} | \mathbf{U}^H \mathbf{U} = \mathbf{I}_m\}.$$

$\mathrm{St}(m, n)$ does not form a vector space, because of which linear algorithms to interpolate and quantize are not effective, as these operations will result in violation of the $\mathbf{U}^H \mathbf{U} = \mathbf{I}_m$ constraint. That is, to interpolate between $\mathbf{A}$ and $\mathbf{B}$, with $\mathbf{A}, \mathbf{B} \in \mathrm{St}(m, n)$, evaluating $\lambda \mathbf{A} + (1-\lambda)\mathbf{B}$, $\lambda \in (0, 1)$ yields a result that does not necessarily belong to $\mathrm{St}(m, n)$. Similarly, one can not use a linear prediction approach for predictive quantization. However, the tangent space local to a point $\mathbf{U} \in \mathrm{St}(m, n)$, defined as, $T_{\mathbf{U}}\mathrm{St}(m, n)$, forms a linear vector space that admits linear operations. The vector space property of the local tangent space facilitates prediction and interpolation algorithms for SVD precoders in the MIMO-OFDM setting.

We represent the considered limited feedback scenario by a time-frequency bins matrix (Fig. 2), where t-th row and i-th column corresponds to the precoding matrix at t-th OFDM frame and i-th subcarrier. The matrix is sparsely fed back (in terms of number of bins fed back) by the receiver to the transmitter, respecting the limited feedback constraint. The transmitter reconstructs the time-frequency bins matrix at non-fed back entries, by utilizing the temporal and frequency

correlations of the nearby fed back points. The metric we choose to quantify the closeness of a reconstruction of the time-frequency bins matrix at the transmitter to the exact time-frequency bins matrix observed at the receiver is $d[t] = \frac{1}{N}\sum_{i=0}^{N-1} d_s(\tilde{\mathbf{U}}_{i,t}, \mathbf{U}_{i,t})$, with d_s being the chordal distance between two points in Stiefel manifold, defined as,

$$d_s(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{j=1}^{N_R} d_g^2(\mathbf{a}_j, \mathbf{b}_j)}, \;\; d_g(\mathbf{u}, \mathbf{v}) = \sqrt{1 - |\mathbf{u}^*\mathbf{v}|^2}, \tag{1}$$

where $\mathbf{A}, \mathbf{B} \in \mathrm{St}(N_T, N_R)$, and $\mathbf{a}_j, \mathbf{b}_j$ are the j-th columns of $\mathbf{A}$, $\mathbf{B}$ respectively, and $d_g(\mathbf{a}_j, \mathbf{b}_j)$ is the Grassmannian chordal distance between $\mathbf{a}_j$ and $\mathbf{b}_j$ [15]. Observe that $d[t_1]$ at a particular time t_1 represents the average chordal distance over the N subcarriers, between the estimates and actual values of row t_1 in the time-frequency bins matrix. Thus, the objective is to come up with a strategy to reconstruct the sparsely fed back time-frequency bin matrix to minimize the distance metric $d[t]$ and bring the performance as close to complete exact (i.e accurate unquantized precoding matrix available at each bin) time-frequency bins matrix.

With the sparsely fed back time-frequency bins matrix from the receiver, and the individual fed back precoders incurring substantial quantization errors, exploiting time and frequency correlations is absolutely imperative for reducing $d[t]$. Independently utilizing the temporal and frequency correlations to reconstruct the time-frequency bins matrix at the transmitter can be performed using a frequency based interpolation scheme, coupled with a time based predictive quantization scheme. We use the prior work on predictive quantization over $\mathrm{St}(N_T, N_R)$ [3] and the interpolation scheme suggested over $\mathrm{St}(N_T, N_R)$ in Section III, to reconstruct the time-frequency bin matrix using the strategy shown in Fig. 2, by sending feedback on regularly spaced subcarrier indices. The strategy depicted in Fig. 2 forms a baseline of comparison for our proposed scheme which exploits time-frequency correlations jointly.

Our proposed joint frequency-time scheme is aided by a hopping strategy (Section IV), where the locations of fed back subcarriers are alternated with time. With the use of the hopping scheme, we effectively combine both temporal and frequency domain information in the tangent spaces of the Stiefel manifold, on which the precoders $\tilde{\mathbf{U}}_{i,t}$ lie (Section IV-A). With the predictive quantization scheme at the receiver, we show that the quantization chordal distance ($d_s(\tilde{\mathbf{U}}_{i,t}, \mathbf{U}_{i,t})$) decreases as t increases, viz. the receiver feeds back refined estimates of each row of time-frequency bins matrix in subsequent OFDM frame. In addition, we also propose a joint time-

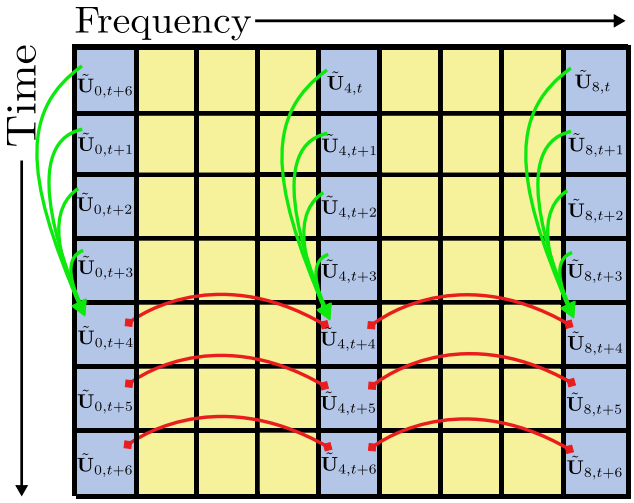


Fig. 2: Exploiting regularly spaced limited feedback scheme for $N = 9$, with (yellow) blue cells indicating the (non) fed back indices. Frequency correlations are captured via the red interpolating curves, and the temporal correlations are used to improve the quantization over time (shown implicitly via green arrows).

frequency interpolation scheme (Section IV-B) at the transmitter, to reconstruct the precoders at the missing subcarriers in every OFDM frame by effectively using the refined fed back precoders at the transmitter. We empirically show that jointly exploiting the frequency and temporal correlations improves $d[t]$ metric as compared to the independent utilization of the said correlations, which in turn allows for lower BER and higher achievable rate.

III. Proposed Interpolation on $St(N_T, N_R)$

In this section, we describe the limited feedback scheme that allows reconstruction of time-frequency bins matrix at the transmitter by exploiting the temporal and frequency correlations independently. To exploit temporal correlations, we utilize the past work in [3], which described a predictive quantization scheme on Stiefel manifold, although that was limited to only a flat fading channel. A natural approach to exploit frequency correlations by interpolating, would be to use geodesics, but true geodesic curves with closed form solutions in the Stiefel manifold are not known [5, 16]. Recently, [14] has proposed a unique quasi-geodesic curve using the Cayley type lifting-retraction maps that enable direct interpolation on $\mathrm{St}(N_T, N_R)$, which we utilize in our proposed interpolation scheme. Interpolation directly on $\mathrm{St}(N_T, N_R)$ has been discussed only in [5], albeit for unquantized precoders. An alternate approach towards interpolating on Stiefel manifold would be to project the matrix in the Stiefel manifold to the Unitary manifold, then use well known $\{\mathrm{expm}(\cdot) - \mathrm{logm}(\cdot)\}$ geodesic curve, and then project back to the Stiefel manifold.

Simulations reveal that quantization plays a key role in performance comparison of the proposed interpolation scheme on Stiefel manifold versus interpolating on the Unitary manifold.

In order to obtain a unitary precoding matrix, we take the full SVD of MIMO channel matrix observed at the receiver, instead of a compact SVD. The full SVD of $\mathbf{H}_{i,t}$ yields $\mathcal{U}_{i,t}\Sigma_{i,t}\mathbf{V}_{i,t}$, where $\mathcal{U}_{i,t} \in \mathbb{C}^{N_T \times N_T}$ is a unitary matrix, $\Sigma_{i,t} \in \mathbb{C}^{N_T \times N_R}$ is a diagonal matrix, and $\mathbf{V}_{i,t} \in \mathbb{C}^{N_R \times N_R}$ is a unitary matrix. The optimal precoder consists of the first N_R columns of $\mathcal{U}_{i,t}$ [2]. Selecting the first N_R rows projects the unitary matrix back onto the Stiefel manifold. We represent $\tilde{\mathcal{U}}_{i,t}$, as the matrix ultimately used by transmitter for precoding (and interpolating across non-fed back points), which is obtained by quantizing $\mathcal{U}_{i,t}$. The usual approach using geodesic interpolation on the Unitary manifold is as follows. Let us suppose that $\tilde{\mathcal{U}}_{i,t}$ have been fed back for subcarriers i_1 and i_2, $i_1 < i_2$. The transmitter interpolates the precoding matrix at i-th subcarrier, with $i_1 < i < i_2$ using the geodesic scheme in (2)

$$\tilde{\mathcal{U}}_{i,t} = \tilde{\mathcal{U}}_{i_1,t}\text{expm}\left(\frac{i}{i_2 - i_1}\text{logm}(\tilde{\mathcal{U}}_{i_1,t}^{-1}\tilde{\mathcal{U}}_{i_2,t})\right), \tag{2}$$

where expm and logm refer to the matrix exponential and matrix logarithm, respectively. The precoding matrices are then obtained by taking the first N_R columns of the $\tilde{\mathcal{U}}_{i,t}$ matrices obtained after interpolation, since we require only the first N_R right singular vectors. The remaining $N_T - N_R$ columns, which were included in the precoding matrix to form a unitary matrix and enable interpolation with geodesics, are actually redundant. Quantization with this redundant information would incur allocation of bits for unneeded information.

We now propose a method to directly interpolate over the Stiefel manifold, thereby obviating the need to quantize the redundant information. Let us suppose that $\tilde{\mathbf{U}}_{i,t}$ have been fed back for subcarriers i_1 and i_2, $i_1 < i_2$. We propose an interpolation method using the Cayley exponential lifting and retraction maps at $\tilde{\mathbf{U}}_{i_i,t}$, denoted as $\text{Exp}^{-1}_{\tilde{\mathbf{U}}_{i_1,t}}(\cdot)$ and $\text{Exp}_{\tilde{\mathbf{U}}_{i_1,t}}(\cdot)$ respectively (for explicit definitions refer Appendix Section VII-A), consistent with the notation in [14]. The transmitter interpolates the precoding matrix at i-th subcarrier, with $i_1 < i < i_2$ using (3)

$$\tilde{\mathbf{U}}_{i,t} = \text{Exp}_{\tilde{\mathbf{U}}_{i_1,t}}\left(\text{Exp}^{-1}_{\tilde{\mathbf{U}}_{i_1,t}}\left(\tilde{\mathbf{U}}_{i_2,t}\right)\right). \tag{3}$$

One thing to note is that the $\text{logm}(\cdot)$ maps $N_T \times N_T$ unitary matrices to $N_T \times N_T$ skew-Hermitian matrices. Similarly, the $\text{Exp}^{-1}(\cdot)$ in Equation 3 maps $N_T \times N_R$ matrices on $\text{St}(N_T, N_R)$ to $N_T \times N_T$ skew-Hermitian matrices, with the lower right $(N_T - N_R) \times (N_T - N_R)$ minor being the null matrix. A more detailed treatment is available in Appendix Section VII-A.

Using the approach described in (3), the receiver needs to quantize and feed back only $N_T \times N_R$ matrices $\mathbf{U}_{i,t}$, with $\mathbf{U}_{i,t}^H \mathbf{U}_{i,t} = \mathbf{I}_{N_R}$ instead of $N_T \times N_T$ unitary matrices $\mathcal{U}_{i,t}$, thereby obviating the need to quantize redundant information by working in lower dimensionsal Stiefel manifold. Improvement in performance (BER and achievable rate) stemming from interpolation in reduced dimensions is quantified by our simulations in Section V.

Alongside this, the interpolation scheme discussed in this section, coupled with predictive quantization scheme in [3], also allows us to populate the time-frequency bins matrix with the strategy depicted in Fig. 2, which forms a baseline of comparison for our proposed joint-time frequency predictive quantization and interpolation scheme, elaborated upon in the next section.

IV. Predictive Quantization and Joint Time-Frequency Interpolation

In this section, we describe a predictive quantization method employed by the receiver, coupled with an interpolation scheme at the transmitter to exploit both the temporal and the frequency correlations, which enables more efficient reconstruction of the time-frequency bins matrix with a lower $d[t]$ metric and improved performance (BER and achievable rate).

We first propose a hopping strategy to feed back the precoding matrices to the transmitter, instead of feeding back precoding matrices at regularly spaced subcarriers, as shown in Fig. 2. In the hopping strategy, the fed back precoding matrices' bins in the time-frequency bins matrix alternate with each OFDM frame. The hopping strategy allows the receiver to exploit frequency correlations, in addition to temporal correlations, for improved predictive quantization, enabling feedback to the transmitter with lower quantization error. The transmitter gets feedback with reduced quantization error, at time-alternated frequency positions in the time-frequency bins matrix, which it then uses to reconstruct the matrices at non fed-back entries of the time-frequency bins matrix, using the proposed joint time-frequency interpolation scheme. Hence, our proposed scheme improves upon both the aspects of the limited feedback precoding scheme, viz. the predictive quantization at the receiver and interpolation algorithm at the transmitter. This allows for more efficient utilization of the bits allocated for feed back of precoding matrices, as compared to the scheme in Fig. 2, which exploited temporal and frequency correlations independently instead of doing it jointly. This section describes the explicit details pertaining to both of these proposed methods.

A. Hopping and Predictive Quantization scheme

The approach for interpolation and prediction, shown in Fig. 2 is limited, in that the channel is fed back only at a fixed set of equally spaced out subcarriers, and the remaining subcarriers' precoders are interpolated. This method has two disadvantages. One is that the precoders for those subcarriers that lie in the middle of the subcarriers whose precoders have been fed back face larger interpolation errors, since they are equally far from the subcarriers where feedback is available. The other is that, when predictively quantizing solely over the time-domain, we ignore the useful past information present in the nearby subcarriers due to frequency correlations. We therefore suggest a "hopping" scheme, in which the fed back frequency bins alternate with each OFDM frame.

We now describe our hopping strategy both mathematically and using a visual illustration (Fig. 3). Let D_f be the frequency separation between two precoders estimated at the receiver, and δ_f be the frequency offset of the hopping scheme, both measured in terms of number of subcarriers (depicted in Fig. 3). We choose D_f as a factor of $N-1$, where N is the total number of subcarriers. We alternate between feeding back the precoder at subcarriers $\{mD_f | m = 0, 1, 2, \ldots (N-1)/D_f\}$ in one OFDM frame and subcarriers $\{\delta_f + mD_f | m = 0, 1, 2, \ldots (N-1)/D_f - 1\}$ in the next. We set $\delta_f = \lfloor D_f/2 \rfloor$ to address the issue of large interpolation errors in the middle of two fed back precoders. This permits us to obtain a more accurate reconstruction of the precoder variation across frequency, enabling more effective predictive quantization. The hopping strategy is visually illustrated in Fig. 3. The yellow boxes in Fig. 3 indicate the time-frequency bins for which the quantized $\tilde{\mathbf{U}}_{i,t} \in \mathrm{St}(N_T, N_R)$ were fed back. The white boxes in the same figure indicate the time-frequency bins for which the precoding matrices were not fed back. In the subsequent discussions, we refer to the value of the precoding matrix at subcarrier i at time t, $\tilde{\mathbf{U}}_{i,t}$, merely by referring to it as the precoder at box (i, t).

Using the above described hopping strategy, we now present our proposed predictive quantization scheme. For the t-th OFDM frame, the transmitter uses the past fed back information to obtain a prediction, say P for those boxes where the receiver would provide feedback in the next time frame. This situation is depicted in Fig. 3 for the $(t+6)$-th OFDM frame. The feedback from the receiver indicates the direction within the local tangent space of P, which the transmitter should take, so that it obtains the precoders closest in terms of $d[t]$ to the actual

precoders. The feedback helps transmitter capture channel innovation (viz. difference between predicted precoder and actual precoder), whereas the prediction utilizes the joint time-frequency correlations to minimize the innovation terms required (viz. reduce the distance as far as possible between the predicted and actual precoder).

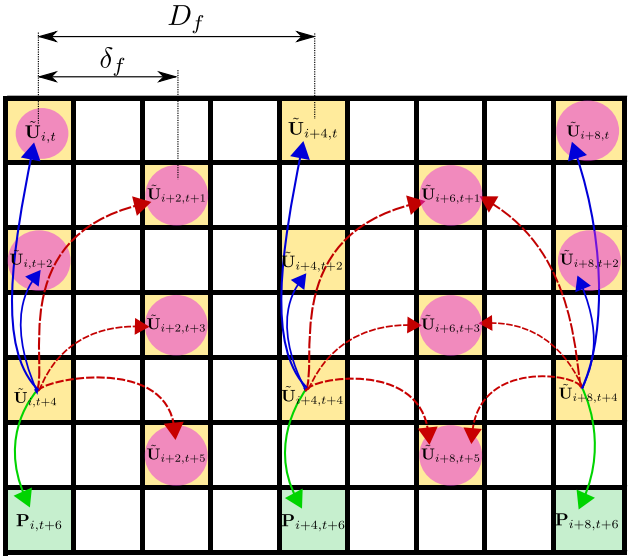


Fig. 3: Hopping strategy for $N = 9, D_f = 4, \delta_f = 2$. The red, blue and green curves represent the tangents on $\mathrm{St}(N_T, N_R)$, illustrated in Fig. 4.

For prediction at both the receiver and the transmitter, we use a lifting map to obtain local tangent spaces, optimize linear functions in the local tangent spaces, and then use a retraction map that corresponds to the chosen lifting map to return to the manifold. We obtain the tangent from box (i,t) to nearby box (j,s), denoted as $\mathbf{T}_{i,t}^{j,s}$, viz. the tangent emanating from $\tilde{\mathbf{U}}_{i,t}$ to $\tilde{\mathbf{U}}_{j,s}$ using the chosen lifting operation, denoted by $\mathbf{T}_{i,t}^{j,s} = \texttt{lift}_{\tilde{\mathbf{U}}_{i,t}}\left(\tilde{\mathbf{U}}_{j,s}\right)$. The corresponding retraction operation to map back to the manifold is given by $\texttt{retract}_{\tilde{\mathbf{U}}_{i,t}}\left(\mathbf{T}_{i,t}^{j,s}\right)$. Since the local tangent space forms a vector space, $\mathbf{T}_{i,t}^{j,s}$ can be approximated by a linear combination of two matrices that represent the tangent matrices in time and frequency domain separately, local to (i,t). In other words, there exist matrices $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$ such that $\mathbf{T}_{i,t}^{j,s} \approx \mathcal{F}_{i,t}\Delta f + \mathcal{T}_{i,t}\Delta t$, where $\Delta f, \Delta t \in \mathbb{Z}$ are small steps in the time and frequency axes respectively. This allows us to effectively combine information from temporal and frequency correlations, captured by $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$ respectively. In our case, we choose $\Delta f = j - i$ and $\Delta t = s - t$, i.e. the signed frequency/time separation between boxes (j,t) and (i,t). We estimate $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ using a least-squares fit over the known tangents to $(j,s) \in \texttt{nbrs}_{i,t}(p)$. $\texttt{nbrs}_{i,t}(p)$ (defined below) refers to the collection of neighbours (j,s) of

(i,t), which are lifted to the tangent space local to i,t for estimating $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$. Let,

$$\texttt{same\_freq\_nbrs}_{i,t}(p) = \{(i, t-2m) | m \in \{1 \dots p-1\}\}, \text{ and,}$$
$$\texttt{diff\_freq\_nbrs}_{i,t}(p,q) = \{(i+q\delta_f, t-2m+1) | m \in \{1 \dots p\}\}$$

where p is the number of past samples considered, $q = +1/-1$, indicating right/left neighbors. To ensure that $\mathbf{\Delta}_{i,t}$ is invertible, we take frequency separations only on one side, which gives,

$$\texttt{nbrs}_{i,t}(p) = \begin{cases} \texttt{same\_freq\_nbrs}_{i,t}(p) \bigcup \texttt{diff\_freq\_nbrs}_{i,t}(p,1), \text{ for } i \neq N-1 \\ \texttt{same\_freq\_nbrs}_{i,t}(p) \bigcup \texttt{diff\_freq\_nbrs}_{i,t}(p,-1), \text{ for } i = N-1 \end{cases} \tag{4}$$

where N is the total number of subcarriers, indexed from 0. In Fig. 3, the purple encircled boxes are $\texttt{nbrs}(\tilde{\mathbf{U}}_{i,t+4}(3))$ and $\texttt{nbrs}(\tilde{\mathbf{U}}_{i+8,t+4}(3))$ respectively. With the neighbouring precoders defined, we frame the following objective function to estimate $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$,

$$\mathcal{F}_{i,t}, \mathcal{T}_{i,t} \leftarrow \text{argmin} \sum_{(j,s) \in \texttt{nbrs}_{i,t}(p)} ||(\mathcal{F}_{i,t}(j-i) + \mathcal{T}_{i,t}(s-t) - \mathbf{T}_{i,t}^{j,s})||_F^2 \tag{5}$$

where $||\cdot||_F$ represents Frobenius norm. We take partial derivatives of the objective function in (5) with respect to $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ and set them to be the null matrix (for a detailed proof of the below equations, refer Appendix Section VII-B) to get the following expressions for $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ after optimizing the objective function:

$$\begin{bmatrix} \mathcal{F}_{i,t} \\ \mathcal{T}_{i,t} \end{bmatrix} = \mathbf{\Delta}_{i,t}^{-1} \begin{bmatrix} \Sigma_{j,s}(j-i)\mathbf{T}_{i,t}^{j,s} \\ \Sigma_{j,s}(s-t)\mathbf{T}_{i,t}^{j,s} \end{bmatrix}$$

$$\mathbf{\Delta}_{i,t} = \begin{bmatrix} \Sigma_{j,s}(j-i)^2 & \Sigma_{j,s}(j-i)(s-t) \\ \Sigma_{j,s}(j-i)(s-t) & \Sigma_{j,s}(s-t)^2 \end{bmatrix}$$

As an example, consider how to predict the precoding matrix at box (k,l): When the temporal correlation of the channel is higher than the frequency correlation (measured in terms of chordal distance to exact precoder at the receiver), $(i,t) = (k, l-2)$, and if the frequency correlation is larger, $(i,t) = (k \pm 2, l-1)$. The chosen (i,t) is likely to be the closest among the past fed back points, in terms of chordal distance, to the new box (k,l). The chosen (i,t) is referred to as the center (anchor point) on whose local tangent space we optimize to obtain $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ using the known neighbouring boxes (j,s) around the center (i,t). We thus obtain the predicted precoder at box (k,l) given by $\mathbf{P}_{k,l} = \text{retract}(\tilde{\mathbf{U}}_{i,t}, \mathcal{F}_{i,t}(k-i) + \mathcal{T}_{i,t}(l-t))$. $\mathbf{P}_{k,l}$ is the outcome of the

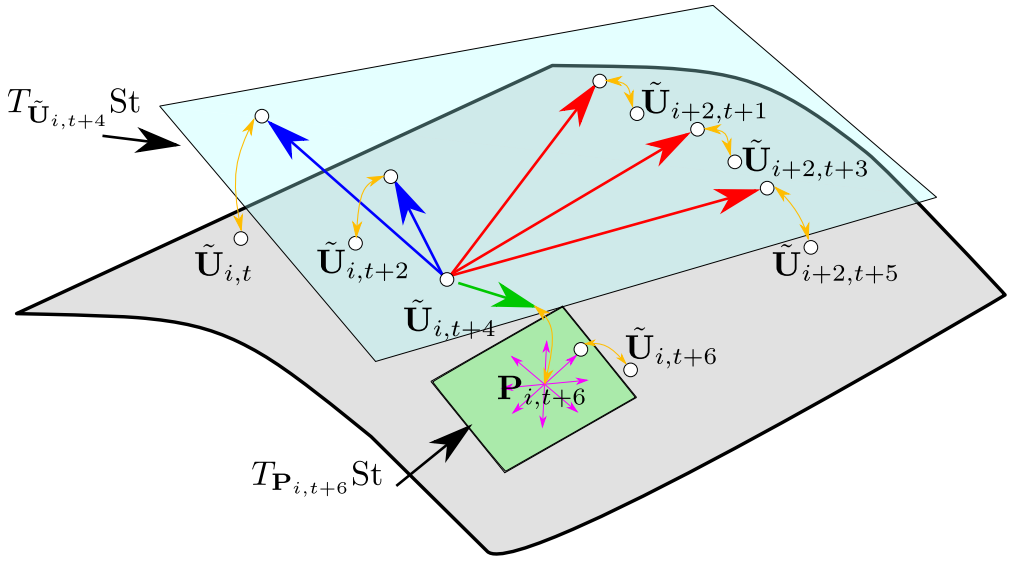


Fig. 4: Predictive Quantization algorithm to obtain $\tilde{\mathbf{U}}_{i,t+6}$, with closest previous value being $\tilde{\mathbf{U}}_{i,t+4}$. The yellow curves represent the lifting/retraction operations. $\mathbf{P}_{i,t+6}$ is the predicted value of $\tilde{\mathbf{U}}_{i,t+6}$. The blue and green planes are the tangent spaces local to $\tilde{\mathbf{U}}_{i,t+4}$ and $\mathbf{P}_{i,t+6}$ respectively. The local tangent space at $\mathbf{P}_{i,t+6}$ is then quantized and fed back to obtain $\tilde{\mathbf{U}}_{i,t+6}$.

predictive quantization algorithm for box (k,l). Note that these operations can be performed at the transmitter independently without any additional feedback.

Having described the proposed prediction approach, we now describe the feed back approach which allows the receiver to signal the channel innovation, so that transmitter can obtain a refined estimate of the actual precoder from the prediction it had made for the same (using information that was fed back in the past to the transmitter). Once we obtain the prediction $\mathbf{P}_{k,l}$, the receiver quantizes the tangent space local to $\mathbf{P}_{k,l}$ using a $\mathcal{B}$ bit codebook and feeds back the optimal index, which the transmitter uses to obtain the matrix which it ultimately uses for precoding, i.e. $\tilde{\mathbf{U}}_{k,l}$. The approach used by us to quantize the tangent space local to the prediction is based on previous work for the same in [3, 4]. A codebook quantizing the local tangent space would have two components, quantization of the directions in the tangent space, and quantization of magnitudes of steps taken in the quantized directions. We adopt the strategy in [3], which controls the magnitude of tangent steps, by having two codebooks $T_p^{\mathcal{C}}, T_m^{\mathcal{C}}$ of different spreads, but same $2^{\mathcal{B}-1}$ base vectors $\in T_{\mathbf{P}_{k,l}}\mathrm{St}(N_T,N_R)$. The base vectors correspond to $2^{\mathcal{B}-1}$ quantized directions, and by spread of codewords, we mean the volume covered by the codebook in the tangent space, with $T_p^{\mathcal{C}}$ covering more volume than $T_m^{\mathcal{C}}$. The two codebooks are concatenated to form a $2^{\mathcal{B}}$ length codebook $T^{\mathcal{C}}$. The receiver finds the optimal index c_n in $T^{\mathcal{C}}$ by comparing the

chordal distance given by (1), of each codeword to the actual precoder $\mathbf{U}_{k,l}$ obtained from the compact SVD of the channel matrix. using (6). The receiver then feeds back c_n to the transmitter using $\mathcal{B}$ bits. The transmitter uses the fed back c_n and (7) to calculate $\tilde{\mathbf{U}}_{k,l}$,

$$c_n \leftarrow \text{argmin}_{i \in 2^{\mathcal{B}}} \left(d_s \left(\mathbf{U}_{k,l}, \text{retract}(\mathbf{P}_{k,l}, T^C[i]) \right) \right) \tag{6}$$

$$\tilde{\mathbf{U}}_{k,l} = \text{retract}(\mathbf{P}_{k,l}, T^C[c_n]) \tag{7}$$

We now describe the algorithm used to control the spread of the codebooks, as discussed in [3]. The base codebook ($\mathcal{T}^B$) consists of $2^{\mathcal{B}-1}$ matrices belonging to the local tangent space at $\mathbf{P}_{k,l}$, which represent the $2^{\mathcal{B}-1}$ quantized directions. The two codebooks, T_p^C, T_m^C, have the same set of base vectors, but different spreads g^{s_p} and g^{s_m}, where g is the growth factor and s_p, s_m control the spread of the two codebooks. Depending on whether $c_n \geq 2^{\mathcal{B}-1}$, i.e. whether the optimum codeword is in T_p^C or T_m^C, the scale parameter $s[k]$, which in turn controls values of s_p, s_m is updated in the following manner,

$$s_p = g^{\min(s[k-1]+1,0)}, s_m = g^{s[k-1]-1}$$

$$s[k] = \begin{cases} \min(s[k-1]+1, 0), \text{ for } c_n \in T_p^C \\ s[k-1]-1, \text{otherwise} \end{cases}$$

with $s[0] = 0$. Intuitively, the algorithm reduces/increases the spread of the codebook till the operation of reduction/increase is no longer beneficial, i.e. the optimum codeword lies in the higher/lower spread codebook instead. The scheme we use for obtaining the quantized directions, however, differs slightly from the one presented in [3]. As discussed in Section III, the Cayley type lifting map takes inputs from the Stiefel manifold and maps them to $N_T \times N_T$ skew Hermitian matrices. This allows for vector quantization of the tangent space, since a skew Hermitian matrix can be easily converted to a vector, by taking the upper triangular entries and stacking them in a vector. Vector quantization for the base codebook, $\mathcal{T}^B$ allows for isotropic $2^{\mathcal{B}-1}$ quantized directions for the tangent space local to $\mathbf{P}_{k,l}$. [3] proposed a randomly initialized $\mathcal{T}^B$ which does not guarantee isotropic $2^{\mathcal{B}-1}$ quantized directions, which can potentially hurt the performance of the system in the cases when random initialization gives directions which are close to each other. More details on codebook construction are provided in Section V. This concludes the proposed predictive quantization algorithm.

The discussion thus far has not assumed a particular lifting-retraction map, and is expected to work for any appropriate lifting-retraction pair. One point to note is that, when considering (i,t) that is closest to the feedback subcarrier (k,l), we perform averaging as suggested in [17] and also used in [3] with $\tilde{\mathbf{U}}_{i,t}$ as the initial center of the averaging algorithm. Averaging over the quantization errors in the anchor point of prediction algorithm, i.e. $\tilde{\mathbf{U}}_{i,t}$, by using appropriate neighbour $\tilde{\mathbf{U}}_{j,s}$ improves the performance of the prediction algorithm. Since the averaging algorithm was suggested for the same orthographic lifting map proposed in [17], for the predictive algorithm, we use the orthographic lifting map. However, the orthographic lifting map has a higher dimension than the Cayley type maps discussed in Section III. Therefore, we use the Cayley type map for the vector quantization of the tangent space.

B. Joint Time-Frequency Interpolation scheme

Once the frequency-time bin matrix has been filled according to the hopping and predictive quantization scheme proposed in Section IV-A, the next step at the transmitter is to fill in the non fed back boxes (k,l) in the time-frequency bins matrix. This is also done by estimating the unit step matrices, viz. $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ at the closest fed back subcarrier (i,t) to subcarrier (k,l) by least-squares optimization. To interpolate precoding matrix at (k,l), we use future nearby feed back boxes of (i,t), as shown in Fig. 5, since past fed back boxes' information has already been captured in the predictive quantization algorithm. Define,

$$\texttt{l\_nbrs}_{i,t}(p) = \texttt{same\_freq\_nbrs}_{i,t}(-p) \bigcup \texttt{diff\_freq\_nbrs}_{i,t}(-p,-1),$$
$$\texttt{r\_nbrs}_{i,t}(p) = \texttt{same\_freq\_nbrs}_{i,t}(-p) \bigcup \texttt{diff\_freq\_nbrs}_{i,t}(-p,1)$$

as the left/right neighbors at (i,t) (marked in dark/light green, purple for $(i+3,t+1)$ in Fig. 5). We obtain different maps for left frequencies $\mathcal{F}^L_{i,t}, \mathcal{T}^L_{i,t}$ from $\texttt{l\_nbrs}_{i,t}(p)$, and right frequencies $\mathcal{F}^R_{i,t}, \mathcal{T}^R_{i,t}$ from $\texttt{r\_nbrs}_{i,t}(p)$, since directly obtaining $\mathcal{F}_{i,t}$ makes $\mathbf{\Delta}_{i,t}$ singular. Hence, we get the following objective functions, for obtaining $\mathcal{F}^L_{i,t}, \mathcal{T}^L_{i,t}, \mathcal{F}^R_{i,t}, \mathcal{T}^R_{i,t}$

$$\mathcal{F}^L_{i,t}, \mathcal{T}^L_{i,t} \leftarrow \text{argmin} \left(\sum_{(j,s)\in \texttt{l\_nbrs}_{i,t}(p)} \left(||(\mathcal{F}_{i,t}(j-i) + \mathcal{T}_{i,t}(s-t) - \mathbf{T}^{j,s}_{i,t})||^2_F \right) \right)$$
$$\mathcal{F}^R_{i,t}, \mathcal{T}^R_{i,t} \leftarrow \text{argmin} \left(\sum_{(j,s)\in \texttt{r\_nbrs}_{i,t}(p)} \left(||(\mathcal{F}_{i,t}(j-i) + \mathcal{T}_{i,t}(s-t) - \mathbf{T}^{j,s}_{i,t})||^2_F \right) \right)$$

Hence, the interpolated estimates for non-fed back boxes (k, l) at the transmitter are given by,

$$\tilde{\mathbf{U}}_{k,l} = \begin{cases} \text{retract}(\tilde{\mathbf{U}}_{i,t}, \mathcal{F}^L_{i,t}(k-i) + \mathcal{T}^L_{i,t}(l-t)), \text{ for } k < i \\ \text{retract}(\tilde{\mathbf{U}}_{i,t}, \mathcal{F}^R_{i,t}(k-i) + \mathcal{T}^R_{i,t}(l-t)), \text{ for } k \geq i \end{cases} \tag{8}$$

The lifting-retraction maps used for interpolation in (8) are the Cayley exponential type lifting maps discussed in Section III. The fed back boxes (i, t) by the receiver obtained by predictive quantization act as centers ($\tilde{\mathbf{U}}_{i,t}$ in (8)) to obtain the interpolated estimate of boxes $(k, l) \in \texttt{cluster}(i, t)$, which is defined as

$$\texttt{cluster}(i,t) = \begin{cases} \left\{(i-\delta_f+p, t+q) | p \in \{0,1,\ldots 2\delta_f\}, q \in \{0,1\}\right\} \setminus (i,t), \text{ for } i \neq \{0, N-1\} \\ \left\{(i+p, t+q) | p \in \{0,1,\ldots \delta_f\}, q \in \{0,1\}\right\} \setminus (i,t), \text{ for } i = 0 \\ \left\{(i-\delta_f+p, t+q) | p \in \{0,1,\ldots \delta_f\}, q \in \{0,1\}\right\} \setminus (i,t), \text{ for } i = N-1 \end{cases}$$

The joint time-frequency interpolation method is visually illustrated in Fig. 5. Having described the strategies used by the receiver to quantize and feed back precoding matrices, and at the transmitter to reconstruct the time-frequency bins matrix, we now explain the strategy used by the receiver to quantize and the transmitter to interpolate the singular values arising from SVD. For optimal power allocation via waterfilling, the singular values should also be fed back to the transmitter. We quantize these using vector quantization and feed them back on regularly spaced subcarrier frequency indices, such as $0, 33, (33 \times 2), \ldots, (33 \times 31 = 1023)$ for 1024 subcarriers. Predictive quantization is not performed for the singular values. At the receiver, these are interpolated at the unknown subcarriers using linear interpolation,

$$\sigma_i = \left(\sigma_{\text{prev(i)}}(\text{next}(i) - i) + \sigma_{\text{next(i)}}(i - \text{prev}(i))\right) / (\text{next}(i) - \text{prev}(i))$$

where $\text{next}(i), \text{prev}(i)$ indicate the previous/next fed back point corresponding to non fed back point i, σ_j denotes the singular values corresponding to subcarrier j.

V. Simulation Results

We simulated our algorithms by using the IT++ library through the Python wrapper pyitpp [18]. Channels were generated using Jake's model with both ITU Vehicular-A and ITU Pedestrian-A profiles, mentioned as appropriate in the following section. The simulations have been performed for $N_T = 4$ and $N_R = 2$.

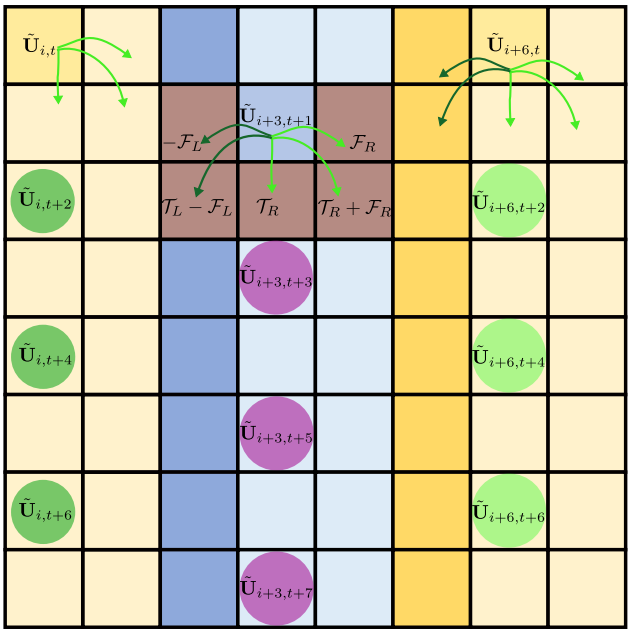


Fig. 5: Joint interpolation strategy: Each fed back subcarrier (i,t) uses neighboring future information to estimate precoder values for $(k,l) \in \texttt{cluster}(i,t)$. The brown shaded cells form $\texttt{cluster}(i+3, t+1)$.

SVD of a matrix is inherently not unique, and hence computation algorithms used to perform SVD of $\mathbf{H}_{i,t}$ to obtain $\mathbf{U}_{i,t}$ as explained in Section II may not provide a continuous sequence of $\mathbf{U}_{i,t}$ matrices. Hence, it is required to correct for abrupt changes while simulating the algorithm, and this was also performed in [3]. We have corrected for abrupt changes in singular vectors obtained via SVD by forcing the first row of the $\mathbf{U}_{i,t}$ (viz. the left matrix obtained from SVD of $\mathbf{H}_{i,t}$) to have positive real numbers, achieved via multiplying $\mathbf{U}_{i,t}$ by diagonal matrix corresponding to conjugate angle exponentials of the complex numbers in the first row of $\mathbf{U}_{i,t}$. This ensures a continuous sequence of $\mathbf{U}_{i,t}$ obtained via SVD of $\mathbf{H}_{i,t}$.

A. Proposed Interpolation on $St(N_T, N_R)$

We consider an OFDM system with $N = 1024$ subcarriers with ITU Vehicular-A channel model for these simulations. Keeping with the labeling in Fig. 3, we use $D_f = 33$, i.e. the receiver feeds back the precoder for 32 equally spaced feedback points at indices $33k, k = 0, 1, 2 \ldots 31$. The transmitter interpolates over 4×2 $\tilde{\mathbf{U}}_{i,t}$ matrices via (3) and over 4×4 $\tilde{\mathcal{U}}_{i,t}$ matrices via (2).

1) Codebook Generation: A 6 bit codebook for both 4×2 and 4×4 matrices is constructed using the Lloyd codebook algorithm [13]. This is done by generating 10,000 such $\mathbf{U}_{i,t}$'s and $\mathcal{U}_{i,t}$'s. We, therefore, use $32 \times 6 = 192$ bits per OFDM frame for precoder feedback for both the

cases.

2) Results: The results in Fig. 6 were obtained by averaging over 100 independent simulation runs. We observe approximately 3 dB gain at 10^{-3} BER with uncoded QPSK and MMSE equalization when using the Cayley Exp. map instead of the unitary geodesic scheme for interpolation. Note that, when interpolation is done over the ideal unquantized precoders (i.e. over $\mathbf{U}_{i,t}$, $\mathcal{U}_{i,t}$, instead of $\tilde{\mathbf{U}}_{i,t}$, $\tilde{\mathcal{U}}_{i,t}$), we do not observe performance benefits.

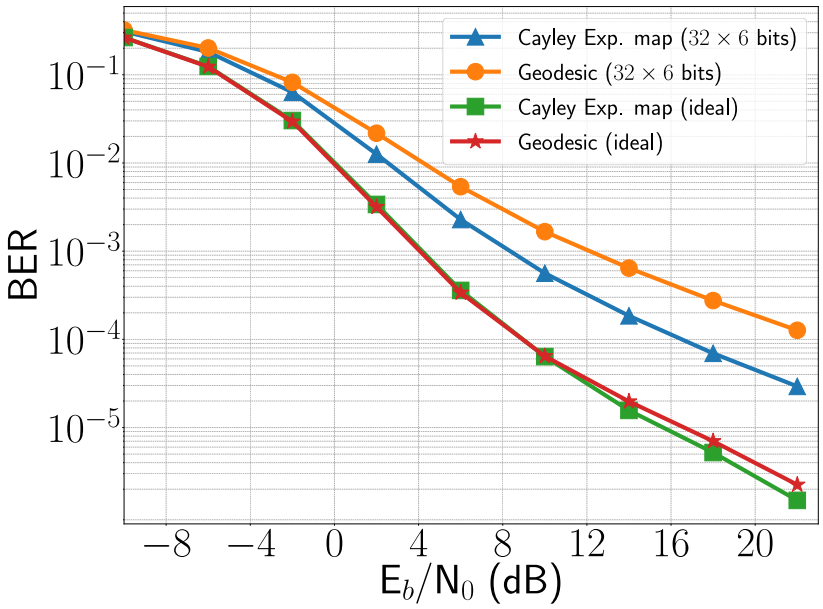


Fig. 6: BER observed in a 4×2 MIMO system when using Geodesic interpolation [2] and Cayley Exponential interpolation as discussed in Section III. The channel had the ITU Vehicular-A profile, and OFDM with 1024 subcarriers was used.

This ascertains the fact that reduction in BER is achieved since we do not quantize redundant information when interpolating directly on $\mathrm{St}(N_T, N_R)$. By dispensing with this redundant information, we can encode more information with lesser quantization error using the same number of bits.

When the singular values are quantized and fed back to enable waterfilling, we observe improvements in achievable sum-rate as well. A 2 bit codebook obtained via k-means clustering of about 1000 independent sample singular values was used to feed back quantized singular values for waterfilling. By allocating 2 bits at each fed back subcarrier for the singular values, we observe a 14% improvement in achievable rate at 0 dB SNR. We see in Fig. 7 that the 14% improvement is not an "average" improvement, and that the proposed scheme is better than geodesic unequivocally at each subcarrier index by approximately 14%.

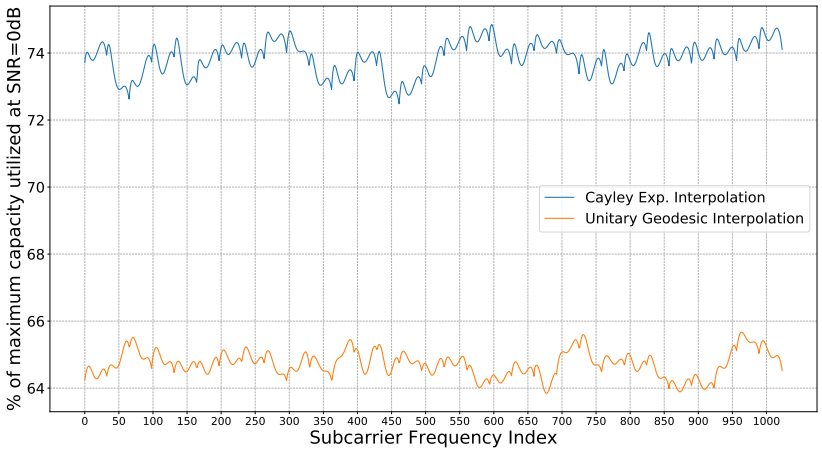


Fig. 7: The achievable rate for a 4×2 MIMO system when using Geodesic interpolation [2] and Cayley Exponential interpolation as discussed in Section III. Other assumptions are similar to those in Fig. 6.

B. Predictive Quantization and joint time-frequency based interpolation scheme

We have simulated our proposed scheme for both ITU Vehicular-A and Pedestrian-A channel models. For the Vehicular channel model, number of subcarriers N were chosen to be 1024 and for Pedestrian channel model, we chose N to be 64. We consider the hopping pattern of feedback indices (shown in Fig. 3), i.e. $\delta_f = 16$ (4), with $D_f = 33$ (9) when N is 1024 (64) respectively. To compare this with the time based predictive quantization and frequency interpolation scheme as in Fig. 2, we feed back indices of the form $33k$ $(9k)$ $(N = 1024$ $(64))$. For interpolation in the time based approach (Fig. 2), we utilize the Cayley Exp. map. The simulations for the algorithm were done for a normalized Doppler $f_d T_s = 3.5 \times 10^{-4}$, until otherwise mentioned.

1) Codebook generation: Since both the time based and hopping based predictive quantization methods predict based on past fed back information, we initialize the system with the 6 bit Lloyd codebook for $\mathrm{St}(4, 2)$ obtained in Section V-A1. Once sufficient past values of $\tilde{\mathbf{U}}_{i,t}$ have been quantized and fed back in advance, the algorithms can be used to start improving upon the quantization chordal error by exploiting the correlations available. An additional 2 bit k-means codebook was used to feed back quantized singular values for waterfilling.

To obtain the base codebook, $\mathcal{T}^B$ for $T_{\mathbf{P}_{k,l}}\mathrm{St}(N_T, N_R)$, we consider 10,000 independent channel evolutions. For each one, we use the evolution of the channel over 10 OFDM frames' duration to obtain a reliable prediction from both methods. We collect the vectors representing the skew Hermitian tangents, obtained by $\mathrm{lift}_{\mathbf{P}_{k,l}}(\mathbf{U}_{k,l})$, for vector quantization (recall that Cayley Exp.

map has skew Hermitian matrices as images (Section III), which form a vector space). This is done to get 10,000 independent predictions that give the same number of independent vectors in the collection. We then apply k-means algorithm to get a 6 bit base codebook, $\mathcal{T}^B$ for $T_{\mathbf{P}_{k,l}}\mathrm{St}(N_T, N_R)$, for both the hopping based and time based predictive quantization schemes. Hence we transmit $(32 \times 6 = 192)$ and $(8 \times 6 = 48)$ bits per frame when $N = 1024$ and 64 respectively for time based predictive quantization scheme. For hopping predictive quantization scheme, the feedback is $(31.5 \times 6 = 189)$ and $(7.5 \times 6 = 45)$ bits per frame when $N = 1024$ and 64, since the feedback alternates between total $32, 31$ and $8, 7$ subcarriers fed back, between two frames. These feedback bits indicate the optimum tangent which the transmitter selects to refine the estimated precoder predictions for appropriate fed back subcarriers.

2) Estimate chordal distance, $d[t]$ results: From Fig. 8 we observe that with time, $d[t]$ for the hopping predictive quantization scheme becomes much lower than that of the time based approach. We observed that after a certain time instant, the chordal distance on the Stiefel Manifold bumps up, which was observed even in [3]. Hence, we propose a scheme in which we reset the prediction algorithm and re-initialize with the 6 bit Lloyd codebook obtained for $\mathrm{St}(4, 2)$ when such a detrimental situation is observed. Note that, even for initialization, we communicate 192 (or 189) and 48 (or 45) bits per frame only, depending whether N is 1024 (or 64, respectively). This allows the algorithm to reset upon facing unpredictable heavy interference, since we do not assume a very heavy feedback requirement for initialization.

3) Results for hopping scenario: Figs. 9a, 9b, shows BER performance (uncoded QPSK, MMSE equalization), averaged over 10 independent channel evolutions, for Vehicular, $N = 1024$ and Pedestrian, $N = 64$ respectively. In each evolution, the algorithm was run till a sudden jump in quantization error was observed (as in [3]). Upon encountering the jump, we reset the algorithm using the independent 6 bit codebook.

For the Vehicular channel model, by performing predictive quantization on our hopping strategy, and doing joint time-frequency interpolation, we have matched the *ideal* 32 subcarrier feedback BER performance. The *ideal* 32 subcarrier feedback corresponds to the case in which the transmitter is assumed to know the ideal unquantized precoding matrices for 32 equally spaced subcarriers $(33k)$ in each OFDM frame and interpolates over frequency to find precoders at non fed back subcarriers, labeled "Cayley Exp. map (32, ideal)" in Fig. 9a. The 63 subcarrier ideal feedback BER curve, labeled "Cayley Exp. map (63, ideal)", serves as a lower bound on the

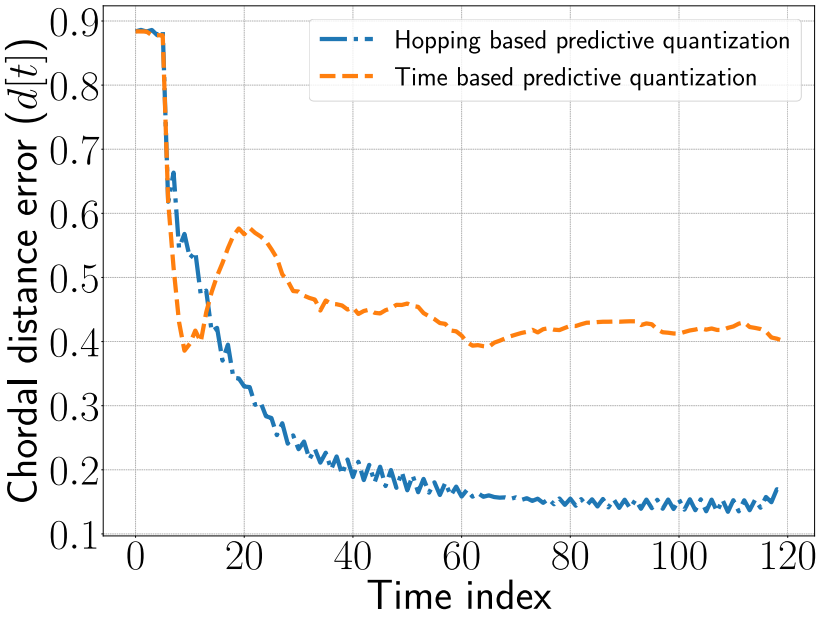


Fig. 8: Chordal distance error ($d[t]$) vs. time for one channel realization, Vehicular channel model, 1024 subcarriers. Here, the hopping based predictive strategy refers to the situation where the fed back points are alternated, as shown in Fig. 3 and interpolation performed via method discussed in Section IV-B. The time based predictive quantization strategy uses the temporal predictive quantization [3], with feedback points as shown in Fig. 2, and interpolation performed via Cayley exponential lifting map as discussed in Section III.

achievable BER performance for our hopping strategy. This scenario indicates that the transmitter has knowledge of ideal unquantized precoders at 32 equally spaced indices along with 31 middle indices of the regular intervals between them, in each OFDM frame, and performs interpolation as before. Hence, matching that curve would mean that we have exploited the temporal and frequency correlations perfectly, which expains why the BER of our scheme is lower bounded by this scenario. We also observe a ~2 dB gain at 10^{-4} BER, ~ 5 dB gain at 10^{-5} BER, when compared to the time based approach.

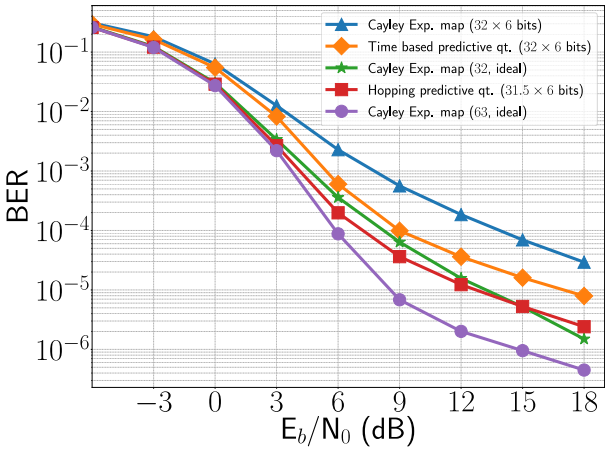
For the Pedestrian channel model, by performing predictive quantization via both the approaches, labeled "Time based predictive qt. (8×6 bits)", "Hopping predictive qt. (7.5×6 bits)" in Fig. 9b, we get huge improvements over the non-predictive independent quantization case, labeled "Cayley Exp. map (8×6 bits)". Observe that doubling the number of subcarriers fed back in the ideal cases does not offer much improvement, labeled "Cayley Exp. map (8/15 ideal)". This can be explained by the fact that the Pedestrian channel is strongly correlated along frequency, and thus doubling the number of subcarriers fed back does not add much to the information content. Therefore, the time domain information is more valuable in the Pedestrian channel

model, and this explains why, both the predictive quantization schemes are able to improve the BER performance significantly. Even though the time domain information is more valuable, our proposed scheme, which combines both time and frequency information, still improves upon the existing approach by around 4 dB at 10^{-4} BER. However, we are not able to match the ideal performance curves, more particularly at higher SNRs, as was the case in Vehicular channel models. This can be explained due to occurrence of the error floors observed in the quantized curves at about 10^{-6} BER.

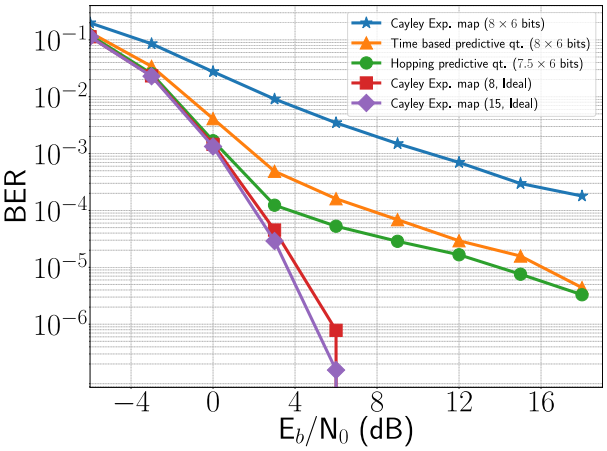
We obtain the achievable rates in Figs. 9c, 9d by averaging similarly, and the singular values are sent and interpolated as described in Section V-A. The results obtained are similar for both the channel models considered. When the channel varies faster, the time-based prediction tracks the channel more efficiently due to higher temporal information. For slower channels, utilizing the frequency correlation is more valuable, since the temporal variation is less significant. The crossover is at about $f_d T_s = 10^{-2}$ for both the channel models considered, which is generally a fast varying channel; we expect the hopping approach to be better in most realistic mobility scenarios.

VI. Conclusions

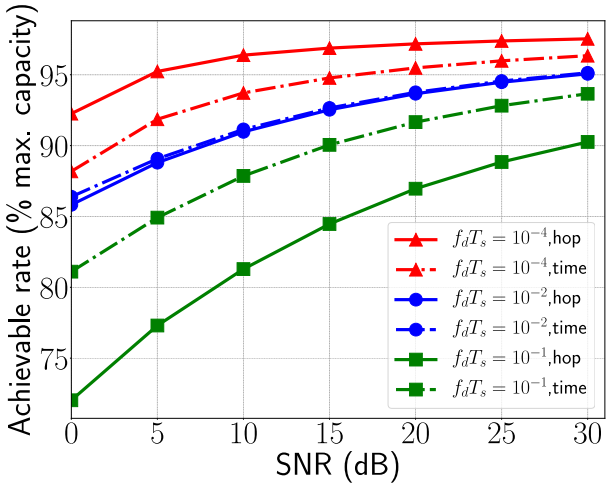
In this paper, we have proposed a new method for interpolating precoder matrices on the Stiefel manifold for wireless MIMO systems, and empirically showed that it offers a better performance than traditional geodesic interpolation. We then proposed a hopping strategy for the location of fed back subcarriers, coupled with a predictive quantization scheme and a joint time-frequency interpolation technique to utilize the time and frequency correlations optimally. The proposed predictive quantization scheme improves quantization performance in subsequent OFDM frames and brings the BER and achievable rates close to that obtained using theoretical unquantized precoders. When compared to past approaches that utilize the same amount of feedback, our approach allows the E_b/N_0 requirement for a the BER to be reduced by $\sim$5 dB. Simultaneously, the achievable rate is also significantly boosted due to the reduced errors in precoder quantization and interpolation. Future work would involve a study with various antenna configurations and generalizing the results to the multiuser-MIMO scenario.



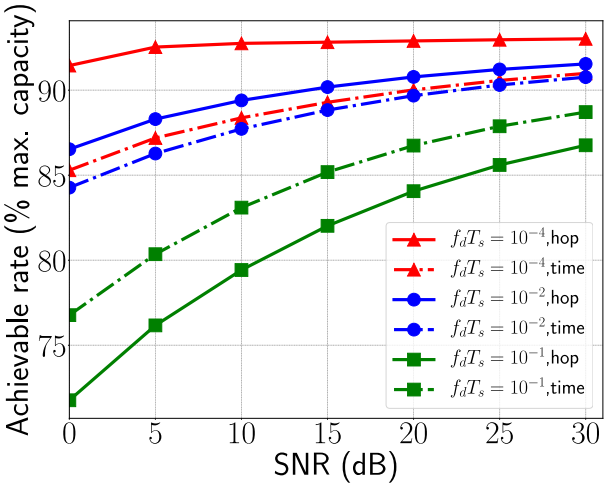
(a) BER results, Vehicular channel model



(b) BER results, Pedestrian channel model



(c) Achievable rate results, Vehicular channel model



(d) Achievable rate results, Pedestrian channel model

Fig. 9: Results for the proposed predictive quantization and joint time-frequency based interpolation scheme. Here, the hopping based predictive strategy refers to the situation where the fed back points are alternated, as shown in Fig. 3 and interpolation is performed using the method discussed in Section IV-B. The time based predictive quantization strategy uses the temporal predictive quantization [3], with feedback points as shown in Fig. 2, and interpolation performed via Cayley exponential lifting map as discussed in Section III.

REFERENCES

[1] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, 2008.

[2] N. Khaled, B. Mondal, R. W. Heath, G. Leus, and F. Petré, "Quantized multi-mode precoding for spatial multiplexing MIMO-OFDM systems," in *IEEE 62nd Vehicular Technology Conference*, vol. 2. IEEE, 2005, pp. 867–871.

[3] S. Schwarz and M. Rupp, "Predictive Quantization on Stiefel Manifold," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 234–238, Feb 2015.

[4] S. Schwarz, R. W. Heath, and M. Rupp, "Adaptive Quantization on a Grassmann-Manifold for Limited Feedback Beamforming Systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4450–4462, Sept 2013.

[5] K. Schober, R. Pitaval, and R. Wichman, "Improved User-Specific Channel Estimation Using Geodesical Interpolation at the Transmitter," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 165–168, April 2015.

[6] R. T. Krishnamachari and M. K. Varanasi, "On the geometry and quantization of manifolds of positive semi-definite matrices," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4587–4599, 2013.

[7] C.-B. Chae, D. Mazzarese, N. Jindal, and R. W. Heath, "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, 2008.

[8] Y. Chou and T. Sang, "Efficient interpolation of precoding matrices in MIMO-OFDM systems," in *Signal Processing Advances in Wireless Communications (SPAWC), 2010*, June 2010, pp. 1–4.

[9] D. Sacristán-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on geodesic curves," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3714–3727, 2010.

[10] T. Li, F. Li, and C. Li, "Manifold-based predictive precoding for the time-varying channel using differential geometry," *Wireless Networks*, vol. 22, no. 8, pp. 2773–2783, Nov 2016.

[11] Y. Li, S. Zhu, H. Tong, and M. Xu, "Enhanced limited rate implicit CSI feedback and its usage in covariance matrix based MU-MIMO," in *Wireless Communications and Networking Conference (WCNC)*. IEEE, 2013, pp. 3067–3071.

[12] Y. Jeon, H. Kim, Y. Cho, and G. Im, "Time-Domain Differential Feedback for Massive MISO-OFDM Systems in Correlated Channels," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 630–642, Feb 2016.

[13] R. Pitaval and O. Tirkkonen, "Joint Grassmann-Stiefel Quantization for MIMO Product Codebooks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 210–222, January 2014.

[14] R. Chakraborty and B. C. Vemuri, "Statistics on the (compact) Stiefel manifold: Theory and Applications," *CoRR*, vol. abs/1708.00045, 2017. [Online]. Available: http://arxiv.org/abs/1708.00045

[15] T. Inoue and R. W. Heath, "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 3076–3079.

[16] K. A. Krakowski, L. Machado, F. S. Leite, and J. Batista, "A modified Casteljau algorithm to solve interpolation problems on Stiefel manifolds," *Journal of Computational and Applied Mathematics*, vol. 311, pp. 84–99, 2017.

[17] S. Fiori, T. Kaneko, and T. Tanaka, "Mixed maps for learning a Kolmogoroff-Nagumo-type average element on the compact Stiefel manifold," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2014, pp. 4518–4522.

[18] V. Saxena, https://github.com/vidits-kth/py-itpp.

VII. Appendix

A. The Cayley Exponential Lifting and Retraction Pairs

Given $\mathbf{X}, \mathbf{Y} \in St(N_t, N_r)$, [14] defines the lifting map $\mathrm{Exp}_{\mathbf{X}}^{-1}(\mathbf{Y}) : St(N_t, N_r) \to T_{\mathbf{X}} St(N_t, N_r)$ by

$$\mathrm{Exp}_{\mathbf{X}}^{-1}(\mathbf{Y}) = \begin{bmatrix} \mathbf{C} & -\mathbf{B}^H \\ \mathbf{B} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{C} = 2(\mathbf{X}_u^H + \mathbf{Y}_u^H)^{-1}\mathtt{sk}(\mathbf{Y}_u^H\mathbf{X}_u + \mathbf{X}_l^H\mathbf{Y}_l)(\mathbf{X}_u + \mathbf{Y}_u)^{-1}$ is a $N_r \times N_r$ skew hermitian matrix and $\mathbf{B} = (\mathbf{Y}_l - \mathbf{X}_l)(\mathbf{X}_u + \mathbf{Y}_u)^{-1}$ is a $(N_t - N_r) \times N_r$ matrix where, $\mathbf{X} = [\mathbf{X}_u, \mathbf{X}_l]^H$, $\mathbf{Y} = [\mathbf{Y}_u, \mathbf{Y}_l]^H$, with $\mathbf{X}_u, \mathbf{Y}_u \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{X}_l, \mathbf{Y}_l \in \mathbb{C}^{N_r \times (N_t - N_r)}$, provided that $\mathbf{X}_u + \mathbf{Y}_u$ is nonsingular, $\mathtt{sk}(\mathbf{M}) = \frac{1}{2}(\mathbf{M}^H - \mathbf{M})$. Note that $\mathrm{Exp}_{\mathbf{X}}^{-1}(\mathbf{Y})$ maps matrices residing in the Stiefel manifold to a $N_t \times N_t$ skew Hermitian matrices.

The corresponding retraction map $\mathrm{Exp}_{\mathbf{X}}(\mathbf{T}) : T_{\mathbf{X}}St(N_t, N_r) \to St(N_t, N_r)$, with $\mathbf{X} \in St(N_t, N_r)$, $\mathbf{T} \in T_X St(N_t, N_r)$ is defined as $\mathrm{Exp}_{\mathbf{X}}(\mathbf{T}) = \mathtt{Cay}(\mathbf{T})\mathbf{X}$, where $\mathtt{Cay}(\mathbf{T}) = (\mathbf{I}_{N_t} + \mathbf{T})(\mathbf{I}_{N_t} - \mathbf{T})^{-1}$ is the Cayley conformal map. Hence the unique geodesic from $\mathbf{X} \in St(N_t, N_r)$ to $\mathbf{Y} \in St(N_t, N_r)$, denoted by $\Gamma_{\mathbf{X}}^{\mathbf{Y}}(t) = \mathrm{Exp}_{\mathbf{X}}(t\mathrm{Exp}_{\mathbf{X}}^{-1}(\mathbf{Y}))$.

Let $\tilde{\mathbf{U}}_{i_1,t}, \tilde{\mathbf{U}}_{i_2,t} \in St(N_t, N_r)$ be the quantized precoding matrices corresponding to the i_1, i_2 subcarriers at the $t-$th time index. Then, the interpolated $\tilde{U}_{i,t}$ with $i_1 < i < i_2$ is given by,

$$\tilde{\mathbf{U}}_{i,t} = \Gamma_{\tilde{\mathbf{U}}_{i_1,t}}^{\tilde{\mathbf{U}}_{i_2,t}}\left(\frac{i}{i_2 - i_1}\right) = \mathrm{Exp}_{\tilde{\mathbf{U}}_{i_1,t}}\left(\frac{i}{i_2 - i_1}\mathrm{Exp}_{\tilde{\mathbf{U}}_{i_1,t}}^{-1}(\tilde{\mathbf{U}}_{i_2,t})\right)$$

B. Detailed steps to derive $\Delta_{i,t}$

Recall that, $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$ are matrices such that $\mathbf{T}_{i,t}^{j,s} \approx \mathcal{F}_{i,t}\Delta f + \mathcal{T}_{i,t}\Delta t$, where $\Delta f, \Delta t \in \mathbb{R}$ are small steps in the time and frequency axes respectively. We have $\Delta f = j - i$ and $\Delta t = s - t$, i.e. the signed frequency/time separation between boxes (j, t) and (i, t). With the neigbour precoders for past p samples ($\mathtt{nbrs}_{i,t}(p)$) defined in (4), we get the following optimization framework to estimate $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$,

$$\mathcal{F}_{i,t}, \mathcal{T}_{i,t} \leftarrow \mathrm{argmin} \sum_{(j,s)\in\mathtt{nbrs}_{i,t}(p)} ||(\mathcal{F}_{i,t}(j - i) + \mathcal{T}_{i,t}(s - t) - \mathbf{T}_{i,t}^{j,s})||_F^2 \tag{9}$$

Now, define objective function $O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t}) = \sum_{(j,s)\in\mathtt{nbrs}_{i,t}(p)} ||(\mathcal{F}_{i,t}(j - i) + \mathcal{T}_{i,t}(s - t) - \mathbf{T}_{i,t}^{j,s})||_F^2$. We use the identities $\frac{\partial}{\partial X}(\mathrm{Tr}(\mathbf{AXB})) = \mathbf{A}^H\mathbf{B}^H$ and $\frac{\partial}{\partial \mathbf{X}}\left(\mathrm{Tr}(\mathbf{AX}^H\mathbf{B})\right) = \mathbf{BA}$, where $\mathbf{A}, \mathbf{X}, \mathbf{B}$ are matrices such that the products $\mathbf{AXB}$ and $\mathbf{AX}^H\mathbf{B}$ exist. Also, recall $|\mathbf{M}|_F = \mathrm{Tr}(\mathbf{MM}^H)$ for any matrix $\mathbf{M}$. For the sake of brevity, we write $\sum_{(j,s)\in\mathtt{nbrs}_{i,t}(p)}$ as $\sum_{(j,s)}$ in the subsequent discussion. Taking the partial derivative of $O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t})$ with respect to $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$ we get the following equations,

$$\frac{\partial}{\partial \mathcal{F}_{i,t}} O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t}) = \sum_{(j,s)}\left((j - i)^2\mathcal{F}_{i,t} + (j - i)(s - t)\mathcal{T}_{i,t} - (j - i)\mathbf{T}_{i,t}^{j,s}\right)$$

$$\frac{\partial}{\partial \mathcal{T}_{i,t}} O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t}) = \sum_{(j,s)} \left((j-i)(s-t)\mathcal{F}_{i,t} + (s-t)^2 \mathcal{T}_{i,t} - (s-t)\mathbf{T}_{i,t}^{j,s} \right)$$

Setting $\frac{\partial}{\partial \mathcal{T}_{i,t}} O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t})$, $\frac{\partial}{\partial \mathcal{F}_{i,t}} O(\mathcal{T}_{i,t}, \mathcal{F}_{i,t})$ to null matrix gives us the following linear equation (Observe that $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$ will come out of the summation since they do not depend on j, s):

$$\begin{bmatrix} \mathcal{F}_{i,t} \\ \mathcal{T}_{i,t} \end{bmatrix} = \mathbf{\Delta}_{i,t}^{-1} \begin{bmatrix} \Sigma_{j,s}(j-i)\mathbf{T}_{i,t}^{j,s} \\ \Sigma_{j,s}(s-t)\mathbf{T}_{i,t}^{j,s} \end{bmatrix}$$

$$\mathbf{\Delta}_{i,t} = \begin{bmatrix} \Sigma_{j,s}(j-i)^2 & \Sigma_{j,s}(j-i)(s-t) \\ \Sigma_{j,s}(j-i)(s-t) & \Sigma_{j,s}(s-t)^2 \end{bmatrix}$$

This completes the detailed proof of $\Delta_{i,t}$