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

Abstract

Precoding transmissions in wireless MIMO systems is essential to enable optimal utilization of the spatial degrees of freedom. However, communicating the precoding matrices 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 to reconstruct the precoders over a wide band, although these techniques have not been used jointly. We propose both a predictive quantization as well as a joint time-frequency interpolation strategy for precoding matrices over the Stiefel manifold. The key insight that we use is that local tangent spaces in the manifold 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, the receiver needs to feedback channel state information (CSI) to the transmitter, which then uses it for precoding (matrix transformation of the input signal) for efficient allocation of power

across transmit antennas by waterfilling, or for efficient signaling to increase rate while reducing bit-error-rate (BER) [1].

Modern wireless OFDM systems support thousands of subcarriers, thus the CSI requirement is rather large. However, only a limited number of bits are reserved for CSI transmission. This necessitates exploiting the underlying manifold structure of the channel/precoding matrices matrix for quantization and interpolation [2] across frequency. Utilizing the frequency/ temporal correlation independently, to improve channel estimates has been considered in [3–8]. The manifold structure permits reconstruction and prediction of precoders at unknown locations.

Typical approaches for quantization over manifolds involves unitary and Grassmannian manifolds, where the quantization is implemented over the full $N_T \times N_T$ space, while the effective dimensions are only $N_T \times N_R$, where $N_T > N_R$. Stiefel manifolds can be used to work with the effective dimensions ($N_T \times N_R$), thereby reducing the CSI overhead, however, computing geodesics (shortest paths) over the Stiefel manifolds is challenging. In recent work, a Lloyd type codebook scheme has been suggested for direct quantization of matrices residing in Stiefel manifold [9] and it has been shown that closed form unique quasi-geodesics can be computed over Stiefel manifold [10]. We exploit both of these to reduce the CSI overhead requirement by quantizing and interpolating in the effective $N_T \times N_R$ space.

In addition to working over the reduced space using the Stiefel manifold quasi-geodesics, we also focus on jointly exploiting the temporal and 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 correlation [11] or in frequency domain [12] 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 [4, 11].

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, we are able to reduce the E_b/N_0 requirement by around $\sim$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 the paper.

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 [11], 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 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 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). We get N orthogonal flat fading channels via the OFDM modulation scheme, which allows for detection of data on each subcarrier independently. From a MMSE decoder and the received $\mathbf{y}_{i,t}$, one obtains an estimate of $\mathbf{x}_{i,t}$, denoted by $\hat{\mathbf{x}}_{i,t}$. This scenario is visually illustrated in Fig. 1. We assume that the channel is estimated ideally at the receiver.

If the accurate channel matrices $\mathbf{H}_{i,t}^H$ are available at the transmitter, the optimal policy is to use the right singular vectors of $\mathbf{H}_{i,t}^H$ as the precoding matrices $\tilde{\mathbf{U}}_{i,t}$ to maximise achievable rate. In particular, if $\text{SVD}(\mathbf{H}_{i,t}) = \mathbf{U}_{i,t}, \Sigma_{i,t}, \mathbf{V}_{i,t}$, then $\tilde{\mathbf{U}}_{i,t} = \mathbf{U}_{i,t}$. Notice that matrices $\mathbf{U}_{i,t}$ reside on the Stiefel manifold $\text{St}(N_T, N_R)$, since the columns of $\mathbf{U}_{i,t}$ form a set of N_R orthogonal vectors in

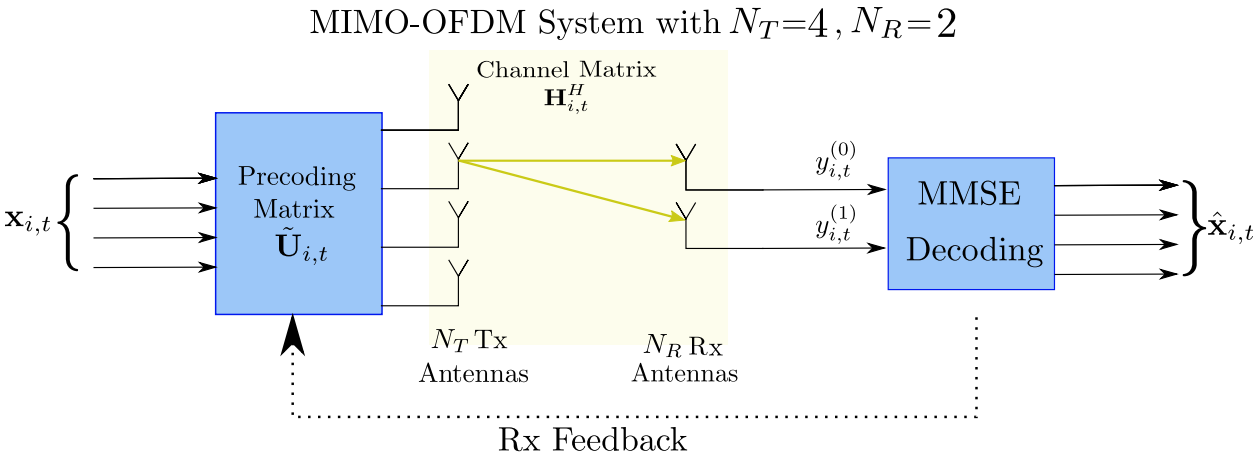


Fig. 1: Sample point-to-point MIMO-OFDM setting that we consider, with unequal number of transmit and receive antennas. The system shown is a 4×2 system, where the precoding matrix $\tilde{\mathbf{U}}_{i,t}$ is a 4×2 with $\tilde{\mathbf{U}}_{i,t}^H \tilde{\mathbf{U}}_{i,t} = \mathbf{I}_2$. Channel estimates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

N_T dimensions. On the other hand, if the transmitter allocates equal power to each antenna, the optimal policy is to use the subspace information conveyed by the right singular vectors, viz. $\tilde{\mathbf{U}}_{i,t} = \mathbf{U}_{i,t}\check{\mathbf{U}}$, where $\check{\mathbf{U}}$ is a $N_R \times N_R$ unitary matrix [13]. Since only the subspace information is required, the optimal precoders for this setting reside on the Grassmannian manifold. In this work, we consider that transmitter has flexibility in allocating power and hence the optimal precoders reside on the Stiefel manifold. Given practical limitations, only limited feedback is available from the receiver, and the objective is to find 'reasonable' estimate of $\mathbf{U}_{i,t}$, given by $\tilde{\mathbf{U}}_{i,t}$ using the available feedback bits.

Before proceeding, we discuss the points concerned with the geometry of the Stiefel manifold. The (compact) Stiefel manifold, $\mathrm{St}(m, n)$ represents set of all ordered m orthogonal frames in the vector space $\mathbb{C}^n$ [11]:

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

Clearly, this does not form a vector space, because of which linear algorithms to interpolate and quantize are not effective. 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 for vector algebraic operations. The vector space property of the local tangent space allows for prediction and interpolation algorithms for SVD precoders in the MIMO-OFDM setting. We consider the chordal distance metric on the

Stiefel manifold that is 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 $\mathbf{a}_j, \mathbf{b}_j$ is the j-th column 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 $\mathbf{b}_j$ [5]. Thus, the objective is to obtain an estimate $\tilde{\mathbf{U}}_{i,t}$ of $\mathbf{U}_{i,t}$ that minimizes the chordal distance between $\tilde{\mathbf{U}}_{i,t}$ and $\mathbf{U}_{i,t}$ on the Stiefel manifold.

Frequency and time dependent channel matrices $\mathbf{H}_{i,t}^H$ (and consequently $\mathbf{U}_{i,t}$), pose an additional challenge with limited feedback, since feeding back information about $\mathbf{U}_{i,t}$ for each i and t is prohibitive, thereby making exploiting the frequency and temporal correlation structure imperative. Feeding back information about $\mathbf{U}_{i,t}$ at regular frequency and time intervals and employing a naïve linear interpolation does not work, since the precoding matrices do not form a vector space. Several approaches in literature exploit the underlying manifold structure but have largely exploited frequency or temporal correlation in isolation for interpolation and prediction [4, 6, 8, 11, 14, 15]. For the case where the precoding matrix is not square, temporal correlation has been used in [11] and an interpolation algorithm has been suggested in [12]. A non-manifold based approach to simultaneously capture frequency and temporal correlation is presented in [16].

In this paper, we present a manifold based approach to estimate the precoding matrices under limited feedback that uses both the frequency and temporal correlations. Our approach exploits the differential geometry of the Stiefel manifold. Throughout the paper, we assume that perfect zero-delay limited feedback or channel state information (CSI) is available at the transmitter.

We represent the considered limited feedback scenario by a time-frequency bin matrix (Fig. 2), where information about time-frequency bins is sparsely fed back by the receiver. The transmitter has to estimate the matrices at each non fed-back point by utilizing the temporal and frequency correlations. The metric we choose to compare two estimates of the time-frequency bin matrix 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 as in (1). Observe that $d[t_1]$ at a particular time t_1 represents the average chordal distance between the estimates and actual values of row t_1 in the frequency-time bin matrix. Using the prior work on predictive quantization over $\mathrm{St}(N_T, N_R)$ [11] along with the modifications suggested in Section III, one can fill the time-frequency bin matrix using the strategy shown in Fig. 2 by sending feedback on regularly spaced subcarrier indices. This strategy captures frequency and time correlations independently. A lower $d[t]$ may potentially

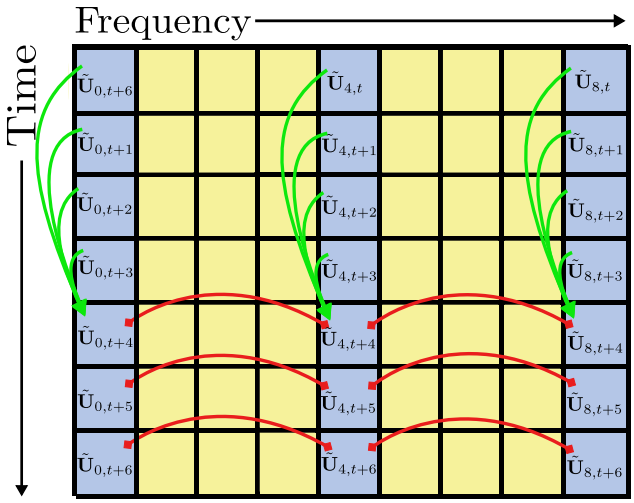


Fig. 2: Exploiting regularly spaced limited feedback scheme for $N = 9$, with (yellow) blue cells indicating the (non) fed back indices. Frequency correlations are captured via the red interpolating curves, and the temporal correlations are used to improve the quantization over time (shown implicitly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

be obtained over time by capturing the time and frequency correlations jointly. This is outlined in Section IV, and can provide higher rates and lower BERs.

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

In this section, we describe the interpolation scheme that allows interpolation along both the time and frequency (independently) domain over the Stiefel manifold. Doing this allows us to populate the time- frequency bins matrix with the strategy depicted in Fig. 2 and hence generalizes predictive quantization strategy in [11] for the MIMO-OFDM case.

Taking the full SVD of MIMO channel matrix $\mathbf{H}_{i,t}$ 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. Here, the optimal precoder consists of the first N_R columns of $\mathcal{U}_{i,t}$ [2]. $\tilde{\mathcal{U}}_{i,t}$ is the quantized value of $\mathcal{U}_{i,t}$. The usual approach using geodesic interpolation 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 via geodesic scheme [2] to obtain the $\tilde{\mathcal{U}}_{i,t}$ between i_1 and i_2 via

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

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 ones, which were included in the setting to form the unitary matrix and enable interpolation with geodesics, and are actually redundant.

We now propose a method to directly interpolate over the Stiefel manifold, thereby obviating the need to quantize the redundant information. A natural approach would be to use geodesics, but true geodesic curves with closed form solutions in the Stiefel manifold are not known [12, 17]. Recently, [10] has proposed a unique quasi geodesic curve using the Cayley type lifting-retraction maps that enable direct interpolation on $\mathrm{St}(N_T, N_R)$. Interpolation directly on $\mathrm{St}(N_T, N_R)$ has been discussed only in [12], albeit for unquantized precoders.

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 to obtain $\tilde{\mathbf{U}}_{i,t}$, similar to the geodesic interpolation in Equation 2 using the Cayley exponential lifting and retraction maps at $\tilde{\mathbf{U}}_{i_1,t}$, denoted as $\mathrm{Exp}^{-1}_{\tilde{\mathbf{U}}_{i_1,t}}(\cdot)$ and $\mathrm{Exp}_{\tilde{\mathbf{U}}_{i_1,t}}(\cdot)$ respectively, consistent with the notation in [10]. The transmitter interpolates via the below equation to obtain the $\tilde{\mathbf{U}}_{i,t}$ between i_1 and i_2

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

One thing to note is that the $\mathrm{logm}(\cdot)$ maps $N_T \times N_T$ unitary matrices to $N_T \times N_T$ skew-Hermitian matrices. Similarly, the $\mathrm{Exp}^{-1}(\cdot)$ in Equation 3 maps $N_T \times N_R$ matrices on $\mathrm{St}(N_T, N_R)$ 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 VII-A.

By using the approach described above the receiver needs to quantize and feed back only $N_T \times N_R$ matrices $\mathbf{U}_{i,t}$, with $\mathbf{U}^H_{i,t}\mathbf{U}_{i,t} = \mathbf{I}_{N_R}$ instead of $N_T \times N_T$ unitary matrices $\mathcal{U}_{i,t}$, thereby obviating the need to quantize redundant information. The results, codebook generation and other nuances for the simulations for these approaches are described in Section V.

IV. Predictive Quantization and Joint Time-Frequency Interpolation

In this section, we describe efficient methods for obtaining and quantizing precoding matrices at the receiver.

A. Hopping and Predictive Quantization scheme

The approach for interpolation and prediction shown in Figure 2 is limited, in that the channel is fed back only at a fixed set of subcarriers, and the remaining subcarriers' precoders are

interpolated. This method has two disadvantages. One is that the precoders for those subcarriers that lie between the subcarriers whose precoders have been fed back face larger errors, since they are equally far away from the subcarriers where feedback is available. The other is that, while predictively quantizing solely over the time-domain, we ignore the useful past information present in the nearby subcarriers present due to frequency correlation. We therefore suggest a "hopping" scheme, in which the fed back frequency bins are alternated in time. 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 $N-1$, where N is the total number of subcarriers. Therefore, 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.

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). Consider the hopping scheme described in Fig. 3. The yellow boxes 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 indicate the time-frequency bins where the precoding matrices that were not fed back.

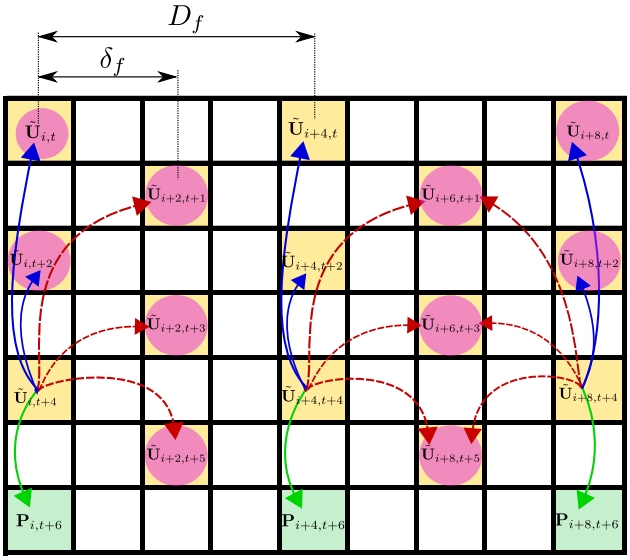


Fig. 3: Hopping strategy for $N = 9, D_f = 4, \delta_f = 2$. The red, blue and green curves represent the tangents on $\mathrm{St}(N_T, N_R)$, illustrated in Fig. 4.

For the t-th OFDM frame, the transmitter uses the past fed back information to obtain a prediction for those boxes where the receiver would provide feedback. This situation is depicted in Fig. 3 for the $(t+6)$-th OFDM frame. The feedback from the receiver indicates the direction in the local tangent space which takes the predicted value closest in terms of $d[t]$ to the actual precoders. For prediction, 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. The lifting map to obtain the tangent space and the retraction map to return to the manifold include the maps described in Section III, though several such maps exist. 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 $\Delta f, \Delta t \in \mathbb{R}$ are small steps in the time and frequency axes 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,

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

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 proceed ahead by taking partial derivatives of the objective function in (5) with respect to $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ and setting them to 0 (For a detailed proof of the below equations, refer Appendix Section VII-B). We 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, to predict the precoding matrix at box (k,l): When the temporal correlation of the channel is higher than the frequency correlation, $(i,t) = (k, l-2)$, 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 predictive quantization algorithm for box (k,l). Note that these operations can be performed at the transmitter independently without any additional feedback.

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 $\tilde{\mathbf{U}}_{k,l}$. The approach to quantize the tangent space local to the prediction has been discussed in [8, 11]. We adopt the strategy in [11], which controls the spread of codewords by having two codebooks T_p^C, T_m^C of different spreads, but same $2^{\mathcal{B}-1}$ base vectors $\in T_{\mathbf{P}_{k,l}}\text{St}(N_T, N_R)$, with T_p^C covering more volume than T_m^C. The two codebooks are concatenated to form a $2^{\mathcal{B}}$ length codebook T^C, and the receiver feeds back the optimal index using $\mathcal{B}$ bits. The receiver finds the optimal index c_n, in the concatenated codebook 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.

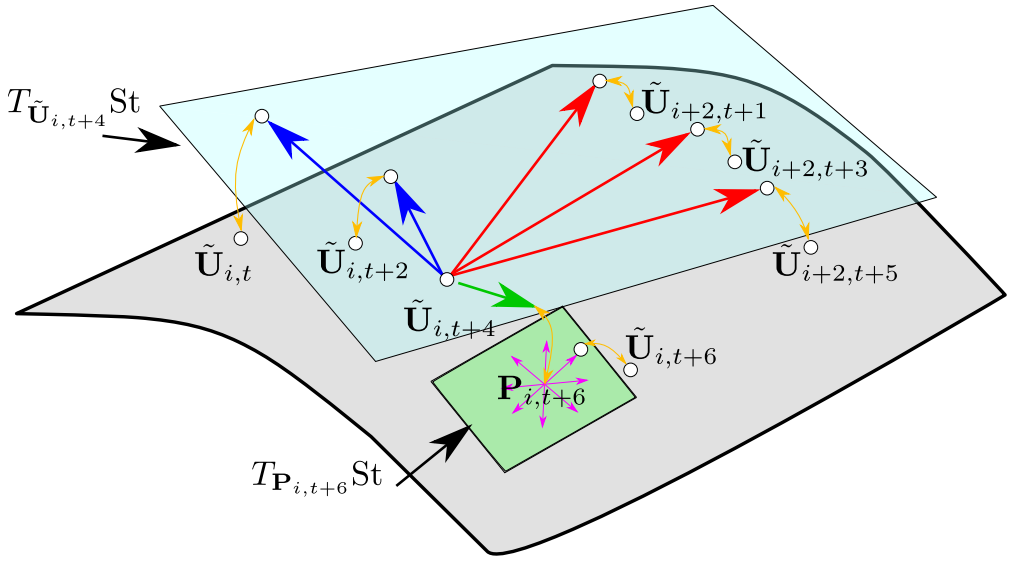


Fig. 4: Predictive Quantization algorithm to obtain $\tilde{\mathbf{U}}_{i,t+6}$, with closest previous value being $\tilde{\mathbf{U}}_{i,t+4}$. The yellow curves represent the lifting/retraction operations. $\mathbf{P}_{i,t+6}$ is the predicted value of $\tilde{\mathbf{U}}_{i,t+6}$. The blue and green 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 $\tilde{\mathbf{U}}_{i,t+6}$.

using (6). The transmitter uses (7) to calculate $\tilde{\mathbf{U}}_{k,l}$,

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

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

The base codebook consists of $2^{\mathcal{B}-1}$ matrices belonging to the local tangent space at $\mathbf{P}_{k,l}$. The two codebooks, $T_p^{\mathcal{C}}, T_m^{\mathcal{C}}$, 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 $T_p^{\mathcal{C}}$ or $T_m^{\mathcal{C}}$, the scale parameter $s[k]$, which in turn controls values of s_p, s_m is updated in the following manner,

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

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

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 codebook generation scheme we use, however, differs slightly from the one presented in [11]. As discussed in Section III, the Cayley type lifting map

has images as $N_T \times N_T$ skew Hermitian matrices, which allows for vector quantization of the tangent space, since a skew Hermitian matrix can be easily converted to a vector how?

. Vector quantization for the base codebook allows for isotropic codewords for the tangent space local to $\mathbf{P}_{k,l}$. [11] proposed a randomly initialized base codebook which does not guarantee isotropic codewords. 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 [18] and also used in [11] 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 [18], 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 box (k,l). This is also done by estimating the $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ at the closest fed back subcarrier (i,t) to subcarrier (k,l) by least-squares optimization. To interpolate we use future nearby feed back points, as shown in Fig. 5, since past fed back points' 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 $\mathcal{F}^L_{i,t}$ from $\texttt{l\_nbrs}_{i,t}(p)$, and right frequencies $\mathcal{F}^R_{i,t}$ from $\texttt{r\_nbrs}_{i,t}(p)$, since directly obtaining $\mathcal{F}_{i,t}, \mathcal{T}_{i,t}$ makes $\mathbf{\Delta}_{i,t}$ singular. Hence, we get the

following objective functions,

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

The lifting-retraction maps used here for interpolation are the Cayley exponential type lifting maps discussed in Section III. The interpolated estimates for non-fed back boxes (k, l) are hence given by,

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

The fed back boxes (i, t) obtained by predictive quantization act as centers to obtain the interpolated estimate of boxes $(k, l) \in \texttt{cluster}(i, t)$, which is defined as,

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

For optimal power allocation via waterfilling, the singular values should also be fed back. 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 py-itpp. Channels were generated using Jake's model with both ITU Vehicular-A and ITU Pedestrian-A profiles mentioned appropriately. The simulations have been performed for $N_T = 4$ and $N_R = 2$. 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

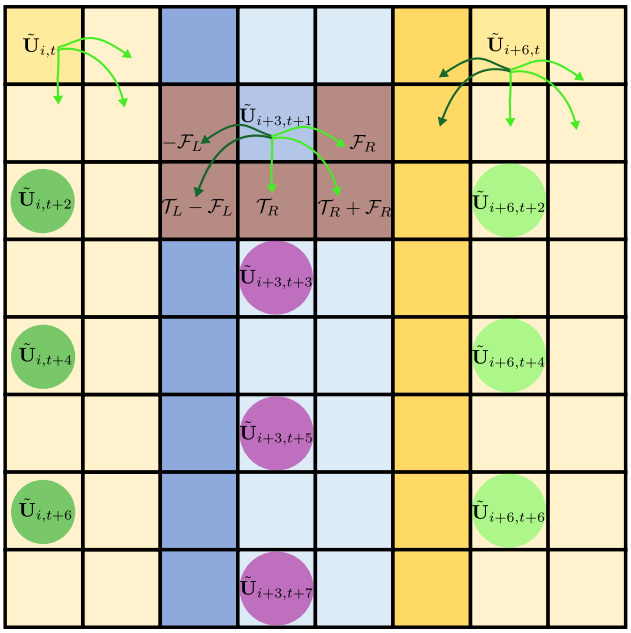


Fig. 5: Joint interpolation strategy: Each fed back subcarrier (i,t) uses neighboring future information to estimate precoder values for `cluster`(i,t). The brown shaded cells form `cluster`$(i+3, t+1)$.

$\mathbf{U}_{i,t}$ matrices. Hence it is required to correct for abrupt changes while simulating, which was also performed in [11]. 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, by 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 [9]. 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 simulations 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.

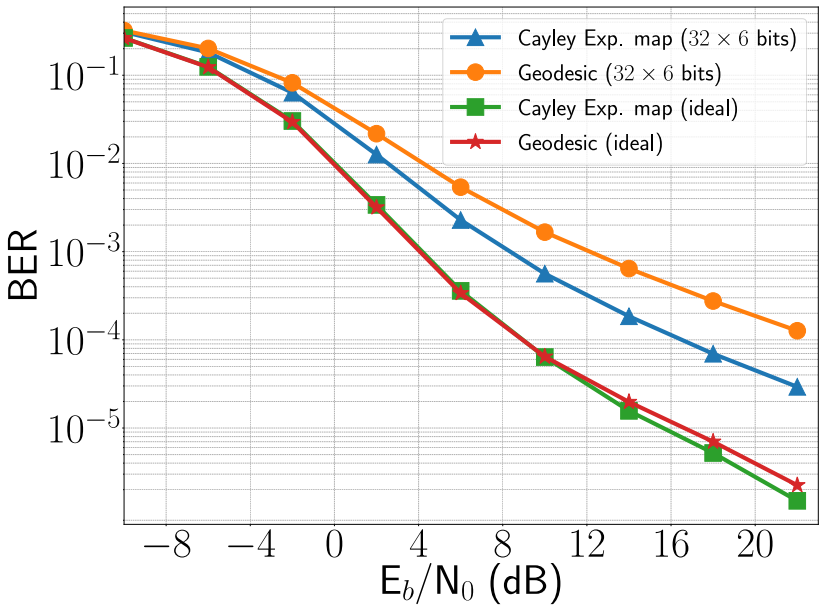


Fig. 6: BER observed in a 4×2 MIMO system when using Geodesic interpolation [2] and Cayley Exponential interpolation as discussed in Section III. The channel had the ITU Vehicular-A profile, and OFDM with 1024 subcarriers was used.

This ascertains the fact that benefits in BER arises since we do not encode 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. By allocating 2 bits in each fed back subcarrier for singular values, we observe a 14% improvement in achievable rate at 0 dB SNR. Observe from Fig. 7 that the 14% improvement is not an averaged improvement, but the proposed scheme is better than geodesic unequivocally at each subcarrier index by approximately 14%.

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 $D_f = 32$ (8) when N is 1024 (64) respectively.

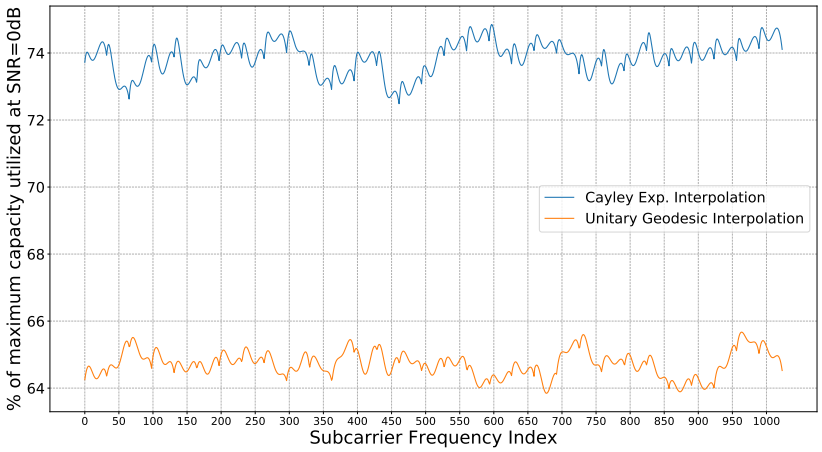


Fig. 7: The achievable rate for a 4×2 MIMO system when using Geodesic interpolation [2] and Cayley Exponential interpolation as discussed in Section III. Other assumptions are similar to those in Fig. 6.

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 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 quantisation (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 for $T_{\mathbf{P}_{k,l}}\mathrm{St}(N_T, N_R)$, for both the hopping based and time based schemes. Hence we transmit (32×6 = 192 and (8×6 = 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 $32, 31$ and $8, 7$ between two frames. These feedback bits indicate the optimum tangent, the transmitter can select to refine the estimated precoder predictions for appropriate fed back subcarriers.

2) Estimate chordal distance, $d[t]$ results: From Fig. 8 we observe that, over time, $d[t]$ for the hopping predictive quantization scheme is 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 [11]. 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).

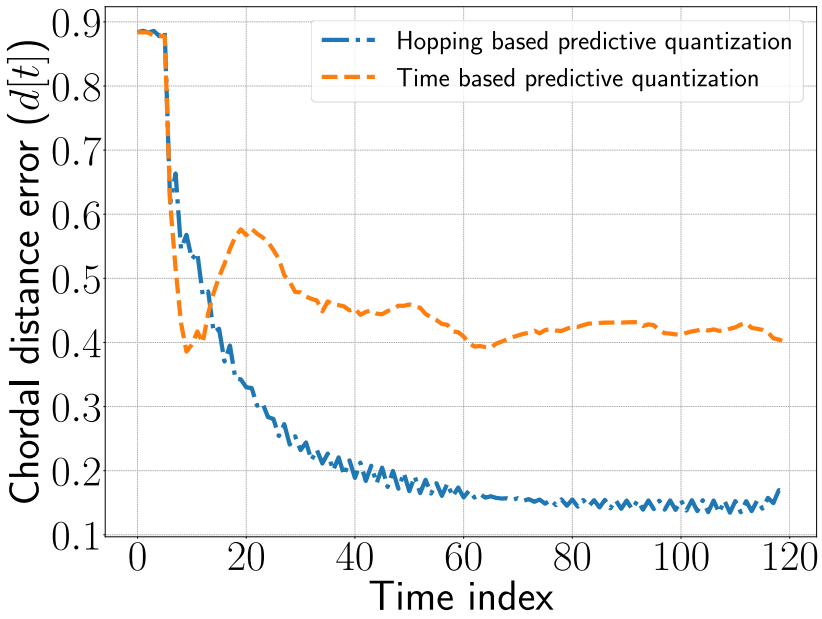


Fig. 8: Chordal distance error ($d[t]$) vs. time for one channel realization, Vehicular channel model, 1024 subcarriers. Here, the hopping based predictive strategy refers to the situation where the fed back points are alternated, as shown in Fig. 3 and interpolation performed via method discussed in Section IV-B. The time based predictive quantization strategy uses the temporal predictive quantization [11], with feedback points as shown in Fig. 2, and interpolation performed via Cayley exponential lifting map as discussed in Section III.

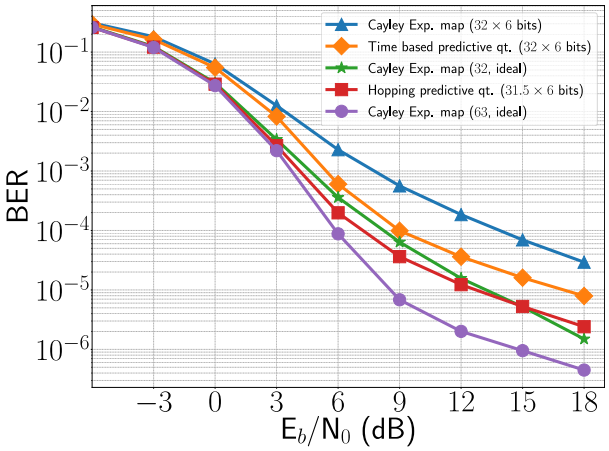
3) Results for hopping scenario: Figs. 9a, 9b, shows BER performance (QPSK, MMSE equalization), averaged over 10 independent channel evolutions, for Vehicular, $N = 1024$ and Pedestrian, $N = 64$ respectively. In each evolution, the algorithm was run till a sudden jump in quantization error was observed (as in [11]). Upon encountering the jump, we reset the algorithm

using the independent 6 bit codebook.

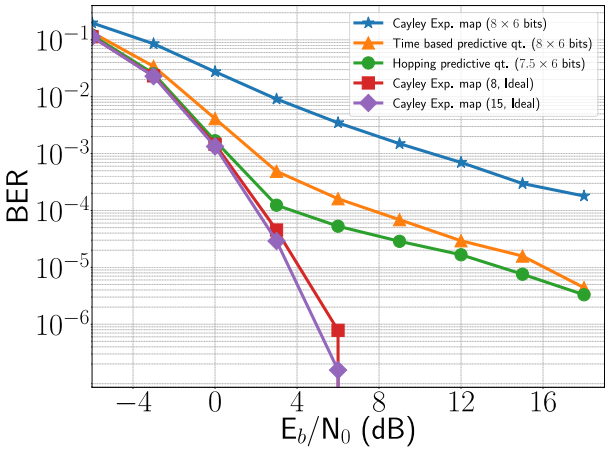
For the Vehicular channel model, by performing predictive quantization on our hopping strategy, and doing joint time-frequency interpolation, we have matched the *ideal* 32 subcarrier feedback BER performance. The *ideal* 32 subcarrier feedback corresponds to the case in which the transmitter is assumed 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 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. 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.

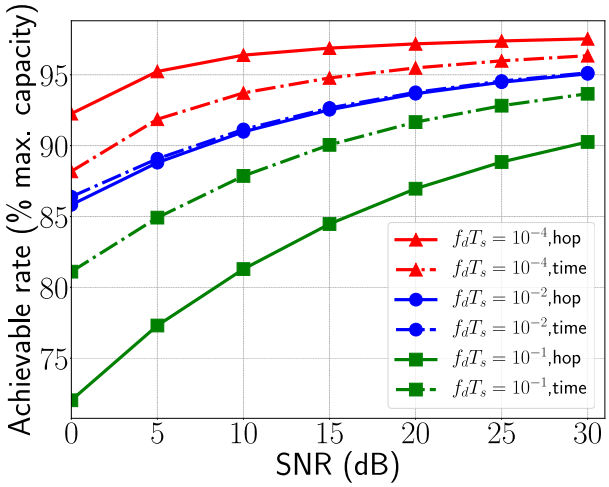
We obtain the achievable rates in Figs. 9c, 9d by averaging similarly, and the singular values are sent and interpolated as described in Section V-A. The results obtained are similar for both



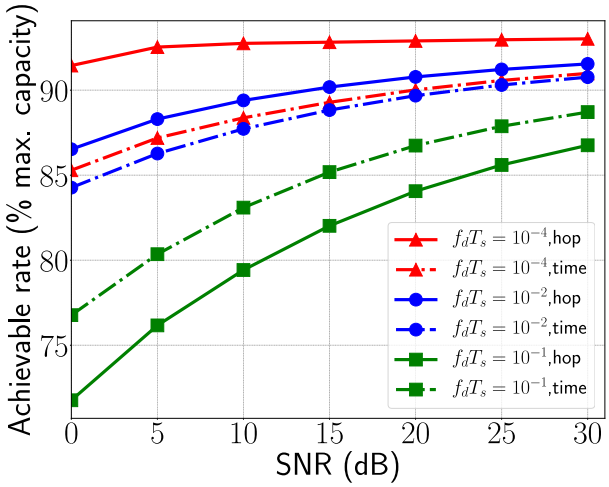
(a) BER results, Vehicular channel model



(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 [11], with feedback points as shown in Fig. 2, and interpolation performed via Cayley exponential lifting map as discussed in Section III.

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 $f_dT_s = 10^{-2}$ 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 ~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.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[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] 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.

[9] 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.

[10] 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

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

[12] 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.

[13] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb 2004.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

VII. Appendix

A. The Cayley Exponential Lifting and Retraction Pairs

Given $\mathbf{X}, \mathbf{Y} \in St(N_t, N_r)$, [10] 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 to a $N_t \times N_t$ skew Hermitian matrix.

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,

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

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 $\mathcal{T}_{i,t}, \mathcal{F}_{i,t}$,

$$\mathcal{F}_{i,t}, \mathcal{T}_{i,t} \leftarrow \mathrm{argmin} \sum_{(j,s)\in \mathtt{nbrs}_{i,t}(p)} ||(\mathcal{F}_{i,t}(j-i) + \mathcal{T}_{i,t}(s-t) - \mathbf{T}_{i,t}^{j,s})||_F^2 \tag{8}$$

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 $\Delta_{i,t}$