The Dissertation Committee for Takao Inoue
certifies that this is the approved version of the following dissertation:

Manifold Signal Processing for MIMO Communications

Committee:

Robert W. Heath, Jr., Supervisor

Jeffrey G. Andrews

Constantine Caramanis

Inderjit S. Dhillon

Haris Vikalo

Manifold Signal Processing for MIMO Communications

by

Takao Inoue, B.S.E.E., M.S.E.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2009

To my family.

Acknowledgments

First and foremost, I would like to express my sincere appreciation to my advisor Professor Robert W. Heath, Jr. for his invaluable guidance during my program at UT Austin. His vision and ethics for highest standard in research have profoundly inspired and helped my research. I am also very grateful for his interest and support he has given me in my entrepreneurial interests.

I would also like to thank my committee members, Professor Jeffrey G. Andrews, Professor Constantine Caramanis, Professor Inderjit S. Dhillon, and Professor Haris Vikalo for asking challenging questions and providing guidance at various points in my program. I am indebted to Professor Irwin W. Sandberg who has kindly provided me with the opportunity to study at UT Austin and taught me useful mathematical foundations. I am also indebted to Professor Elizabeth Keating for giving me an opportunity to work at the Science, Technology, and Society program. I gained a fresh perspective on how science and engineering influence societies.

I am very thankful to have had the chance to work with wonderful colleagues in WSIL. I would not have been here today without the daily interactions, technical discussions, support, and best of all, comradely with them. I want to especially thank the former WSIL colleagues Chan-Byoung Chae, Johann Chiang, Kaibin Huang, Caleb Lo, and Rahul Vaze for frequent discussions, collaborations, advices, and fun conversations. I also want to thank the current WSIL members, Salam Akoum, Omar Al Ayache, Ramya Bhagavatula, Robert Daniels, Insoo Hwang, Alvin Leung, Minchan Lee, Behrang Nosrat-Makouei, Steven Peters, Kien Trung Troung, Ali Yazdan-Panah, and Yongseok Yoo for being great group-mates during my program and various projects that we have undertaken together.

Finally, I am thankful to all of my family members. I would like to especially thank my parents, Shuji and Yuko Inoue, for their constant support and encouragement, and most importantly, believing in me throughout this journey. I would like to thank my siblings, Haruo and Tomoko Inoue, and my grandparents, Istuo and Kazuko Inoue and Dr. Yasuo Naito and Ayako (a.k.a. Yumiko) Naito, for their understanding and being my mental support. It is with deep regrets that I cannot personally share this dissertation with my grandparents except for Kazuko. Last but not least, I am greatly indebted to my dearest wife Chiho Sunakawa Inoue. I would not have completed my PhD study without her unconditional support and encouragement.

Takao Inoue

Manifold Signal Processing for MIMO Communications

Publication No. ______________

Takao Inoue, Ph.D.
The University of Texas at Austin, 2009

Supervisor: Robert W. Heath, Jr.

The coding and feedback inaccuracies of the channel state information (CSI) in limited feedback multiple-input multiple-output (MIMO) wireless systems can severely impact the achievable data rate and reliability. The CSI is mathematically represented as a Grassmann manifold or manifold of unitary matrices. These are non-Euclidean spaces with special constraints that makes efficient and high fidelity coding especially challenging. In addition, the CSI inaccuracies may occur due to digital representation, time variation, and delayed feedback of the CSI. To overcome these inaccuracies, the manifold structure of the CSI can be exploited. The objective of this dissertation is to develop a new signal processing techniques on the manifolds to harvest the benefits of MIMO wireless systems.

First, this dissertation presents the Kerdock codebook design to represent the CSI on the Grassmann manifold. The CSI inaccuracy due to digital representation is addressed by the finite alphabet structure of the Kerdock codebook. In addition, systematic codebook construction is identified which reduces the resource

requirement in MIMO wireless systems. Distance properties on the Grassmann manifold are derived showing the applicability of the Kerdock codebook to beam-forming and spatial multiplexing systems.

Next, manifold-constrained algorithms to predict and encode the CSI with high fidelity are presented. Two prominent manifolds are considered; the Grassmann manifold and the manifold of unitary matrices. The Grassmann manifold is a class of manifold used to represent the CSI in MIMO wireless systems using specific transmission strategies. The manifold of unitary matrices appears as a collection of all spatial information available in the MIMO wireless systems independent of specific transmission strategies. On these manifolds, signal processing building blocks such as differencing and prediction are derived. Using the proposed signal processing tools on the manifold, this dissertation addresses the CSI coding accuracy, tracking of the CSI under time variation, and compensation techniques for delayed CSI feedback. Applications of the proposed algorithms in single-user and multiuser systems show that most of the spatial benefits of MIMO wireless systems can be harvested.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A recent breakthrough in wireless communication is the use of multiple antennas at the transmitter and the receiver, or multiple-input multiple-output (MIMO) technology shown in Fig. 1.1. Driven by application in MIMO wireless systems, and in part by applications to commercial wireless systems [1, 41], a new class of quantization problems on the Grassmann manifold has recently emerged [15, 74, 79]. The Grassmann manifold arise, for example, as the mathematical space representing channel state information (CSI) under certain performance metrics in single user and multiuser limited feedback MIMO wireless systems [66]. The motivation for quantization on the Grassmann manifold is to encode the CSI at the receiver so that it can be efficiently communicated back to the transmitter through a finite rate feedback link [39, 67]. Communication theoretic performance measures such as capacity and bit error rate of MIMO wireless systems are known to depend on the number of quantization levels or the codebook size used to encode the CSI [67].

The manifold of unitary matrices plays an important role as a mathematical space representing CSI. While the Grassmann manifold appears as the space representing the CSI for specific transmission strategies, the manifold of unitary matrices appears as the collection of all spatial dimensions available in a given MIMO

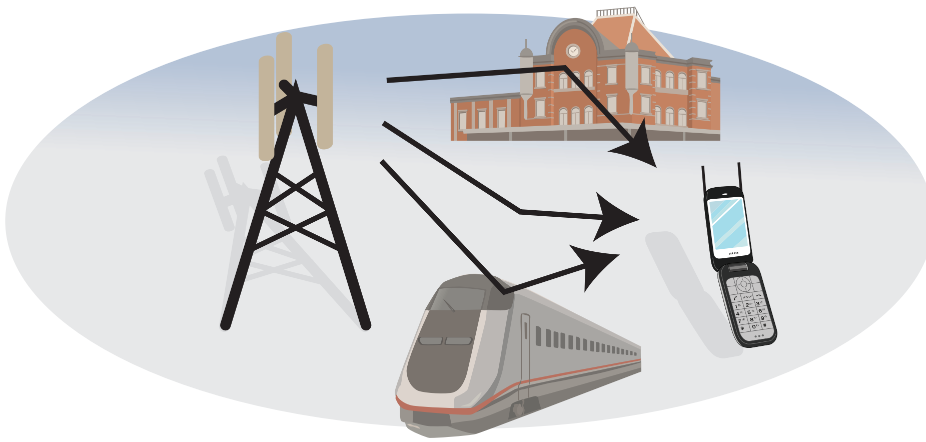

Figure 1.1: Illustration of single user MIMO with various propagation paths.

wireless system [65, 94]. Availability of all spatial dimensions can enable capacity achieving multimode precoding [65]. The quantization of these manifolds is especially challenging because of its large dimensions and non-Euclidean structure with special constraints.

In practice, the wireless communication channel may exhibit correlation due to mobility in the propagation channel, thus resulting in time varying CSI [38, 56]. In addition to quantization error, the time varying CSI and feedback delay may further aggravate the CSI inaccuracy. The focus of this dissertation is to develop new signal processing techniques on the manifolds with the goal of improving the CSI accuracy and provide robustness to time varying CSI and feedback delay.

In the remainder of this chapter, an overview of single user MIMO wireless systems is given in Section 1.1, an overview of multiuser MIMO wireless systems is given in Section 1.2, an overview of limited feedback is given in Section 1.3,

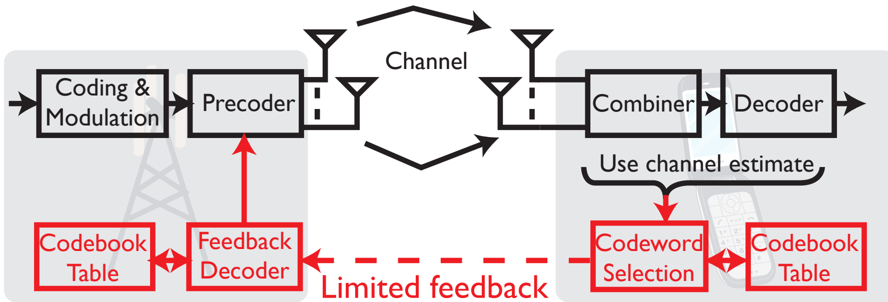

Figure 1.2: Block diagram illustrating limited feedback for single user MIMO system.

and an overview and motivation for signal processing on manifolds in Section 1.4, followed by the thesis statement, contributions, the organization of the dissertation, and notations in Sections 1.5, 1.6, 1.7, and 1.8, respectively.

1.1 Single User MIMO Wireless Systems

A single user limited feedback MIMO wireless systems consists of a transmitter and a receiver using multiple antennas. A general block diagram of the single user limited feedback MIMO wireless system is shown in Fig. 1.1. It represents a basic model of point-to-point communication encountered, for example, in cellular communication systems [2], metropolitan wireless networks [41, 43], and wireless local area networks [42]. It has been shown that N times the data rate of the same link with single antenna system is possible when at least N antennas are used at the transmitter and the receiver [21].

The benefits and challenges of MIMO wireless systems both arise from the multiple propagation path between each pair of antennas from the transmitter to the receiver, called the MIMO channel. The baseband equivalent sampled model of the MIMO channel may be written as a matrix with coefficients representing the effects of fading between the pairs of antennas. The MIMO channel matrix contains the CSI, which if used at the transmitter, can improve the downlink data rates and reliability by customizing the transmit signal to the specific MIMO channel [24, 64, 67, 94].

One transmission strategy is to use limited feedback unitary precoding where the signal at the transmitter is multiplied by a unitary vector or a unitary matrix from a codebook shared between the receiver and the transmitter [64, 67]. Beamforming corresponds to transmission of single stream of data precoded by an $n \times 1$ unitary vector [67]. Unitary precoded spatial multiplexing corresponds to sending $p < n$ data streams precoded by an $n \times p$ unitary matrix [64]. The selected precoder is used to customize the transmit signal to exploit the desired number of strongest spatial modes available in the MIMO channel without changing the transmit power. In general, the precoder must be designed with the knowledge of the MIMO channel, usually available at the receiver. Thus, the receiver is tasked with a quantization and feedback of the precoder information.

It has been shown that under signal to noise ratio metric, the unitary precoders are invariant to unitary rotations thus corresponding to the Grassmann manifold [64, 67, 79]. The square unitary right singular matrix of a MIMO channel corresponds to the manifold of unitary matrices representing all the available spa-

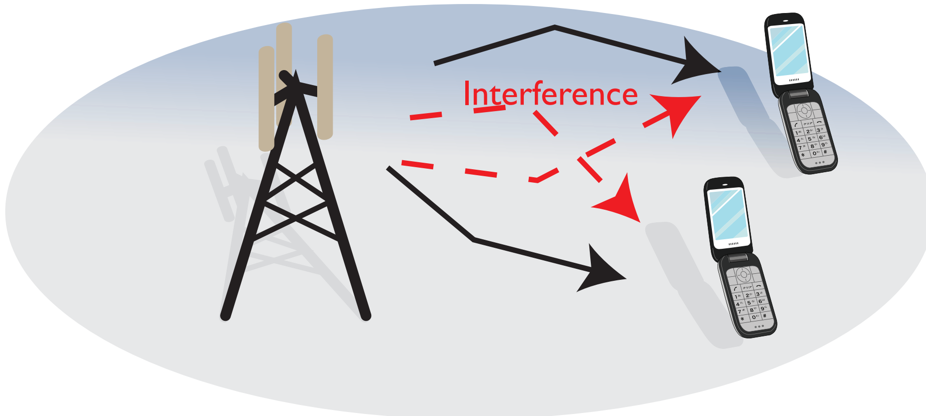

Figure 1.3: Illustration of multiuser MIMO with various propagation paths and interference.

tial modes in the MIMO channel. The main theme of this dissertation is to develop high fidelity and practical CSI encoding techniques on these manifolds.

1.2 Multiuser MIMO Wireless Systems

In recent years, multiuser MIMO wireless system has emerged as a new paradigm in using multiple antennas [23]. It involves a transmitter and multiple receivers as shown in Fig. 1.3. In the downlink, from the transmitter to multiple users, the data stream intended for each user is multiplexed spatially while sharing the same time and frequency. Thus multiuser systems promises improved spectral resource usage. It has been shown that the achievable sum rate, i.e., the sum of all data rate delivered to the receivers, multiplies proportional to number of antennas at the transmitter [23]. For maximizing the sum rate, dirty paper coding has been known to be optimal [14]. Unfortunately, this technique requires non-causal

CSI which is not realizable in practice. A more practical approach uses transmit beamforming based on zero forcing. If perfect CSI is available at the transmitter, zero forcing perfectly avoids interferences between users [51]. The method for CSI feedback in multiuser MIMO wireless systems differs from single user cases because the precoders used for zero forcing must be designed with the scheduled user's CSI. The transmitter collects the CSI from all users to design zero forcing precoder. The consequence is that CSI quantization error is magnified by the zero forcing computation, i.e., matrix inverse, resulting in residual interuser interference. It has been shown using random codebook argument that the feedback bits should be increased linearly with signal to noise ratio (SNR) to achieve the full sum rate benefits [51]. Therefore, to harvest the achievable sum rate available in multiuser MIMO systems, search for high fidelity CSI feedback technique remains to be an active research area. In Chapter 3, a new Grassmannian predictive coding strategy is shown to provide significant improvement in sum rate using a comparable number of feedback bits.

1.3 Limited Feedback

As described in Section 1.1 and 1.2, feedback of CSI is important in obtaining achievable benefits of MIMO wireless systems. Limited feedback is a practical approach to obtain the CSI at the transmitter at the possible cost of CSI inaccuracy due to quantization error [66]. The quantization of CSI is especially challenging because of its manifold structure. In addition, CSI inaccuracies may arise due to time variations and delayed feedback. These sources of inaccuracies motivate to take a more sophisticated signal processing approach on the manifold to fully reap benefits of the MIMO wireless systems.

The most widely employed limited feedback strategy uses *codebooks* known to the transmitter and the receiver, shown in Fig. 1.2. The codebook consists of multiple precoders representing the quantized CSI with indices assigned to them. In single user systems, the precoder with a maximum product norm with the MIMO channel matrix is selected [64, 67]. In multiuser system, the codeword with minimum chordal distance between the codeword and the normalized channel vector is selected. Then, the index of the selected precoder is communicated to the transmitter, resulting in the limited feedback. For memoryless feedback, the size of the codebook determines the number of bits to be fed back. Fortunately, using just a few bits of feedback, limited feedback precoded systems has been shown to provide substantial SNR gains over non-precoded systems [79].

Motivated by this simple and effective strategy, current wireless standards such as IEEE 802.16e [41] and 3GPP LTE [2] have adopted the limited feedback strategies. Future wireless standards such as IEEE 802.16m [43] and 3GPP LTE-Advanced [3] continue to actively consider viable limited feedback strategies. These practical systems are usually implemented on digital integrated circuits where CSI inaccuracy due to quantization may be aggravated by round off errors in digital systems. To address CSI inaccuracies due to the implementation, new codebooks suitable for digital systems are presented in Chapter 2.

1.4 Signal Processing on Manifolds

The key property of the $n \times p$ unitary precoder under SNR metric in single user MIMO systems, is that it can be right multiplied by an arbitrary $p \times p$ unitary matrix without affecting the system performance [64, 67, 79]. The mathematical space for such rotationally invariant precoders is represented by the Grassmann manifold. It is a collection of p-dimensional subspaces embedded in n-dimensional Euclidean space in which calculus can be performed. Thus each point on the Grassmann manifold represents a subspace. There are other equivalent definitions for Grassmann manifold each with its own insights. For example, it may be defined as a quotient space of unitary group, i.e., $\mathcal{U}_n/(\mathcal{U}_{n-p} \times \mathcal{U}_p)$, or as collection of projection matrices. This seemingly complicated mathematical space has attracted research in analysis [8], quantization [74], coding [104], and optimization [5] on manifolds.

The manifold of unitary matrices is essentially the unitary group. The manifold designation is used to distinguish the fact that the differential geometric properties from Lie theory are used. Lie group of unitary matrices is a unitary group with differential geometric structure that is compatible with group operations [89]. This is an important distinction from conventional unitary group because the differential geometric structure provides more freedom to move about on the manifold.

Early work on limited feedback codebook designs has focused on quantizing the Grassmann manifold to obtain a fixed optimal codebook. The optimal codebook on the Grassmann manifold for block fading Gaussian channel is given by isotropically distributed points on the manifold. It turns out that this is a classical problem in algebraic geometry called sphere packing [13]. Most codebook designs employ numerical search or vector quantization. Suboptimal systematic codebooks do exist based on Fourier construction [35]. In Chapter 2 the problems of systematic codebook construction and generation are addressed.

Real world propagation channels are dynamic due to mobility in the channel. This has motivated a different direction in codebook designs that is dynamic and adaptive to the changing environment, thereby reducing the CSI inaccuracies. Unfortunately, due to the manifold structure of CSI, conventional signal processing and adaptive techniques do not immediately extend to the Grassmann manifold. A novel approach to adapt the codebook is to exploit the group theoretic structure of the Grassmann manifold. Using translation and scaling defined on the Grassmann manifold, adaptive codebooks have been shown to improve system performances over fixed codebook regime [34, 85]. The approach taken in this dissertation is to view the time evolution of the CSI as a time series evolving on the Grassmann manifold. This is graphically depicted in Fig. 1.4. As depicted in Fig. 1.4, the Grassmann manifold can be envisioned with a hyper-spherical geometry. Thus linear signal processing operations such as addition and multiplication are not well defined. This motivates the need for intrinsic signal processing techniques on the manifold. Questions arise; How can we quantify the difference or error between two points? What is a sensible way to predict given two points? In Chapter 3, we develop basic signal processing tools to arrive at predictive coding algorithm on the Grassmann manifold.

Finally, the Grassmann manifold characterization of CSI is associated with

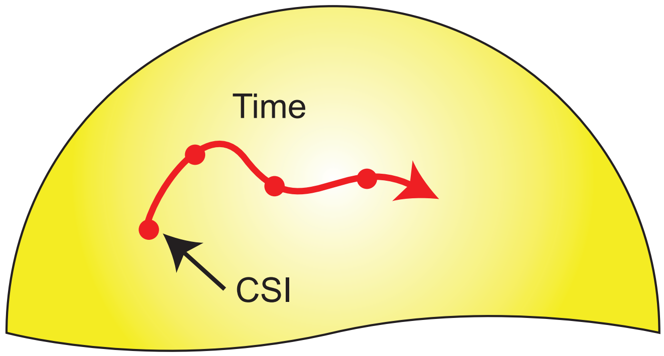

Figure 1.4: Graphical depiction of CSI evolution over time on the Grassmann manifold with non-Euclidean structure.

specific transmission strategy, e.g., beamforming and spatial multiplexing. The drawback is that multiple codebooks or multiple adaptation techniques have to be implemented for each mode of transmission strategies. If, however, a single representation for all the available spatial features in the MIMO channel can be made available, the MIMO wireless systems can further benefit from capacity achieving techniques such as multimode precoding [65]. The collection of all the spatial features of a given MIMO channel gives rise to the manifold of unitary matrices. The manifold of unitary matrices, considered as a manifold in the space of invertible complex matrices, is a collection of square unitary matrices in which calculus can be performed. Motivated by the same questions for the Grassmann manifold, basic signal processing tools, differential coding, and predictive coding techniques on the manifold of unitary matrices are developed in Chapter 4.

1.5 Thesis Statement

Correlation and structure of the signals evolving on the Grassmann manifold and the manifold of unitary matrices can be exploited to obtain high resolution predictive coding techniques and to improve throughput of MIMO communication systems.

1.6 Contributions

The contributions of this dissertation may be summarized as follows.

- **Kerdock Codebook Design** A new single user codebook design suitable for both beamforming and unitary precoded spatial multiplexing systems is proposed. The proposed codebook has systematic construction, reduced storage, and search enabled by finite alphabet structure. Sylvester-Hadamard construction and power method to systematically generate the codebooks are shown. Closed form distance properties of the codebook are derived which shows that it performs similarly to previously known codebooks. Numerical results show that symbol error rate and achievable rate similar or better than previously known floating point codebooks are obtained.

- **Signal Processing on the Grassmann Manifold** Basic signal processing building blocks such as differencing, mapping onto the Grassmann manifold, parallel transport, and optimal prediction frameworks are derived. Using these building blocks, a Grassmannian predictive coding is proposed. Furthermore, a predictive coding algorithm suitable for delayed feedback system is developed with step size prediction to optimize the predicted estimate. Based on the geometric interpretation of the algorithm, distortion bounds are derived. Distortion bounds show that the proposed framework provides substantial distortion improvement over memoryless techniques. Application to single user limited feedback beamforming system shows that symbol error rate approaching the perfect CSI case can be obtained. Furthermore, for limited feedback multiuser MIMO system with zero forcing, the proposed algorithm achieves multiplexing gain as a function of temporal correlation exceeding that of memoryless approach.

- **Signal Processing on the Manifold of Unitary Matrices** Basic signal processing building blocks such as differencing, mapping onto the manifold of unitary matrices, and optimal prediction frameworks are derived. Using these building blocks, a differential coding and predictive coding are proposed. Error quantization is performed on the tangent space to the unitary manifold which reduces the number of parameter by a half. Applications to single user limited feedback unitary precoded system operating in temporally correlated channels show that the proposed algorithm provides high fidelity CSI that is independent of the rank. Unitary precoders derived from the feedback for various ranks show symbol error rate performance approaching the perfect CSI case. Furthermore, the proposed predictive coding algorithm applied to limited feedback-based block diagonalization in multiuser MIMO system is shown to provide substantial sum rate improvement for mild temporal correlation.

1.7 Organization

In Chapter 2, design of Kerdock codebook, derivations, and simulation results are provided. In Chapter 3, manifold-constrained signal processing tools and predictive coding results for Grassmann manifold are presented. In Chapter 4, manifold-constrained signal processing tools, differential and predictive coding techniques, and predictions for delayed feedback on unitary manifold are presented. Finally, concluding remarks and future work are presented in Chapter 5.

1.8 Notations

Lower case bold letters e.g., $\mathbf{v}$, are used to denote vectors and upper case bold letters e.g., $\mathbf{H}$, are used to denote matrices. The norms of vectors or matrices are denoted by $\|\cdot\|$ (or $\|\cdot\|_2$) and $\|\cdot\|_F$ for the usual 2-norm and Frobenius norm, respectively. The trace of a matrix is denoted tr$(\cdot)$. The space of integers, real numbers, and complex numbers are denoted $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively, with the appropriate superscript to denote the dimensions of the space. The real part of a complex number is denoted by $\Re(\cdot)$. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$. Superscripts ^{T}, *, and † are used to denote the transposition, Hermitian transpose, and pseudo inverse, respectively. The n-th column entry of a matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_n$ and a subset of columns by $[\mathbf{A}]_{1:n}$. The expectation is denoted $\mathbb{E}[\cdot]$.

Chapter 2

Kerdock Codebook for Limited Feedback MIMO Systems

2.1 Prior Work

Channel state information (CSI) at the transmitter can provide considerable capacity and resilience to channel fading in multiple-input multiple-output (MIMO) systems [66]. A practical solution to provide CSI to the transmitter is a codebook based feedback strategy, known as the limited feedback [64, 66, 67, 78, 79]. In a limited feedback system, the receiver searches for the appropriate transmit precoder from a finite set of precoders, called the *codebook*, shared by the transmitter and the receiver. Then, the receiver sends the index of the codeword back to the transmitter resulting in the feedback of quantized CSI. In practice, three to six bits of feedback are common to balance the benefits of limited feedback and feedback overhead tradeoff [2, p. 39], [41, pp. 457-466].

There are several codebook designs in the literature such as those based on vector quantization [52,79,87], Grassmannian packing [64,67,78], discrete Fourier transform [64, 67], and quadrature amplitude modulation [90]. Unfortunately, the codebook design often requires numerical iterations and optimizations. Furthermore, codebooks for beamforming and spatial multiplexing needs to be separately stored in a memory. In today's hand-held devices with limited memory, size, and power [80], limited feedback codebooks with smaller memory footprint will help to reduce implementation costs. Furthermore, reduced search computation will ease stringent computational timing requirement in real-time system and allow the system to quickly adapt to highly mobile environments. The codebooks adopted for recent standards illustrates trends towards systematic finite alphabet codebooks [2, p. 39], [41, pp. 457-466].

2.2 Contributions

In this chapter, a new single user codebook design is proposed for limited feedback unitary precoded MIMO systems, called the Kerdock codebook, due to the Kerdock code construction with quaternary alphabet [28,33,54]. The main contribution is to identify the Kerdock codebook as a new avenue of codebook design with additional benefits of systematic construction, reduced storage, and search enabled by the finite alphabet construction. Reduced storage is made possible in two parts: 1) by finite alphabet, and 2) by deriving spatial multiplexing codebook from beamforming codebook. The distance properties for Kerdock codebook are derived showing that it performs similarly to previously known codebooks. Two practical examples of codebooks are shown for two and four transmit antennas using two different constructions: a Sylvester-Hadamard construction [33] and a power construction [26]. The Sylvester-Hadamard construction gives a good solution for the two antenna case while the structure in the power construction gives a better solution for the four antenna case and permits closed form derivation of subspace distance prop-

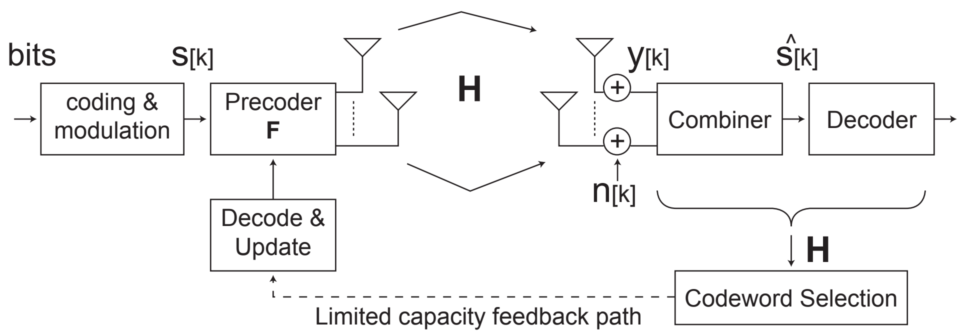

Figure 2.1: Block diagram of general limited feedback MIMO System

erties. The proposed construction can be extended to matrices with dimensions that are power of two. Compared with prior work in [64, 67, 87, 90], our approach provides systematic codebook construction with finite alphabet where single codebook can be used for both beamforming and spatial multiplexing transmissions.

2.3 System Model

2.3.1 Discrete-time System Model

A limited feedback precoded MIMO wireless system with N_t transmit antennas and N_r receive antennas is shown in Fig. 2.1. Let k denote the time index and N_s denote the number of spatial streams being used. The case when $N_s = 1$ is called beamforming and the general case when $1 < N_s \leq N_t$ is called N_s-stream spatial multiplexing. The transmit bit stream is sent to the encoder and modulator which outputs a complex transmit vector, $\mathbf{s}[k] = [s_1[k], s_2[k], \ldots, s_{N_s}[k]]^T$. The average power is assumed to be constrained as $\mathbb{E}_{\mathbf{s}}[\mathbf{s}\mathbf{s}^*] = \frac{\mathcal{E}_s}{N_s}\mathbf{I}_{N_s}$ where $\mathbb{E}_{\mathbf{s}}$ is used to denote the expectation with respect to the transmit vector s and $\mathcal{E}_s$ is used to denote the total transmit power. The transmit vector $\mathbf{s}[k]$ is multiplied by the unitary precoder $\mathbf{F}[k] \in \mathbb{C}^{N_t \times N_s}$ ($\mathbf{f}[k] \in \mathbb{C}^{N_t \times 1}$ for beamforming) with unitary constraint, $\mathbf{F}^*[k]\mathbf{F}[k] = (1/N_s)\mathbf{I}_{N_s}$, producing a length N_t transmit vector $\mathbf{x}[k] = \sqrt{\mathcal{E}_s/N_s}\mathbf{F}[k]\mathbf{s}[k]$. The precoder $\mathbf{F}[k]$ is selected based on the limited feedback information.

Assuming perfect synchronization, sampling, and a linear memoryless channel, the equivalent baseband input-output relationship is

$$\mathbf{y}[k] = \sqrt{\mathcal{E}_s/N_s}\mathbf{H}[k]\mathbf{F}[k]\mathbf{s}[k] + \mathbf{n}[k]$$

where $\mathbf{H}[k] \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix and $\mathbf{n}[k]$ is the additive noise vector. The entries of $\mathbf{n}[k]$ are assumed to be complex Gaussian independent and identically distributed (i.i.d.) according to $\mathcal{CN}(0, N_0)$. The receive vector $\mathbf{y}[k]$ is then decoded by assuming a perfect knowledge of $\mathbf{H}[k]\mathbf{F}[k]$ at the receiver to produce the output vector ŝ.

2.3.2 Codeword Search

Based on the estimate of the channel and the receiver structure, the receiver chooses the best precoding codeword $\hat{\mathbf{F}}[k]$ from a set of N possible codewords in the codebook $\mathcal{F} = \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N\}$ shared by the transmitter and the receiver. The codeword index is represented by $b = \lceil \log_2 N \rceil$ bits resulting in b-bit feedback.

The beamformer that minimizes the probability of symbol error for maximum ratio combining receiver is [67]

$$\hat{\mathbf{f}}[k] = \arg\max_{\mathbf{f} \in \mathcal{F}} \|\mathbf{H}[k]\mathbf{f}\|_2^2. \tag{2.1}$$

For spatial multiplexing with a zero forcing receiver, the minimum singular value selection criteria is often used [64]

$$\hat{\mathbf{F}}[k] \quad = \quad \arg\max_{\mathbf{F}\in\mathcal{F}} \lambda_{\min}\{\mathbf{H}[k]\mathbf{F}\} \tag{2.2}$$

where $\lambda_{\min}\{\cdot\}$ denotes the minimum singular value of the argument. This selection criteria approximately maximizes the minimum substream signal to noise ratio (SNR).

The Grassmannian beamforming criterion states that the beamforming codebook should be designed such that the minimum pairwise chordal distance is maximized [67]. Therefore, the chordal distance is used to analyze the distance property of the beamforming codebook. The chordal distance between codeword vectors, $\mathbf{f}_1$ and $\mathbf{f}_2$, is given by

$$d_{\text{ch}}(\mathbf{f}_1, \mathbf{f}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\mathbf{f}_1^* \mathbf{f}_2|^2}. \tag{2.3}$$

For spatial multiplexing, projection 2-norm distance, among many possible distance metrics [64], is used to evaluate the spatial multiplexing codebooks. It was shown in [64] that the codebook should be designed by maximizing the minimum projection 2-norm distance

$$\begin{aligned} d_{p2}(\mathbf{F}_1, \mathbf{F}_2) &= \|\mathbf{F}_1\mathbf{F}_1^* - \mathbf{F}_2\mathbf{F}_2^*\| \\ &= \sqrt{1 - \lambda_{\min}\{\mathbf{F}_1^*\mathbf{F}_2\}} \end{aligned} \tag{2.4}$$

between a pair of codewords to approximately maximize the minimum substream SNR.

2.4 Kerdock Codebook Design

Quaternary alphabet Kerdock codes were originally proposed as error correcting codes [28] and are known to be mutually unbiased bases (MUB) [84]. MUB contains orthonormal bases satisfying mutually unbiased property. Some of the known MUB constructions can be found in [26, 59, 84]. It was shown in [60] that many of the MUB constructions are equivalent and that these constructions have a close connection with complex projective space and uniform tight frames, both of which have been used for the construction and analysis of quantized codebooks for limited feedback MIMO systems. Based on these connections, the utility of Kerdock codes and MUB as limited feedback codebooks are studied.

If $\mathbf{S} = [\mathbf{s}_1 \ \dots \ \mathbf{s}_{N_t}]$ and $\mathbf{U} = [\mathbf{u}_1 \ \dots \ \mathbf{u}_{N_t}]$ are two $N_t \times N_t$ orthonormal bases (i.e. $\mathbf{S}^*\mathbf{S} = \mathbf{I}_{N_t}$), the column vectors drawn from each orthonormal basis are said to satisfy the *mutually unbiased* property if $|\langle \mathbf{s}_n, \mathbf{u}_m \rangle| = 1/\sqrt{N_t}$ for $n, m = 1, \dots, N_t$. An MUB is the set $\mathcal{S} = \{\mathbf{S}_0, \mathbf{S}_1, \dots\}$ satisfying the mutually unbiased property. The maximum number of orthonormal bases, i.e. $|\mathcal{S}|$, has been shown to be less than or equal to $N_t + 1$ for any N_t; a sufficient condition for equality is that N_t is a power of a prime [59]. It is presently unknown whether equality occurs when N_t is not a power of prime and this question remains to be an active area of research [26]. Several approaches for the construction of size $N_t + 1$ MUB for prime powers have been proposed [84].

2.4.1 Sylvester-Hadamard Construction

One approach for Kerdock code construction was proposed in [33]. The construction consists of generating an $N_t \times N_t$ diagonal matrices $\mathbf{D}_n$ for $n = 0, 1, \ldots, N_t - 1$ which are used to transform $N_t \times N_t$ Sylvester-Hadamard matrix. Each transformed matrix $\mathbf{S}_n$ becomes the orthonormal basis. The benefit of this approach is that algebraic construction of the diagonal matrix using $\mathbb{Z}_4$ quadratic maps is available.

Let $\hat{\mathbf{H}}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ denote the Sylvester-Hadamard matrix. The $N_t \times N_t$ Sylvester-Hadamard matrix such that $N_t = 2^B$ is given by the Kronecker product of B Sylvester-Hadamard matrix

$$\hat{\mathbf{H}}_{N_t} = \underbrace{\hat{\mathbf{H}}_2 \otimes \hat{\mathbf{H}}_2 \cdots}_{B\text{times}}. \tag{2.5}$$

The general strategy for the Kerdock codebook construction is:

1. Construct the diagonal generator matrices $\mathbf{D}_n$ for $n = 0, 1, \ldots, N_t - 1$.
2. Compute the basis $\mathbf{S}_n = (1/\sqrt{N_t})\mathbf{D}_n \hat{\mathbf{H}}_{N_t}$.
3. Let $\mathcal{S} = [\mathbf{S}_0 \; \mathbf{S}_1 \; \cdots \; \mathbf{S}_{N_t-1}]$.

For brevity, the details of the construction according to [33] are omitted.

2.4.2 Power Construction

Another attractive MUB construction using a single generator matrix was recently proposed by Gow [26]. The construction uses advanced concepts from finite groups and representation theory. In particular, the generating basis arises as an automorphism of an extra-special 2-group which is a group structure that also appears in the construction of Kerdock codes [33]. Let N_t be a power of two. The following theorem was proved by Gow [26]:

Theorem 1. *If* $\mathbf{D}$ *is an invertible unitary* $N_t \times N_t$ *matrix that satisfies* $\mathbf{D}^{N_t+1} = \mathbf{I}$ *and the determinant of* $\mathbf{D}$ *is equal to* 1*, then the powers* $\mathbf{D}, \mathbf{D}^2, \ldots, \mathbf{D}^{N_t+1} = \mathbf{I}$ *generates* $N_t + 1$ *pairwise mutually unbiased bases. Furthermore, all entries of* $\mathbf{D}$ *are in the quaternary alphabet.*

Theorem 1 is an existence theorem which states that if $\mathbf{D}$ satisfies the indicated mild conditions, then the powers $\mathbf{D}, \mathbf{D}^2, \ldots, \mathbf{D}^{N_t+1} = \mathbf{I}$ generates $N_t + 1$ pairwise mutually unbiased bases with quaternary entries. The contribution here is that a generator $\mathbf{D}$ obtained from the Sylvester-Hadamard construction in Section 2.5 is identified.

From the limited feedback codebook design perspective, Theorem 1 represents a powerful result when the number of transmit antennas are power of 2. Only the generating base $\mathbf{D}$ needs to be stored and the rest of the codebook can be generated by taking the powers. Note also the inclusion of the identity element which corresponds to the case of antenna subset selection [32]. Prior codebook designs do not include the identity element as part of the unified codebook design.

2.4.3 Codebook Arrangement

The construction of multi-stream codebooks using the special structure of $\mathcal{S}$ is shown next. For beamforming, the codebook is constructed by taking the columns of each basis

$$\mathcal{F} = \{\mathbf{f}_1 = [\mathbf{S}_0]_1, \mathbf{f}_2 = [\mathbf{S}_0]_2, \ldots, \mathbf{f}_N = [\mathbf{S}_{N_t}]_{N_t}\} \tag{2.6}$$

where $N \leq N_t \cdot (N_t + 1)$.

For spatial multiplexing, unique column combinations are selected from each $\mathbf{S}_n$ to form the codebook. Note that every column combination yields a unitary matrix. Specifically, for an N_s-stream spatial multiplexing codebook, the largest codebook is derived by taking all N_s-column combinations from each $\mathbf{S}_n$. There are $\binom{N_t}{N_s}$ column subset combinations in each $\mathbf{S}_n$. The maximum number of codewords that the MUB can take is $(N_t+1) \times \binom{N_t}{N_s}$. Smaller codebook can be obtained, for example, by taking the subset of the largest codebook which maximizes minimum distance between codewords. In Section 3.5, the distance properties of this codebook are derived. In Section 4.7, this codebook is shown to perform favorably to same sized Grassmannian codebook through Monte Carlo simulations.

2.5 Kerdock Codebook Examples

2.5.1 Two Transmit Antenna Construction

For the two antenna MIMO system, the Sylvester-Hadamard construction is used. The resulting bases are

$$\mathbf{S}_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \mathbf{S}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}, \mathbf{S}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.7}$$

where S_0 is the scaled Sylvester-Hadamard matrix. The beamforming codebook is constructed as $\mathcal{F} = \{\mathbf{f}_1 = [\mathbf{S}_0]_1, \mathbf{f}_2 = [\mathbf{S}_0]_2, \ldots, \mathbf{f}_6 = [\mathbf{S}_2]_2\}$.

2.5.2 Four Transmit Antenna Construction

For the four antenna MIMO system, a power construction based codebook is used. Starting with the Sylvester-Hadamard construction and making a slight modification to one of the bases, the following generator matrix satisfies Theorem 1

$$\mathbf{D} = \frac{1}{2}\begin{bmatrix} -j & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -j & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}\left(\hat{\mathbf{H}}_2 \otimes \hat{\mathbf{H}}_2\right). \tag{2.8}$$

Finally, computing $\mathbf{S}_n = \mathbf{D}^{n+1}$ for $n = 0, \cdots, 4$ yields the following bases

$$\begin{aligned} \left[\mathbf{S}_0 \vdots \mathbf{S}_1\right] &= \frac{1}{2}\left[\begin{array}{cccc:cccc} -j & -j & -j & -j & -1 & -1 & -j & j \\ 1 & -1 & 1 & -1 & -j & -j & -1 & 1 \\ -j & -j & j & j & -j & j & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & j & j \end{array}\right] \\ \left[\mathbf{S}_2 \vdots \mathbf{S}_3\right] &= \frac{1}{2}\left[\begin{array}{cccc:cccc} -1 & j & j & 1 & j & 1 & j & -1 \\ -1 & j & -j & -1 & j & -1 & j & 1 \\ j & -1 & -1 & -j & j & 1 & -j & 1 \\ -j & 1 & -1 & -j & j & -1 & -j & -1 \end{array}\right] \\ \mathbf{S}_4 &= \mathbf{I}_4. \end{aligned} \tag{2.9}$$

For the beamforming system, a codebook of size $N = 20$ (5-bit codebook) is obtained. Without antenna selection, the identity element $\mathbf{S}_4$ can be deleted reducing the codebook size to 4 bits. For $N_s = 2$ spatial multiplexing system, $5 \times \binom{4}{2} = 30$ codewords or 5-bit codebook is obtained. For $N_s = 3$, $5 \times \binom{4}{3} = 20$ codewords (5-bit codebook) is obtained. Thus, a finite alphabet codebook which can be shared for beamforming and spatial multiplexing is obtained. Furthermore, the proposed codebooks satisfy the per antenna power constraints because equal energy is distributed across the transmit antennas. In contrast, Grassmannian codebooks may distribute energy unevenly among the transmit antennas.

2.6 Codebook Storage and Search Complexity

The on-chip memory of baseband processors, often on the order of few kilobytes to a megabyte, are extensively used for storing instructions and intermediate data that are being processed. The on-chip memory may not have sufficient room to maintain all the codebooks. Consequently, it is likely that larger codebooks will be stored on an off-chip memory which takes time to load. Therefore the storage requirement of a codebook has a significant impact on the implementation. Furthermore, it is well known that a multiplication take more clock cycles than an addition. Consequently, reducing multiplication in the baseband processor helps to meet the stringent timing requirement and computational load. The arithmetic logic unit often supports additional modes of operations such as addition, shift and sign change which can be used to exploit the benefits of proposed codebook and reduce the computational clock cycles. While the specific benefits achieved by multiplier-free codebook are implementation specific, the storage and search complexity of the proposed codebook in terms of storage bits and arithmetic operations are quantified next.

2.6.1 Storage

For codebook storage, the number of real elements needed to store a codebook for each mode of transmission is considered. If N_b is the number of bits available in the system to represent a real number, the storage required for a single N-entry codebook is $2N_bNN_tN_s$ bits. Some reduction may be possible due to specific values taken on by the codeword entries, but only the worst case scenario is considered for comparison. The Grassmannian codebook [63, 64, 67], without any structure, result in $2N_bNN_tN_s$ bits of storage for each codebook. The Fourier codebook [35] requires the generator matrix and the discrete Fourier transform matrix to be stored resulting in $2N_b(N_t + N_tN_s)$ bits for each codebook. Note that the storage requirement is independent of the codebook size because the generator matrix is designed for a given codebook size.

For the $N_t = 2$ Kerdock codebook, a total of 8 bits of storage is required where 4 bits each are used for the Sylvester-Hadamard matrix and $\mathbf{D}_1$. For $N_t = 4$, total of 12 bits of storage is required where 4×2 bits are used for the diagonal matrix and 4 bits are used for the Hadamard matrix in (2.8). Note that the Kerdock codebook storage is independent of N_b and the same codebook can be used for beamforming and spatial multiplexing. For a fair comparison, Table 2.1 shows the number of bits required to store the Kerdock, Fourier, and Grassmannian codebooks for $N_t = 4$ using $N = 16$ for beamforming and $N = 8$ for 2-stream spatial multiplexing. The Kerdock codebook provides a small fixed storage requirement independent of the system specific word size.

Table 2.1: Number of bits required for storing proposed Kerdock, Fourier, and Grassmannian codebooks for $N_t = 4$ and using $N = 16$ for beamforming and $N = 8$ for 2-stream spatial multiplexing. A system dependent number of bits which are used to represent a real number is denoted by N_b.

MUB	Fourier	Grassmannian
12	$40N_b$	$256N_b$

2.6.2 Search Complexity

For search complexity, the number of arithmetic computation required to arrive at the desired codeword is considered. It is assumed that (2.1) is tested for beamforming and (2.2) is tested for spatial multiplexing with the estimated channel matrix. Since the norm computations are common for all codebook entries, the computation required to compute **Hf** for (2.1) and **HF** for (2.2) for each codeword in the codebook are compared. The proposed Kerdock codebook with quaternary alphabet reduces the complex multiplication into either a sign change or swapping the real and imaginary part with sign change, eliminating the need for complex multiplication.

For beamforming, the Grassmannian and Fourier based codebooks require NN_tN_r complex multiplies and $NN_r(N_t-1)$ complex additions to find all the candidate effective channel gains. Meanwhile, the proposed Kerdock codebook does not require any complex multiplication and it only requires $NN_r(N_t-1)$ complex additions. Similarly, for spatial multiplexing, the Grassmann and Fourier based codebooks require $NN_sN_r{}^2$ complex multiplies and $NN_r{}^2(N_s-1)$ complex additions while the proposed Kerdock codebook does not require any complex multiplication with same number of complex additions. Therefore the Kerdock codebook eliminates complex multiplications which helps to reduce computational cycles for resource limited mobile terminals.

2.7 Relationship with Previous Designs

In this section, the distance properties of rank one, two and three Kerdock codebooks are first derived. Next, the proposed Kerdock codebook is shown to have full diversity. Finally, an achievable rate analysis of the proposed Kerdock codebook is provided.

2.7.1 Distance Properties

The distance properties of the codebook can be derived from the mutually unbiased property.

Lemma 2. *For any pair of beamforming Kerdock codewords $\mathbf{f}_k$ and $\mathbf{f}_l$ for $k, l = 1, 2, \ldots, N$ the chordal distance is either 1 when $\mathbf{f}_k$ and $\mathbf{f}_l$ are from the same basis or $\sqrt{1 - \frac{1}{N_t}}$ when $\mathbf{f}_k$ and $\mathbf{f}_l$ are from different bases.*

Next, consider the derived spatial multiplexing codebook and examine the projection 2-norm distance property. For any N_t that is power of two and $N_s = 2$ spatial multiplexing codebook based on the power construction, the following property is obtained.

Property 3. *Let $\mathbf{F}_k$ and $\mathbf{F}_l$, $k \neq l$, be $N_t \times 2$ matrices composed by taking two* *columns from any power of* $\mathbf{D}$. *Then,*

$$|\det(\mathbf{F}_k^*\mathbf{F}_l)| = \begin{cases} 0, & \textit{when } \mathbf{F}_k \textit{ and } \mathbf{F}_l \textit{ are from the same basis} \\ 1/\sqrt{N_t}, & \textit{otherwise}. \end{cases} \quad (2.10)$$

See Appendix 2.10.1 for the proof. The projection 2-norm distance (2.4) increases as the minimum singular value of $\mathbf{F}_1^*\mathbf{F}_2$ is decreased. Property 3 indicates that the proposed Kerdock codebook for $N_s = 2$ has only two possible projection 2-norm distance between the codewords.

Now consider $N_t = 4$ and $N_s = 3$ spatial multiplexing codebook.

Property 4. *Let* $\mathbf{F}_k$ *and* $\mathbf{F}_l$*,* $k \neq l$*, be* 4×3 *matrices by selecting any* 3 *columns from each* $\mathbf{S}_n$. *Then,*

$$|\det(\mathbf{F}_k^*\mathbf{F}_l)| = 1/2. \quad (2.11)$$

See Appendix 2.10.2 for the proof. The result may appear trivial from mutually unbiased property, but the fact the determinant exhibits this property guarantees that the codebook exhibits fixed projection 2-norm distance between the codewords. Unfortunately, the proof only applies for $N_t = 4$.

2.7.2 Diversity

The diversity order is an important performance metric that indicates the probability of symbol error trends for high SNR regime. In this chapter, the diversity definition in [67, 105] is used. The Kerdock codebook arranged as in (2.6) is easily verified to have full rank.

Theorem 5. *The proposed Kerdock codebook using at least one basis has full diversity order.*

Proof. The proof follows that found in [67] using the fact that the Kerdock codebook is of full rank since it is composed of unitary matrices. Thus, maximum diversity is achieved by the Kerdock codebook. □

2.7.3 Achievable Rate

The achievable rate of the system using a quantized codebook is an important indicator of the quality of the codebook [67, 87]. The ergodic achievable rate of the system with a unitary precoder is

$$\mathcal{R} = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathbf{I}_{N_s} + \frac{\mathcal{E}_s}{N_s N_o}\mathbf{F}^*(\mathbf{H})\mathbf{H}^*\mathbf{H}\mathbf{F}(\mathbf{H})\right)\right]. \tag{2.12}$$

where $\mathbb{E}_{\mathbf{H}}$ denotes the expectation with respect to $\mathbf{H}$ and $\mathbf{F}(\mathbf{H})$ is the selected precoder as a function of $\mathbf{H}$ according to the selection criteria (2.1) or (2.2). Perfect channel knowledge at the receiver and uncorrelated Gaussian signaling for each stream are assumed. This is the achievable rate upper bound when there are no channel estimation errors and feedback delay, but not the true capacity since the transmit covariance is not optimized for power allocation (i.e. water filling solution). For a fair comparison, the achievable rate with respect to perfect CSI at the transmitter case of an equal size Grassmannian, Fourier, and Kerdock codebook are compared in Fig. 2.2. For both the beamforming case (dashed line) and spatial multiplexing case (solid line), the Grassmannian, Fourier, and Kerdock codes have

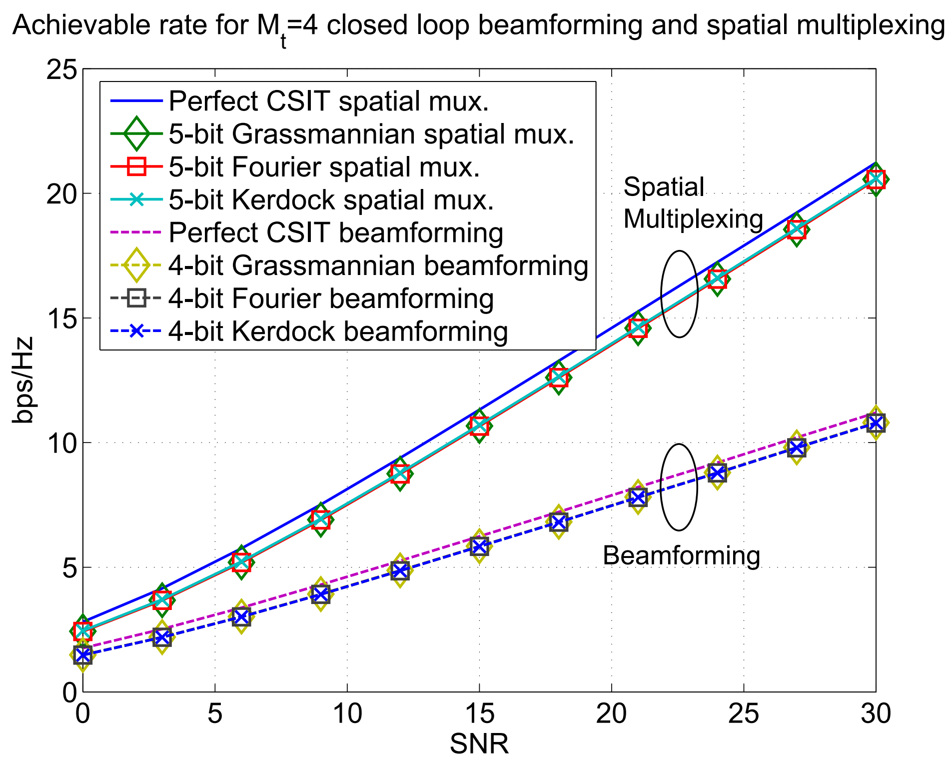

Figure 2.2: Achievable rate for $N_t = 4$ beamforming system and unitary precoded spatial multiplexing system using, perfect CSI at the transmitter, Grassmannian, Fourier, and Kerdock codebooks

the same achievable rates. Therefore there is no loss in achievable rate using the Kerdock codebook.

2.8 Numerical Results

In this section, simulation results for 1) vector symbol error rate (VSER) performance of limited feedback beamforming system, and 2) VSER performance of two stream unitary precoded spatial multiplexing system using the Grassman-

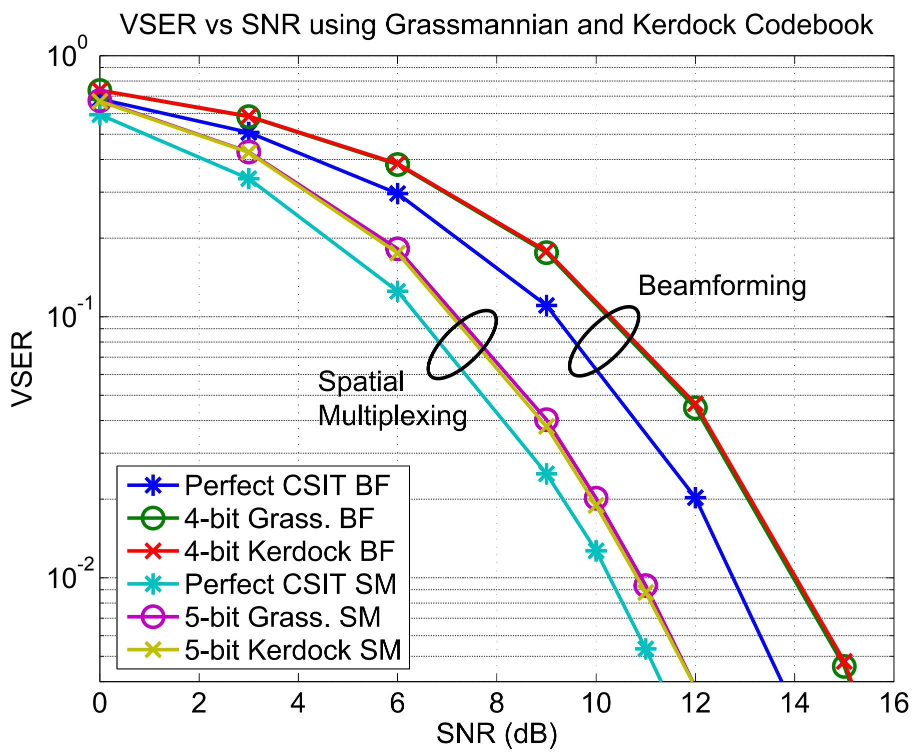

Figure 2.3: Vector symbol error rate performance of $N_t = 4$ beamforming and two stream spatial multiplexing systems for perfect CSI case, Grassmannian codebook and Kerdock codebook

nian and Kerdock codebooks are given. All simulations are performed for $N_t = 4$ assuming delay and error free feedback. No forward error correction is used.

The VSER performance of beamforming system using perfect CSI, Grassmannian codebook and Kerdock codebook are shown in Fig. 2.3. In all cases, 64-QAM is used for modulation. The case with perfect CSI at the transmitter provides the achievable lower bound of VSER. The proposed Kerdock codebook provides VSER performance closely matching the Grassmannian codebook. Similarly, the

VSER performance for two stream spatial multiplexing system using 5-bit codebooks are shown in Fig. 2.3. In all cases, 16-QAM modulation and a zero forcing receiver are used. Remarkably, the proposed Kerdock codebook with 30 codeword entries slightly outperforms the Grassmannian codebook with 32 entries.

To clearly see the performance difference among the codebook designs, Fig. 2.4 shows the SNR gap between the perfect CSI at the transmitter case and the limited feedback approach using Grassmannian, Fourier and Kerdock codebook at VSER $= 10^{-2}$ in two stream spatial multiplexing system. As expected, the Grassmannian codebook outperforms the Fourier codebook. The Kerdock codebook shows worse performance for the 3-bit codebook because only 8 of 30 possible codewords are used. As the codebook size is increased from 4 and 5 bits, however, the Kerdock codebook outperforms the Grassmannian codebook which is quite remarkable considering the fact that the codebook contains only quaternary alphabet. This observation indicates that Grassmannian codebook is not exactly optimal.

Overall, the results indicate that the proposed Kerdock codebook performs very close or better than previously known codebooks with additional benefit of 1) structured construction, 2) finite alphabet, 3) reduced search complexity, and 4) shared codebook between beamforming and spatial multiplexing.

2.9 Summary

In this chapter, a new avenue of codebook design, called the Kerdock codebook, for limited feedback unitary precoded MIMO systems was proposed. The Kerdock codebook is systematically generated with elements drawn from a qua-

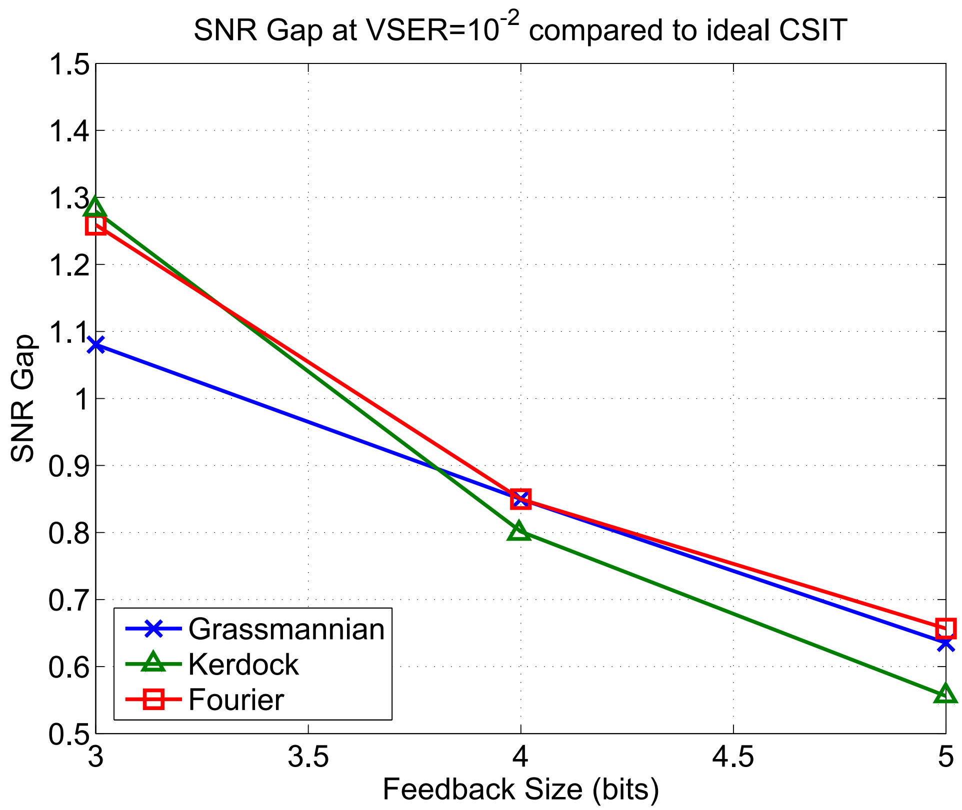

Figure 2.4: SNR gap between perfect CSI at the transmitter case and Grassmann, Fourier, and Kerdock codebook designs for 2-stream Spatial Multiplexing System at VSER = 10^{-2}

ternary alphabet resulting in reduced storage and search complexity. Furthermore, it was shown that the structure of the Kerdock codes can be used to derive spatial multiplexing codebooks from the beamforming codebook. Analysis and simulation results verified that the Kerdock codebooks provides favorable performance to the previously known codebooks. Limitations of this work are that the codebook can only be constructed for number of transmit antennas which are powers of two and the number of available bases are limited to $N_t + 1$. An open problem remains in constructing odd dimension codebook with quaternary alphabet entries. Our future work will consider effects of space or time correlated channels and extensions to multiuser scenarios. In particular, Kerdock codes are also applicable to a multiuser MIMO system using a unitary basis sets, known as PU^{2}RC [37, 91].[1]

2.10 Appendix

2.10.1 Proof of Property 3

Proof. Each $\mathbf{F}_k$ and $\mathbf{F}_l$ can be written as $\mathbf{F}_k = \mathbf{D}^p\mathbf{E}_k$ and $\mathbf{F}_l = \mathbf{D}^q\mathbf{E}_l$ where $\mathbf{E}_k$ and $\mathbf{E}_l$ are $N_t \times 2$ column selection matrices. Then $\mathbf{F}_k^*\mathbf{F}_l = \mathbf{E}_k^T\mathbf{D}^{q*}\mathbf{D}^p\mathbf{E}_l = \mathbf{E}_k^T\mathbf{D}^r\mathbf{E}_l$ where $r = (q-p)*$ when $q > p$ and $r = (p-q)$ when $p > q$. Due to the construction $\mathbf{D}^r$ is one of the member basis. The left and right multiplication of $\mathbf{D}^r$ by $\mathbf{E}_k^T$ and $\mathbf{E}_l$ selects a 2×2 sub-matrix of $\mathbf{D}^r$. Any member $\mathbf{D}^r$ has a structure such that any 2×2 sub-matrix selected this way always contains 1) all reals, 2) a pair of reals and a pair of imaginary, or 3) all imaginary, from the quaternary alphabet. It is easy to verify, by listing all possibilities, that the determinant of such 2×2 matrix can only take values 0 or $1/\sqrt{N_t}$. □

2.10.2 Proof of Property 4

Proof. The $\det(\mathbf{F}_k^*\mathbf{F}_l)$ is given by the determinant of a 3×3 sub-matrix of some basis $\mathbf{S}_n$. Recall that the adjoint of a square matrix $\mathbf{D}$, denoted adj$(\mathbf{D})$, is given

[1] After this work was submitted for [47], a similar codebook design with nested property for multiuser MIMO was proposed in [77]. This work differs in that two codebook design strategies are proposed in addition to the shared codebook structure for single user beamforming and spatial multiplexing cases with provable distance properties.

by adj$(\mathbf{D}) = \mathbf{D}^{-1} \cdot \det(\mathbf{D})$. Since $\mathbf{D}$ is unitary, $\mathbf{D}^{-1} = \mathbf{D}^*$ and $\det(\mathbf{D}) = \mathbf{I}$. So, adj$(\mathbf{D}) = \mathbf{D}^*$. Therefore, the adjoint matrix also has quaternary alphabet. The elements of adjoint matrix is the cofactors which are minors, or determinant of 3×3 sub-matrix, with appropriate signs. This shows that every determinant of 3×3 sub-matrix is in the set $\{\pm 1, \pm j\}$ with scaling $1/\sqrt{N_t} = 1/2$ and the result follows. $\square$

Chapter 3

Signal Processing on the Grassmann Manifold

3.1 Prior Work

Predictive vector quantization (PVQ) is a class of memory based coding techniques used in applications such as speech, image, and video processing [22, 29, 30, 55]. In PVQ, the error signal between the current observed vector and the predicted vector based on past observations is quantized. When the observed data to be encoded are correlated, usually in time or space, and a suitable prediction function is available, quantizing the error signal leads to lower distortion compared with memoryless vector quantization [22]. The effectiveness of PVQ rests on the correlation exhibited by the data, the prediction function, and the quantization technique employed. In speech coding, for example, small blocks of speech signal exhibits temporal correlation due to human speech production mechanisms [17]. Classical PVQ technique has been applied for signals in linear vector space where the usual difference, addition, and prediction are well understood. Unfortunately, when the signal to be encoded is in a special mathematical space such as the Grassmann manifold, the usual linear operations are not well defined and a new set of tools must be established. In particular, an extension of predictive coding for Grassmann manifold has not appeared in prior literature.

Motivated by applications in MIMO wireless communication, there has been research in analyzing [8], quantizing [7, 15, 74], and coding [12, 53, 104] on the Grassmann manifold driven in part by applications to commercial wireless systems [1, 41]. Prior work exists for designing suitable memoryless quantization codebooks such as Grassmannian line packing [79], vector quantization [88], Grassmannian frames [102], and Kerdock codebooks [47]. In practice, though, the wireless communication channel often exhibits temporal correlation due to mobility in the propagation channel [38, 40, 56, 75, 85, 98]. In [38, 40], modeling the feedback state transitions allow the net feedback rate to be reduced. Resolution, however, is fixed by the codebook size. To improve the quantization error, an adaptive codebook approach was proposed which can adapt to a given channel distribution [75]. A feedback overhead to retrain or synchronize the pre-computed codebooks may be needed when the channel distribution changes. Alternatively, a hierarchical codebook strategy uses two codebooks, coarse and fine, for layered feedback in temporally correlated channel [56]. Codeword describing the coarse encoding region is updated infrequently and a finer local codebook is used for frequent feedback. A more flexible approach is to use a progressive refinement strategy in which rotation and scaling are applied to structured codebook so as to provide high resolution feedback [34, 85]. Finally, a closely related result to this chapter is a complex Householder transform based predictive vector quantization technique for correlated normalized channel vectors in multiple-input single-output communication systems [62]. The current vector channel is decomposed into previous vector channel and weighted sum of orthogonal subspaces to represent the temporal variation. While the algorithm is presented in the form of predictive vector quantization, the actual operation is successive decomposition and projection using the complex Householder transform with unit delay as the predictor which was shown to be optimal for the specific application.

Feedback delay in practical systems, e.g., due to protocol overhead and channel estimation interval, may further aggravate CSIT accuracy in time-varying channels [38, 73, 99, 103]. The eigenmode and singular value coherence times were evaluated and the sensitivity of throughput in multiuser MIMO systems to feedback delay was illustrated in [73]. In [38], it was shown that feedback throughput gain decreases exponentially with increasing feedback delay, suggesting the importance of considering feedback delay. Indeed, delayed feedback characterized as an additional error in CSI has shown to degrade the sum rate performance in multiuser MIMO systems, especially for fast time-varying channels [99]. More recently, in [103], a mode switching strategy between single user and multiuser MIMO was proposed conditioned on SNR, normalized Doppler frequency and codebook size. They showed that the operating region for multiuser MIMO to be in low normalized Doppler region with large codebook size, suggesting the vulnerability of multiuser MIMO techniques in time-varying channels with feedback delays.

3.2 Contributions

In this chapter, a predictive coding algorithm is proposed for correlated data on the Grassmann manifold, called the Grassmannian predictive coding (GPC) algorithm. The proposed algorithm is motivated by the need for higher resolution quantization in limited feedback MIMO wireless systems. The GPC algorithm is derived using the intrinsic geometry and natural operations defined on the manifold. The main contributions of this chapter are as follows.

- *Grassmannian predictive coding algorithm*: A framework for predictive coding on the Grassmann manifold is proposed. The key idea of the proposed approach is in using *tangent vector* to establish the notion of an error signal and prediction on the manifold. The error tangent vector is decomposed into an error tangent direction and an error tangent magnitude and several special codebook designs are proposed. Furthermore, a prediction function is proposed using *parallel transport* corresponding to a one step prediction. Formulating higher order prediction functions remain as an open problem. The tangent and parallel transport on the Grassmann manifolds has been used in [18], but it has not been exploited to develop a predictive coding concept.

- *Grassmannian predictive coding for delayed feedback:* To address the limited feedback delay in MIMO systems, a new predictive coding architecture, called adaptive step size GPC algorithm, is proposed where the output of the predictor is used as the output of the decoder. Adaptive step size Grassmannian predictor is used at the encoder which feeds back the predicted step size to the decoder. The encoder thus use the observed CSI to compute the prediction error as well as the predicted step size. Both the prediction error and predicted step size are quantized and transmitted to the decoder via finite rate feedback channel. An efficient step size feedback strategy is proposed to maintain practical feedback rates. The mean squared chordal distance error performance of the proposed approach is compared with memoryless codebook approach and the GPC algorithm that illustrates significantly improved error performance.

- *Distortion bounds:* Based on a geometric interpretation of the proposed GPC algorithm, a simple model of the quantization region is obtained. Using metric volume computations on the Grassmann manifold [74], lower and upper bounds on the quantization error are derived. The obtained bounds are compared with distortion obtained in simulations. Furthermore, the distortion for the proposed GPC algorithm is shown to be lower than the lower bound of memoryless quantizer distortion for a given codebook size.

- *Application to limited feedback beamforming systems*: The proposed GPC algorithms are applied to single user limited feedback beamforming system in a temporally correlated wireless channel with and without feedback delay. The proposed GPC algorithm without feedback delay is compared with memoryless Grassmannian codebook and Householder transform based PVQ with limited feedback. It is shown that the proposed GPC algorithm provides better symbol error and achievable rates than prior methods under the same feedback rate constraints. The adaptive step size Grassmannian predictive coding algorithm in the presence of feedback delay was shown to provide improved symbol error rate performance over memoryless codebook approach.

- *Application to limited feedback multiuser MIMO systems*: The proposed GPC algorithms are applied for limited feedback zero-forcing multiuser MIMO systems with multiple transmit antennas and a single receive antenna at each mobile terminal [51]. To isolate the effect of the limited feedback, it is assumed that the users are scheduled a priori. The proposed GPC algorithm without feedback delay is shown to provide substantial sum rate improvement over memoryless random codebook technique with same feedback rate [51]. The sum rate improvement, however, depends on the channel correlation. When the channel is highly correlated, the proposed GPC algorithm is shown to provide sum rate close to the system with perfect CSI at the transmitter, i.e., infinite feedback. The adaptive step size GPC algorithm was shown to provide substantial sum rate improvement even in the presence of feedback delay. Thanks to the GPC structure, sum rate improvements as a function of temporal correlation are shown in the presence of feedback delay.

3.3 Grassmann Manifold: Preliminaries

Geometric and linear algebraic properties of the Grassmann manifold will be fundamental in derivation of the proposed algorithm. In this section, an overview of the Grassmann manifold is given and necessary tools such as the distance metric, tangent, mapping from tangent onto the manifold, parallel transport, and prediction are derived.

Let $\mathcal{U}_n = \{\mathbf{X} \in \mathbb{C}^{n \times n} : \mathbf{X}^*\mathbf{X} = \mathbf{I}_n\}$ be the unitary group formed by $n \times n$ unitary matrices. The Stiefel manifold is the space of unitary $n \times p$ matrices $(p < n)$

defined as $\mathcal{V}_{n,p} = \{\mathbf{X} \in \mathbb{C}^{n \times p} : \mathbf{X}^*\mathbf{X} = \mathbf{I}_p\}$. Since $p < n$ and there are $n - p$ free dimensions with respect to the n-dimensional space, it may be equivalently identified as the quotient group $\mathcal{U}_n/\mathcal{U}_{n-p}$. The Grassmann manifold, $\mathcal{G}_{n,p}$, is the set of subspaces spanned by the columns of $\mathcal{V}_{n,p}$. It may also be identified as a quotient space of the Stiefel manifold, $\mathcal{V}_{n,p}/\mathcal{U}_p$, or as a quotient space of the unitary group, $\mathcal{U}_n/(\mathcal{U}_{n-p} \times \mathcal{U}_p)$. A point $\mathbf{X} \in \mathcal{G}_{n,p}$ may be considered as an equivalence class, i.e., $[\mathbf{X}] := \{\mathbf{X}\mathbf{U}_p : \mathbf{U}_p \in \mathcal{U}_p\}$. For notational brevity, $\mathbf{X} \in \mathcal{G}_{n,p}$ is used to mean the equivalence class of matrices whose columns span the same p-dimensional subspace. For numerical computation, $\mathbf{X} \in \mathcal{G}_{n,p}$ is interpreted to be one representative of the equivalence class. Also, the Grassmann manifold is a smooth topological manifold with locally Euclidean property and smooth tangent space structure, both of which will be essential in the derivation of the proposed algorithm [61]. In this chapter, the Grassmann manifold $\mathcal{G}_{n,1}$ is considered; the general case for $p > 1$ is treated in Chapter 4.

First, the notion of distance between points on the Grassmann manifold plays a basic role as a metric for error or similarity between points. Let the inner product of $\mathbf{x}[1],\ \mathbf{x}[2] \in \mathcal{G}_{n,1}$ be denoted by $\rho = \mathbf{x}^*[1]\mathbf{x}[2]$. Let $\theta = \cos^{-1}(|\rho|)$ be the subspace angle between $\mathbf{x}[1]$ and $\mathbf{x}[2]$ [25]. The chordal distance metric for $\mathcal{G}_{n,1}$ is given by [18, 67]

$$\begin{aligned} d(\mathbf{x}[1], \mathbf{x}[2]) &= \sqrt{1 - |\rho|^2} \\ &= |\sin\theta| \end{aligned} \tag{3.1}$$

which is same as (2.3) but with the inner product inside the squared absolute value denoted by ρ. For notational brevity, the notation d is used without the arguments when there is no confusion. Unlike the arc length, given by $|\theta|$, the chordal distance is differentiable everywhere and provides a close approximation of the arc length when the points are close [13]. The advantages of using chordal distance for the packing problem on Grassmann manifold are discussed in [8]. For the codebook construction problem using the generalized Lloyd algorithm in limited feedback MIMO systems, it was shown that the Lloyd algorithm is feasible only when the chordal distance is chosen as the distance measure [106]. Based on these observations, the chordal distance is used as the distance measure on the Grassmann manifold throughout this chapter.

Using the chordal distance metric, the correlation of two sequences

$$\{\mathbf{x}[k]\}_{k\in\mathbb{N}}, \{\mathbf{y}_i\}_{i\in\mathbb{N}} \in \mathcal{G}_{n,1}$$

is defined by $\zeta_{\mathbf{x},\mathbf{y}}(n) = \mathbb{E}_k[d(\mathbf{x}[k], \mathbf{y}[k+n])]$ which can be interpreted as the mean chordal distance between two sequences on the Grassmann manifold. Two applications where such sequences arise are described in Section 3.6.3 and 3.6.4.

Based on the smooth manifold structure of the Grassmann manifold, it is possible to relate two points $\mathbf{x}[k], \mathbf{x}[k+1] \in \mathcal{G}_{n,1}$ by considering the tangent vector emanating from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$. The tangent has been used successfully in the development of Newton and conjugate gradient algorithms with orthogonality constraints [18, 68–70]. The tangent method is utilized for its computational benefits and geometric insight to the problem.

Theorem 6 (Tangent)**.** *If* $\mathbf{x}[k],\ \mathbf{x}[k+1] \in \mathcal{G}_{n,1}$*, then the tangent vector emanating* *from* $\mathbf{x}[k]$ *to* $\mathbf{x}[k+1]$ *is*

$$\mathbf{e} \;=\; \tan^{-1}\left(\frac{d}{|\rho|}\right)\frac{\mathbf{x}[k+1]/\rho - \mathbf{x}[k]}{\|\mathbf{x}[k+1]/\rho - \mathbf{x}[k]\|_2} \tag{3.2}$$

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

Proof. See Appendix 3.8.1. □

Theorem 6 provides a compact formula for the tangent vector on $\mathcal{G}_{n,1}$. For notational brevity, the tangent operation is denoted by $\mathbf{e} = L(\mathbf{x}[k], \mathbf{x}[k+1])$. It can be considered as a length preserving unwrapping of the arc between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$ onto the tangent space at $\mathbf{x}[k]$. Furthermore, it is conveniently expressed as the product of magnitude component and the normalized directional component. This decomposition will be exploited for quantization. The tangent vector describes the shortest distance path along the arc from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$, called the *geodesic* [18]. The geodesic can be parameterized by one parameter $t \in [0, 1]$ using the tangent vector as the next theorem shows.

Theorem 7 (Geodesic)**.** *If* $\mathbf{x}[k],\ \mathbf{x}[k+1] \in \mathcal{G}_{n,1}$ *and* $\mathbf{e}$ *is the tangent vector emanating from* $\mathbf{x}[k]$ *to* $\mathbf{x}[k+1]$*, then the geodesic path between* $\mathbf{x}[k]$ *and* $\mathbf{x}[k+1]$ *is*

$$G(\mathbf{x}[k], \mathbf{e}, t) \;=\; \mathbf{x}[k]\cos(\|\mathbf{e}\|_2 t) + \vec{\mathbf{e}}\sin(\|\mathbf{e}\|_2 t) \tag{3.3}$$

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

Proof. See Appendix 3.8.2. □

Theorem 7 provides a convenient formula to map a tangent vector back to the Grassmann manifold. Unfortunately, the geodesic path is only defined between $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$. In order to introduce the notion of prediction, a tangent vector with respect to $\mathbf{x}[k+1]$ such that it extends the geodesic path from $\mathbf{x}[k]$ and $\mathbf{x}[k+1]$ is needed. Translation of the tangent vector is accomplished by the *parallel transport*.

Theorem 8 (Parallel Transport)**.** *Let* $\mathbf{x}[k],\ \mathbf{x}[k+1] \in \mathcal{G}_{n,1}$ *and* $\mathbf{e}$ *be the tangent vector emanating from* $\mathbf{x}[k]$ *to* $\mathbf{x}[k+1]$*. Then, the parallel transported tangent vector emanating from* $\mathbf{x}[k+1]$ *along the geodesic direction* $\mathbf{e}$ *is*

$$\hat{\mathbf{e}} = \tan^{-1}\left(\frac{d}{|\rho|}\right)\frac{\mathbf{x}[k+1]\rho^* - \mathbf{x}[k]}{d}. \tag{3.4}$$

Proof. See Appendix 3.8.3. □

Theorem 8 provides a convenient expression for transporting the *base* of the tangent vector from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$. It can be interpreted as transforming the tangent vector onto another tangent space connected by the geodesic. The next theorem shows the general prediction procedure combining the parallel transport and geodesic formula.

Proposition 9 (General Grassmannian Prediction)**.** *Let* $\mathbf{x}[k-1], \mathbf{x}[k] \in \mathcal{G}_{N,1}$ *and let* $\hat{\mathbf{e}}$ *be the parallel transported tangent vector emanating from* $\mathbf{x}[k]$*. The predicted* *vector* $\tilde{\mathbf{x}}[k+1] \in \mathcal{G}_{N,1}$ *extending the geodesic direction from* $\mathbf{x}[k-1]$ *to* $\mathbf{x}[k]$ *is*

$$\begin{aligned} \tilde{\mathbf{x}}[k+1] &= \mathbf{x}[k]\cos(\|\hat{\mathbf{e}}\|t) \\ &\quad + \frac{\mathbf{x}[k]\rho^* - \mathbf{x}[k-1]}{d}\sin(\|\hat{\mathbf{e}}\|t) \\ &:= P_G(\mathbf{x}[k-1], \mathbf{x}[k], t) \end{aligned} \tag{3.5}$$

where t *is the step size parameter and*

$$\|\hat{\mathbf{e}}\| = \tan^{-1}\left(\frac{d}{|\rho|}\right) \tag{3.6}$$

with $\rho = \mathbf{x}^*[k-1]\mathbf{x}[k]$ *and* $d = \sqrt{1-|\rho|^2}$.

Proof. See Appendix 3.8.4. □

Proposition 9 provides a compact form to perform prediction as a function of two previous points in $\mathcal{G}_{N,1}$ with $t \in [0,1]$ as the step size control parameter. The parameter t can be used to control how far in the parallel transported tangent direction to move. In particular, when a full step, i.e., $t = 1$, is taken, a remarkable simplification is obtained.

Theorem 10. *Let* $\mathbf{x}[k],\ \mathbf{x}[k-1] \in \mathcal{G}_{n,1}$. *The one step predicted vector* $\tilde{\mathbf{x}} \in \mathcal{G}_{n,1}$ *along the geodesic direction from* $\mathbf{x}[k]$ *to* $\mathbf{x}[k+1]$ *is*

$$\tilde{\mathbf{x}}[k+1] = |\rho|\mathbf{x}[k] + \rho^*\mathbf{x}[k] - \mathbf{x}[k-1] \tag{3.7}$$

such that $d(\mathbf{x}[k], \tilde{\mathbf{x}}[k+1]) = d(\mathbf{x}[k-1], \mathbf{x}[k])$.

Proof. See Appendix 3.8.5. □

The surprising outcome of Theorem 10 is that the predicted vector $\tilde{\mathbf{x}}[k+1]$ can be easily computed by the knowledge of $\mathbf{x}[k]$ and $\mathbf{x}[k-1]$ using linear operations. Indeed, if $\alpha, \beta \in \mathbb{C}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{G}_{N,1}$, then

$$(\alpha\mathbf{x} + \beta\mathbf{y})^*(\alpha\mathbf{x} + \beta\mathbf{y}) = |\alpha|^2 + 2\Re(\alpha^*\beta\mathbf{x}^*\mathbf{y}) + |\beta|^2$$

and setting this equal to 1, we may verify that $\alpha = |\mathbf{x}^*\mathbf{y}| + \mathbf{y}^*\mathbf{x}$ and $\beta = -1$ is a solution. Therefore, (4.17) always results in a unit norm vector despite the simple form. This greatly simplifies the computation required in the formulation of the GPC algorithm in section 3.4.

Taking a full step simplifies the calculation of the predictor but there is no reason a full step must be taken. Thus we consider optimizing the step size t in Proposition 9 to minimize the mean squared chordal distance error between the predicted vector and the observed vector. The mean squared chordal distance error between the predicted vector and the observed vector is

$$\begin{aligned} & \mathbb{E}[d^2(\tilde{\mathbf{x}}[k+1], \mathbf{x}[k+1])] \\ = \; & \mathbb{E}[1 - |\tilde{\mathbf{x}}^*[k+1]\mathbf{x}[k+1]|^2] \\ = \; & \mathbb{E}[1 - |P_G^*(\mathbf{x}[k-1], \mathbf{x}[k], t)\mathbf{x}[k+1]|^2]. \end{aligned}$$

The optimal step size t_{opt} is given by

$$\begin{aligned} t_{\text{opt}} &= \arg\min_{t\in[0,1]} \mathbb{E}[1 - |P_G^*(\mathbf{x}[k-1], \mathbf{x}[k], t)\mathbf{x}[k+1]|^2] \\ &= \arg\max_{t\in[0,1]} \mathbb{E}[|P_G^*(\mathbf{x}[k-1], \mathbf{x}[k], t)\mathbf{x}[k+1]|^2]. \end{aligned} \tag{3.8}$$

Unfortunately, (3.8) does not readily yield a closed form solution. Thus we approximate the optimal step size by computing the average of the instantaneous step size that minimizes the chordal distance error. The observed vector $\mathbf{x}[k+1]$ can be expressed with respect to $\mathbf{x}[k]$ using the tangent vector $\mathbf{e}[k+1]$ emanating from $\mathbf{x}[k]$ to $\mathbf{x}[k+1]$ as

$$\mathbf{x}[k+1] = \mathbf{x}[k]\cos(\|\mathbf{e}[k+1]\|) + \frac{\mathbf{e}[k+1]}{\|\mathbf{e}[k+1]\|}\sin(\|\mathbf{e}[k+1]\|). \tag{3.9}$$

Using (3.6) and (3.9), the term inside the absolute value in (3.8) becomes

$$\begin{aligned}
& \tilde{\mathbf{x}}^*[k+1]\mathbf{x}[k+1] \\
= \; & \cos(\|\hat{\mathbf{e}}\|t)\cos(\|\mathbf{e}[k+1]\|) \\
+ \; & \frac{\hat{\mathbf{e}}^*\mathbf{e}[k+1]}{\|\hat{\mathbf{e}}\|\|\mathbf{e}[k+1]\|}\sin(\|\hat{\mathbf{e}}\|t)\sin(\|\mathbf{e}[k+1]\|)
\end{aligned}$$

where we have used the fact that $\mathbf{x}[k]\perp\hat{\mathbf{e}}$ and $\mathbf{x}[k]\perp\mathbf{e}[k+1]$ since both $\hat{\mathbf{e}}$ and $\mathbf{e}[k+1]$ lies in the tangent space at $\mathbf{x}[k]$. Then, the objective is to maximize

$$\begin{aligned}
& |\tilde{\mathbf{x}}^*[k+1]\mathbf{x}[k+1]|^2 \\
= \; & \cos^2(\|\hat{\mathbf{e}}\|t)\cos^2(\|\mathbf{e}[k+1]\|) \\
+ \; & \frac{|\hat{\mathbf{e}}^*\mathbf{e}[k+1]|^2}{\|\hat{\mathbf{e}}\|^2\|\mathbf{e}[k+1]\|^2}\sin^2(\|\hat{\mathbf{e}}\|t)\sin^2(\|\mathbf{e}[k+1]\|) \\
+ \; & 2\frac{\Re(\hat{\mathbf{e}}^*\mathbf{e}[k+1])}{\|\hat{\mathbf{e}}\|\|\mathbf{e}[k+1]\|}\cos(\|\mathbf{e}[k+1]\|t)\sin(\|\mathbf{e}[k+1]\|t)\times \\
& \cos(\|\mathbf{e}[k+1]\|)\sin(\|\mathbf{e}[k+1]\|) \\
= \; & J_1(\hat{\mathbf{e}}, \mathbf{e}[k+1]) + J_2(\hat{\mathbf{e}}, \mathbf{e}[k+1], t)
\end{aligned}$$

where $J_1(\hat{\mathbf{e}}, \mathbf{e}[k+1]) \geq 0$ is a positive quantity independent of t and

$$\begin{aligned}
& J_2(\hat{\mathbf{e}}, \mathbf{e}[k+1], t) \\
= & \big(\cos^2(\|\mathbf{e}[k+1]\|) \\
- & \frac{|\hat{\mathbf{e}}^* \mathbf{e}[k+1]|^2}{\|\hat{\mathbf{e}}\|^2 \|\mathbf{e}[k+1]\|^2} \sin^2(\|\mathbf{e}[k+1]\|)\bigg) \cos(2\|\hat{\mathbf{e}}\|t) \\
+ & \frac{\Re(\hat{\mathbf{e}}^* \mathbf{e}[k+1])}{\|\hat{\mathbf{e}}\| \|\mathbf{e}[k+1]\|} \sin(2\|\mathbf{e}[k+1]\|) \sin(2\|\hat{\mathbf{e}}\|t).
\end{aligned} \tag{3.10}$$

Thus, the instantaneous optimal step size is obtained by maximizing $J_2(\hat{\mathbf{e}}, \mathbf{e}[k+1], t)$ given by (3.10). Unfortunately, a unique solution t that maximizes (3.10) is not available due to the periodicity of the trigonometric functions. To obtain a closed form solution, we recognize that the range of $\|\hat{\mathbf{e}}\|$ is typically small for our applications, e.g., $\|\hat{\mathbf{e}}\| \leq 0.1$. For small $\|\hat{\mathbf{e}}\|$, the cost function is well approximated by the first order Taylor series expansion as a function of t. Therefore, we may write

$$\begin{aligned}
& J_2(\hat{\mathbf{e}}, \mathbf{e}[k+1], t) \\
= & \cos^2(\|\mathbf{e}[k+1]\|) - \frac{|\hat{\mathbf{e}}^* \mathbf{e}[k+1]|^2}{\|\hat{\mathbf{e}}\|^2 \|\mathbf{e}[k+1]\|^2} \sin^2(\|\mathbf{e}[k+1]\|) \\
+ & 2\frac{\Re(\hat{\mathbf{e}}^* \mathbf{e}[k+1])}{\|\hat{\mathbf{e}}\| \|\mathbf{e}[k+1]\|} \sin(2\|\mathbf{e}[k+1]\|) \|\hat{\mathbf{e}}\| t \\
- & 2\left(\cos^2(\|\mathbf{e}[k+1]\|) - \frac{|\hat{\mathbf{e}}^* \mathbf{e}[k+1]|^2}{\|\hat{\mathbf{e}}\|^2 \|\mathbf{e}[k+1]\|^2} \sin^2(\|\mathbf{e}[k+1]\|)\right) \\
\times & \|\hat{\mathbf{e}}\|^2 t^2
\end{aligned} \tag{3.11}$$

which is a quadratic function of t. Since the coefficient for t^2 term is always negative for small enough values of $\|\mathbf{e}[k+1]\|$, the optimal step size t is obtained by taking the derivative of (3.11) with respect to t and setting it equal to zero. Therefore, the optimal step size is given by

$$t_{\text{opt}} = \frac{\Re(\hat{\mathbf{e}}^* \mathbf{e}[k+1]) \sin(2\|\mathbf{e}[k+1]\|)}{2\|\mathbf{e}[k+1]\|\|\hat{\mathbf{e}}\|^2 \cos^2(\|\mathbf{e}[k+1]\|) - \frac{|\hat{\mathbf{e}}^* \mathbf{e}[k+1]|^2}{\|\hat{\mathbf{e}}\|\|\mathbf{e}[k+1]\|} \sin^2(\|\mathbf{e}[k+1]\|)} \tag{3.12}$$

Let $t_{\text{opt}}[k+1]$ denote the optimal step size to predict $\mathbf{x}[k+1]$. The step size optimization criterion is stated as follows.

Step Size Optimization Criterion: Pick t such that

$$t_{\text{opt}}[k+1] = \arg\max_{t \in [0,1]} J_2(\hat{\mathbf{e}}, \mathbf{e}[k+1], t). \tag{3.13}$$

where the closed form solution to $t_{\text{opt}}[k+1]$ is given by (3.12).

In practice, the optimal step size will depend on the statistics of the channel which may vary in time, e.g. do to mobiles moving at different velocities. Thus it is desirable to estimate and update the step size according to the observed channel as opposed to using a fixed precomputed value. In this paper, we propose a least mean square (LMS) based step size predictor which computes the predicted step size based on the past observations of $\mathbf{x}$ and instantaneous t_{opt}. The LMS algorithm was chosen for its computational simplicity as it is undesirable to introduce further computational delay by employing more advanced techniques such as recursive least squares.

The main idea is to compute $t_{\text{opt}}[k+1]$ after observing $\mathbf{x}[k+1]$ and then predict the next step size $\tilde{t}[k+2]$ based on the past knowledge of t_{opt}. Let $\tilde{t}[\cdot]$ denote the predicted step size. Then, the step size prediction error is

$$e_t[k+1] = t_{\text{opt}}[k+1] - \tilde{t}[k+1]. \tag{3.14}$$

We use the mean squared prediction error $\mathbb{E}\left[|e_t|^2\right]$ criterion and design a linear predictor to minimize the mean squared prediction error. Since $t_{\text{opt}}[k]$ is a scalar process, an M-th order linear predictor

$$\tilde{t}[k+1] = \sum_{n=0}^{M-1} a_n t_{\text{opt}}[k-n] \tag{3.15}$$

with filter coefficients $\{a_n\}_{n=0}^{M-1}$ is used. If $\mathbf{R}_t$ is an $M \times M$ autocorrelation matrix for $t_{\text{opt}}[k]$, $\mathbf{a} = \begin{bmatrix} a_0 & \cdots a_{M-1} \end{bmatrix}^T$ is a vector of filter coefficients, and $\mathbf{r}$ is an $M \times 1$ vector of cross correlation between the desired optimized step size t_{opt} and the predicted step size $\tilde{t}$, the Wiener-Hopf equation is [31]

$$\mathbf{R}_t \mathbf{a} = \mathbf{r} \tag{3.16}$$

and the optimal filter coefficients can be found by computing $\mathbf{R}_t^{-1}\mathbf{r}$. Since the sample history of $t_{\text{opt}}[k]$ may not be available, we use the least mean square (LMS) algorithm to adapt the filter coefficients $\{a_n\}_{n=0}^{M-1}$ based on instantaneous correlation estimates [31]. The pseudo code is shown in Algorithm 7.

Algorithm 1 Adaptive Step Size Prediction

Input: $\mathbf{t}_{\text{opt}}[k+1] = \begin{bmatrix} t_{\text{opt}}[k] & \cdots & t_{\text{opt}}[k-M-1] \end{bmatrix}^T$

1: Initialize $\mathbf{a}[1] = \begin{bmatrix} a_0[1] & \cdots & a_{M-1}[1] \end{bmatrix}^T$
2: **for all** k=1,2,... **do**
3: $\quad \tilde{t}[k+1] = \mathbf{a}^T[k]\mathbf{t}_{\text{opt}}$
4: $\quad e_t[k+1] = t_{\text{opt}}[k+1] - \tilde{t}[k+1]$
5: $\quad \mathbf{a}[k+2] = \mathbf{a}[k+1] + \mu e_t[k+1]\mathbf{t}_{\text{opt}}[k+1]$
6: **end for**

Output: $\tilde{t}[k+1]$

Therefore, we arrive at an adaptive step size Grassmannian predictor which consists of step size optimization, LMS algorithm to predict the step size, and the

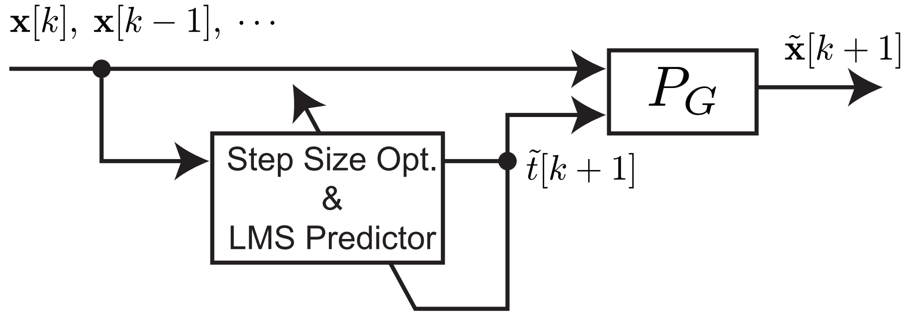

Figure 3.1: Block diagram of proposed fully adaptive Grassmannian predictor with step size optimization, least mean square (LMS) step size predictor, and the general Grassmannian prediction (Proposition 9).

general Grassmannian prediction in Proposition 9. A high level block diagram for the fully adaptive Grassmannian predictor is shown in Fig. 4.1.

3.4 Grassmannian Predictive Coding

In this section, the proposed GPC algorithm is described. First, a general overview of the algorithm is provided. Second, codebook designs for encoding the error tangent vector is described. Finally, methods for initialization are described.

3.4.1 GPC Algorithm

Let $\{\mathbf{x}[k]\}_{k\in\mathbb{N}} \in \mathcal{G}_{n,1}$ be the correlated input sequence with time index k. The general operation of the proposed GPC algorithm closely follows that of the conventional predictive vector quantization technique [22]. Linear operations such as difference, quantization, addition, and prediction are replaced by equivalent operators on Grassmann manifold derived in Section 3.3. The main idea of predictive coding is to quantize the error $\mathbf{e}[k]$ between the predicted vector $\tilde{\mathbf{x}}[k]$ and the current observed vector $\mathbf{x}[k]$. This is in contrast to quantizing $\mathbf{x}[k]$ directly in conventional one-shot approach. Then, the quantized error is applied to predicted vector to construct the estimate $\hat{\mathbf{x}}[k]$ of the current observed vector. The current and previous estimated vectors, $\hat{\mathbf{x}}[k]$ and $\hat{\mathbf{x}}_{k-1}$, are used to compute the predict vector $\tilde{\mathbf{x}}[k+1]$. Since both the encoder and decoder uses estimated vectors for prediction, they both obtain the same predicted vectors.

Algorithm 2 GPC encoder algorithm

Input: $\mathbf{x}[k]$
1: Initialize $\tilde{\mathbf{x}}[1]$ and $\hat{\mathbf{x}}[0]$
2: **for all** k=1,2,... **do**
3: $\quad \mathbf{e}[k] = L(\tilde{\mathbf{x}}[k], \mathbf{x}[k])$
4: $\quad (\ell[k], i[k]) = Q(\mathbf{b}[k])$
5: $\quad \hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1)$
6: $\quad \tilde{\mathbf{x}}[k+1] = P(\hat{\mathbf{x}}[k-1], \hat{\mathbf{x}}[k])$
7: **end for**
Output: $\ell[k], i[k]$

Fig. 3.2 illustrates the proposed GPC encoder; the pseudo code is provided in Algorithm 2. At time k, an error tangent vector is computed from the predicted vector $\tilde{\mathbf{x}}[k]$ to the current observed vector $\mathbf{x}[k]$. Using (4.12), the error tangent vector emanating from $\tilde{\mathbf{x}}[k]$ to $\mathbf{x}[k]$ is computed as

$$\mathbf{e}[k] = \tan^{-1}\left(\frac{d}{|\rho|}\right)\frac{\mathbf{x}[k]/\rho - \tilde{\mathbf{x}}[k]}{\|\tilde{\mathbf{x}}[k]/\rho - \mathbf{x}[k]\|_2} \tag{3.17}$$

where $\rho = \tilde{\mathbf{x}}^*[k]\mathbf{x}[k]$ and $d = \sqrt{1-|\rho|^2}$. Quantization is performed in two steps. Let $\mathcal{C}_m = \{c_{m,\ell}\}_{\ell=1}^{N_m}$ with N_m codewords denote the codebook of error tangent magnitudes in nonnegative reals and $\mathcal{C}_d = \{\mathbf{c}_{d,i}\}_{i=1}^{N_d}$ with N_d codewords denote

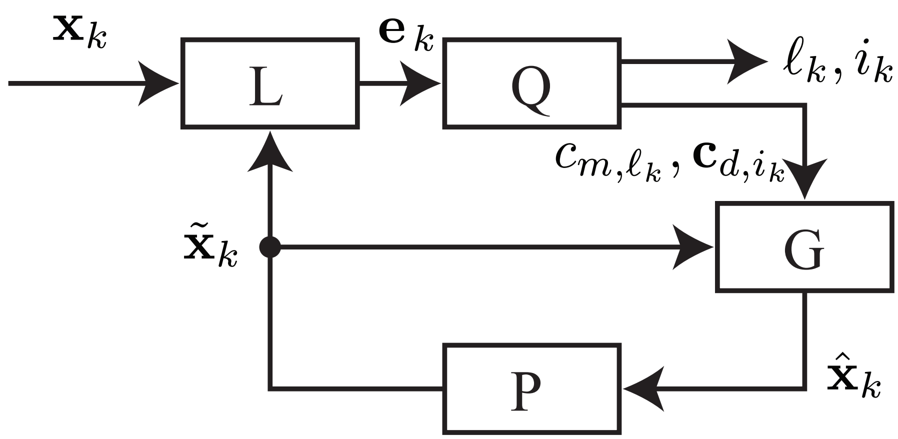

Figure 3.2: Block diagram of predictive encoder on the Grassmann manifold.

the codebook of unit norm error tangent directions in $\mathbb{C}^n$. First, the error tangent magnitude is quantized according to

$$\ell[k] = \underset{i \in \{1,2,\ldots,N_m\}}{\arg\min} \left| \|\mathbf{e}[k]\|_2 - c_{m,i} \right| \tag{3.18}$$

where $\ell[k]$ is the index of the selected error tangent magnitude codeword in $\mathcal{C}_m$. This sets the radius of the quantized error tangent vector. Next, the error tangent direction codeword that minimizes the chordal distance error between the estimated and observed vector is computed as

$$i[k] = \underset{i \in \{1,2,\ldots,N_d\}}{\arg\min} d(G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i}, 1), \mathbf{x}[k]) \tag{3.19}$$

where $G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i}, 1)$ computes the estimated vector according to the selected error tangent magnitude codeword $c_{m,\ell[k]}$ and error tangent direction codeword $\mathbf{c}_{d,i}$ from the codebook $\mathcal{C}_d$. The unit norm tangent direction codeword that yields the

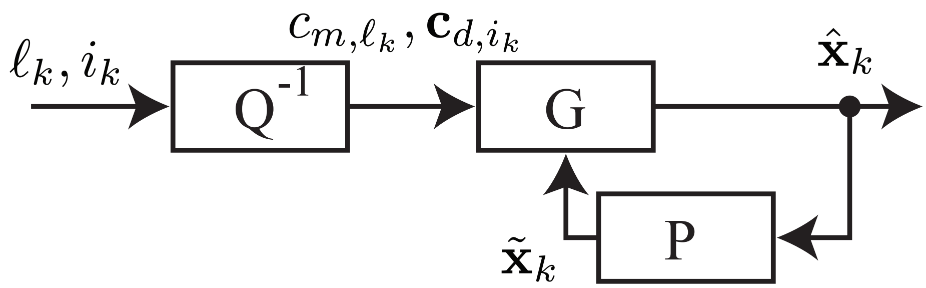

Figure 3.3: Block diagram of predictive decoder on the Grassmann manifold.

estimate with the minimum chordal distance is selected. For notational brevity, the two step quantization is denoted by $Q : \mathbb{C}^n \rightarrow \mathbb{N} \times \mathbb{N}$ which takes the error tangent vector and outputs two codeword indices, i.e., $(\ell[k], i[k]) = Q(\mathbf{e}[k])$. The indices of the direction and magnitude codewords are transmitted to the decoder via finite rate communication channel. Continuing at the encoder, the estimated vector is

$$\hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1). \tag{3.20}$$

Finally, the prediction, Theorem 10, is performed using two previous estimates

$$\tilde{\mathbf{x}}[k+1] = |\rho|\hat{\mathbf{x}}[k] + \rho^*\hat{\mathbf{x}}[k] - \hat{\mathbf{x}}_{k-1} \tag{3.21}$$

where $\rho = \hat{\mathbf{x}}^*[k]\hat{\mathbf{x}}_{k-1}$. For notational brevity, the prediction operation is denoted by a map $P : \mathcal{G}_{n,1} \times \mathcal{G}_{n,1} \rightarrow \mathcal{G}_{n,1}$ which takes current and previous estimate vectors and outputs the predicted vector, i.e., $\tilde{\mathbf{x}}[k+1] = P(\hat{\mathbf{x}}_{k-1}, \hat{\mathbf{x}}[k])$. The encoding procedure is repeated for each time $k+1, k+2, \ldots$.

Fig. 3.3 illustrates the proposed GPC decoder; the pseudo code is shown in Algorithm 3. The same error tangent magnitude and direction codebooks as the encoder are assumed to be available. The received indices are decoded in Q^{-1} to

Algorithm 3 GPC decoder algorithm

Input: $\ell[k], i[k]$

1: Initialize $\tilde{\mathbf{x}}[1]$ and $\hat{\mathbf{x}}[0]$
2: **for all** k=1,2,… **do**
3: $\quad c_{m,\ell[k]} \cdot \mathbf{c}_{d,i[k]} = Q^{-1}(\ell[k], i[k])$
4: $\quad \hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1)$
5: $\quad \tilde{\mathbf{x}}[k+1] = P(\hat{\mathbf{x}}[k-1], \hat{\mathbf{x}}[k])$
6: **end for**

Output: $\hat{\mathbf{x}}[k]$

recover $\mathbf{c}_{d,i[k]}$ and $c_{m,\ell[k]}$. The predicted vector $\tilde{\mathbf{x}}[k]$ is mapped to the estimated vector using the codewords as in (3.28). The decoder output is $\hat{\mathbf{x}}[k]$. Similarly to the encoder, the prediction is performed using (4.17) to obtain $\tilde{\mathbf{x}}[k+1]$ for the next time period. Note that for the first iteration of the decoder, the knowledge of $\tilde{\mathbf{x}}[k]$, or equivalently $\hat{\mathbf{x}}[k-1]$ and $\hat{\mathbf{x}}[k-2]$, is needed. Synchronizing the initial vectors with the encoder is important because if $\tilde{\mathbf{x}}[k]$ is different from the encoder, the received error tangent direction and magnitude no longer represents the error tangent vector. In Section 3.4.3, an efficient strategy is provided for initialization over finite rate communication channel. With appropriate initialization, symmetric operation at the encoder and decoder yields the same predicted vector $\tilde{\mathbf{x}}[k]$ for each time k. Therefore, initialization only needs to be performed once.

3.4.2 Codebook Design

The codebook design is performed exploiting the structure of the tangent space. The error tangent vector $\mathbf{e}[k]$ at time k is written as the product of the unit norm tangent direction $\vec{\mathbf{e}}[k]$ and the tangent magnitude $\|\mathbf{e}[k]\|_2$. Separating the magnitude and direction for quantization is known to provide better dynamic range coverage [22].

The tangent magnitude $\|\mathbf{e}[k]\|_2$ is dependent on the distance between the predicted vector and the observed vector, which in turn is dependent on the rate of change of the input vectors. A simple uniform quantization or more sophisticated scalar quantization may be used [22]. In this chapter, the uniform quantization approach is used and more advanced quantization techniques will be considered in future work. The tangent direction codebook can be designed using Lloyd algorithm [22]; the pseudo code is shown in Algorithm 4. An iterative codebook modification, i.e., Lloyd iteration, is used on a training set consisting of error tangent direction vector samples to improve the average distortion of the codebook. Let $\tau = \{\mathbf{x}_\ell\}_{\ell=1}^{M}$ be the given training set of error tangent directions and $\mathcal{C}_d^{(m)}$ denote the codebook at th m-th iteration. The initial codebook $\mathcal{C}_d^{(1)}$ may be obtained by generating random unit norm vectors. The Lloyd iteration proceeds by obtaining a partition or cluster using the nearest neighbor condition (3.22). The centroid (3.23) of each cluster is used as the new codeword. The Lloyd iteration ends when some pre-specified threshold ϵ for change in the distortion is achieved.

There are two drawbacks in designing codebook using Lloyd algorithm for practical applications. First, in applications such as MIMO communication systems, the receiver may need to perform the codebook training for the observed channel. Due to limitations in computational resources and time, such training may be difficult in real time. The Lloyd algorithm generally requires large number of iterations to arrive at codebook with low quantization error. Second, the final code-

Algorithm 4 Lloyd algorithm for error tangent direction codebook

Input: Training set $\tau = \{\mathbf{x}_\ell\}_{\ell=1}^{M}$ and initial codebook $\mathcal{C}_d^{(1)}$

1: **repeat**

2: Partition τ into cluster sets using nearest neighbor condition with suitable tie breaking rule

$$\mathcal{R}_i = \{\mathbf{x}_\ell \in \tau : d(\mathbf{x}_\ell, \mathbf{c}_{d,i}^{(m)}) \leq d(\mathbf{x}_\ell, \mathbf{x}_{d,n}^{(m)}), \forall n \neq i\} \tag{3.22}$$

3: Compute the centroid $\mathbf{x}_c$ for the each cluster set $\mathcal{R}_i$ [74]

$$\mathbf{x}_c(\mathcal{R}_i) = \arg\min_{\mathbf{x}} \sum_{\ell=1}^{M} d^2(\mathbf{x}_\ell, \mathbf{x}),\ \mathbf{x}_\ell \in \mathcal{R}_i \tag{3.23}$$

4: Form the new codebook $\mathcal{C}_d^{(m+1)} = \{\mathbf{x}_c(\mathcal{R}_i)\}_{i=1}^{N_d}$

5: Compute the average distortion for the new codebook

$$\mathcal{D}_{\text{ave}}(m+1) = \frac{1}{M} \sum_{i=1}^{N_d} \sum_{\ell=1}^{M} d^2(\mathbf{x}_\ell, \mathbf{x}_c(\mathcal{R}_i))\mathcal{I}(\mathbf{x}_\ell) \tag{3.24}$$

where $\mathcal{I}$ is the indicator function given by

$$\mathcal{I}(\mathbf{x}_\ell) = \begin{cases} 1, \text{ if } \mathbf{x}_\ell \in \mathcal{R}_i \\ 0, \text{ otherwise} \end{cases} \tag{3.25}$$

6: **until** $|\mathcal{D}_{\text{ave}}(m+1) - \mathcal{D}_{\text{ave}}(m)| < \epsilon$

Output: $\mathcal{C}_d^{(m+1)}$

book needs to be conveyed to the transmitter for the GPC decoder. Communicating the codebook takes away the communication resource. Both problems can be overcome by training offline and using a predefined fixed codebook. In Section 4.7, it is shown that the Grassmannian codebooks [67] work well for the error tangent direction codebook.

3.4.3 Initialization

To maintain synchronous operation of the encoder and the decoder, it is important to initialize the encoder and the decoder with the same initial estimated vectors in the memory. Two approaches may be considered for initialization. One approach is to perform an initialization process so that the two estimated vectors $\hat{\mathbf{x}}_{k-1}$ and $\hat{\mathbf{x}}_{k-2}$ are communicated from the encoder to the decoder. Since the complete description of $\hat{\mathbf{x}}_{k-1}$ and $\hat{\mathbf{x}}_{k-2}$ must be communicated to the decoder, there is system dependent communication overhead. Another approach is to use the one-shot memoryless quantization technique to set the two vectors. This approach is attractive since the codeword can be efficiently communicated to the decoder. In particular, if the same codebook is used for the error tangent direction codebook and one-shot memoryless quantization codebook, there are no codebook memory overhead resulting in efficient implementation. A consequence of using memoryless quantization approach for initialization is that there may be an initial transient period in which the quantization error is larger than the steady state condition. This is because the memoryless quantization generally results in a larger quantization error, as shown in Section 3.5.

3.4.4 Delayed Feedback Model

Feedback delay is a performance limiting factor for CSI sensitive systems such as MU-MIMO systems [38, 99, 103]. For example, in commercial wireless systems, the channel is typically estimated using preambles or specially designed pilot sequence spaced in time and frequency [2, 38, 41, 93]. Therefore, the conventional assumption of zero delay feedback with update at every channel realization is essentially a non-causal assumption which is not practical.

In this chapter, a periodic channel estimation interval T_H, which may be multiple symbols or frames in length, is also considered [38]. This is reasonable considering the time required for the receiver to acquire the signal, compute the channel estimate and feedback information, place the feedback information into the reverse link protocol, and process the feedback information at the transmitter for its intended use. Therefore, the feedback delay, denoted T_d, is modelled as the time between the beginning of a channel and the time in which the CSI becomes available at the transmitter. Fig. 3.4 graphically illustrates the limited feedback scenario with and without delay. With delayed limited feedback, the shaded region in Fig. 3.4 corresponds to the time when there is a mismatch in the beamformer used at the transmitter. In this chapter, the case where $T_d = T_H$, referred to as unit delay, is considered. Finally, feedback rate $\mathcal{R}_f$ is used to mean the number of bits used for feedback every T_H.

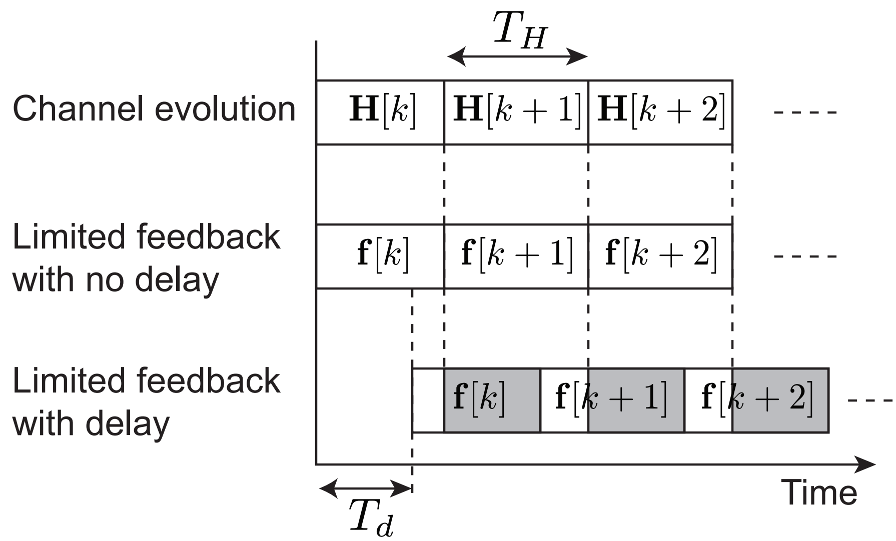

Figure 3.4: The timing diagram illustrates practical limited feedback delay in the system where the beamformer mismatch (shaded region) may occur. The channel observation and estimation interval is denoted T_H and the cumulative feedback delay is denoted T_d.

3.4.5 Predictive Coding for Delayed Feedback

In this section, a new form of predictive coding is proposed in which the predicted vector is used as the output of the decoder using the adaptive step size Grassmannian predictor derived in Section 3.4.1. The main idea is to use predictive coding framework to encode prediction error while simultaneously updating the predictor with optimal step size. By using the step size optimized predicted vector as the output, a unit delay between the encoder and the decoder can be accounted for while providing high resolution output.

The pseudo code and the block diagram for the proposed predictive encoder are provided in Algorithm 5 and Fig. 3.5, respectively. At time k, an error tangent

Algorithm 5 Proposed adaptive step size GPC encoder algorithm

Input: $\mathbf{x}[k]$

Initialize $\tilde{\mathbf{x}}[1]$ and $\hat{\mathbf{x}}[0]$

for all k=1,2,... **do**

  $\mathbf{e}[k] = L(\tilde{\mathbf{x}}[k], \mathbf{x}[k])$

  $(\ell[k], i[k]) = Q(\mathbf{e}[k])$

  $\hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1)$

  $\tilde{\mathbf{x}}[k+1] = P_G(\hat{\mathbf{x}}[k-1], \hat{\mathbf{x}}[k], t_{\text{opt}})$

end for

Output: $\ell[k], i[k]$

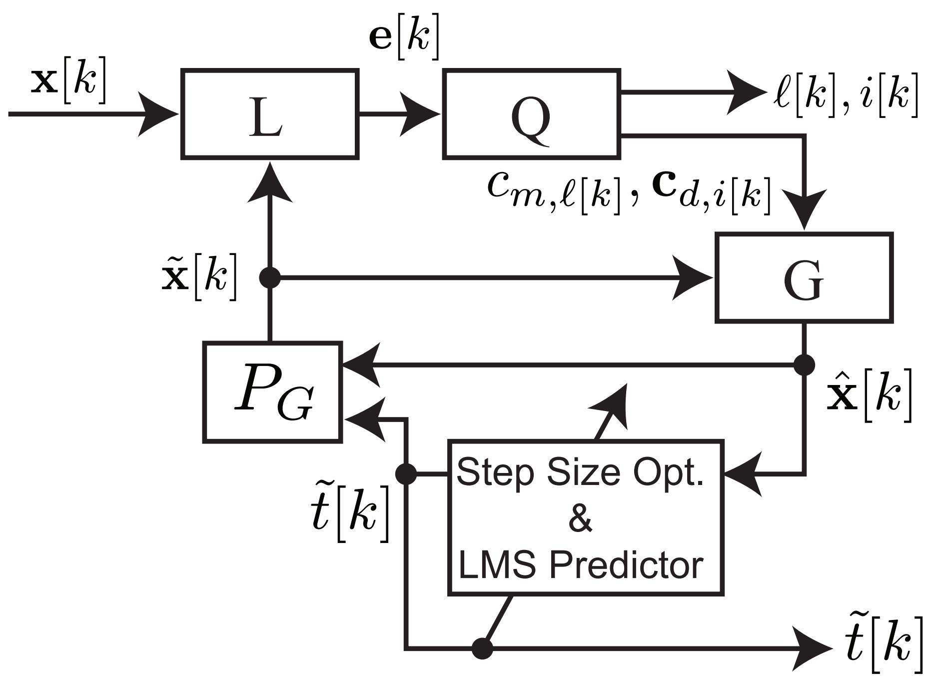

Figure 3.5: Block diagram of proposed adaptive step size predictive encoder on the Grassmann manifold.

vector $\mathbf{e}[k]$ emanating from the predicted vector $\tilde{\mathbf{x}}[k]$ to the current observed vector $\mathbf{x}[k]$ is computed using (4.12). Quantization is performed in two steps using two codebooks. Let $\mathcal{C}_m = \{c_{m,\ell}\}_{\ell=1}^{N_m}$ with N_m codewords denote the codebook of error tangent magnitudes in nonnegative reals and $\mathcal{C}_d = \{\mathbf{c}_{d,i}\}_{i=1}^{N_d}$ with N_d codewords denote the codebook of unit norm error tangent directions in $\mathbb{C}^n$. Please see [48] for details of the codebook design. First, the error tangent magnitude is quantized according to

$$\ell[k] \quad = \quad \underset{i\in\{1,2,\ldots,N_m\}}{\arg\min} \; |\|\mathbf{e}[k]\| - c_{m,i}| \tag{3.26}$$

where $\ell[k]$ is the index of the selected error tangent magnitude codeword in $\mathcal{C}_m$. This sets the radius of the quantized error tangent vector. Next, the error tangent direction codeword that minimizes the chordal distance error between the estimated and observed vector is computed as

$$i[k] \quad = \quad \underset{i\in\{1,2,\ldots,N_d\}}{\arg\min} \; d(G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i}, 1), \mathbf{x}[k]). \tag{3.27}$$

The unit norm tangent direction codeword that yields the estimate with the minimum chordal distance is selected. For notational brevity, we denote the two step quantization by an operator $Q(\mathbf{e}[k]) = (\ell[k], i[k])$ mapping the tangent vector to the codeword indices. The indices of the direction and magnitude codewords are transmitted to the decoder via finite rate communication channel resulting in feedback rate of $\log_2(N_d) + \log_2(N_m)$ bits. Continuing at the encoder, the estimated vector is

$$\hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1). \tag{3.28}$$

Finally, the adaptive step size Grassmannian prediction is used to obtain the predicted vector $\tilde{\mathbf{x}}[k+1]$ based on past estimates $\hat{\mathbf{x}}[k-1]$ and $\hat{\mathbf{x}}[k]$ and predicted

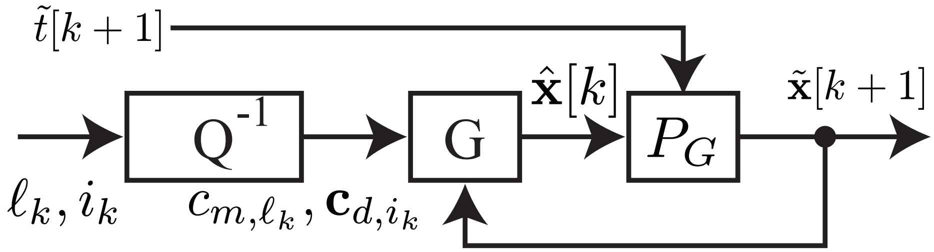

Figure 3.6: Block diagram of proposed predictive decoder on the Grassmann manifold for delayed feedback systems.

$\tilde{t}[k+1]$. Unfortunately, the predicted step size parameter results in an additional feedback requirement. We show in Section 4.7 that the predicted step size can be fed back with small overhead, either feeding back every T_H or less frequently. Another alternative is to feedback the LMS filter coefficients and treat the step size predictor as error whitening filter [17] but we have found that this provides no additional performance gain for increased feedback overhead.

Algorithm 6 Proposed adaptive step size GPC decoder algorithm

Input: $\ell[k], i[k]$
1: Initialize $\tilde{\mathbf{x}}[1]$ and $\hat{\mathbf{x}}[0]$
2: **for all** k=1,2,… **do**
3: $c_{m,\ell[k]} \cdot \mathbf{c}_{d,i[k]} = Q^{-1}(\ell[k], i[k])$
4: $\hat{\mathbf{x}}[k] = G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1)$
5: $\tilde{\mathbf{x}}[k+1] = P_G(\hat{\mathbf{x}}[k-1], \hat{\mathbf{x}}[k], t_{\text{opt}})$
6: **end for**
Output: $\tilde{\mathbf{x}}[k+1]$

The pseudo code and the block diagram for the proposed decoder are provided in Algorithm 6 and Fig. 3.6, respectively. The error tangent magnitude and direction codebooks are assumed to be shared between the encoder and the decoder. The received indices are decoded in Q^{-1} to recover $\mathbf{c}_{d,i[k]}$ and $c_{m,\ell[k]}$. The updated estimate $\hat{\mathbf{x}}[k]$ is computed using the received codeword indices using (3.28). Similarly to the encoder, the prediction is performed using the general Grassmannian predictor with the received step size $\tilde{t}[k+1]$ to obtain $\tilde{\mathbf{x}}[k+1]$. The predicted vector $\tilde{\mathbf{x}}[k+1]$ is the output of the decoder which accounts for the unit delay in the feedback path.

Finally, initialization is important in memory-based architecture. For the first iteration of the decoder, the knowledge of $\tilde{\mathbf{x}}[k]$, or equivalently $\hat{\mathbf{x}}[k-1]$, $\hat{\mathbf{x}}[k-2]$, and $\tilde{t}[k]$ are needed. Otherwise, the the codeword represents the error tangent vector with possibly wrong base point. Two approaches for initializing $\hat{\mathbf{x}}$ were considered in [48]. A practical approach is to use the one-shot memoryless quantization technique to set the two vectors $\hat{\mathbf{x}}[k-1]$ and $\hat{\mathbf{x}}[k-2]$. This approach is attractive since the codewords can be efficiently communicated to the decoder. In particular, if the same codebook is used for the error tangent direction codebook and one-shot memoryless quantization codebook, there are no codebook memory overhead resulting in efficient implementation. Since the predictor in the decoder only needs to know the current step size, quantized $\tilde{t}[k]$ can be fed back without additional initialization overhead.

3.5 Analysis

In this section, the quantization error using small angle approximation is analyzed, distortion bounds are proven, and prediction gain and coding gain expressions are derived for the proposed GPC algorithm on the Grassmann manifold.

3.5.1 Small Angle Approximation

If $\mathbf{x}[2] \in \mathcal{G}_{n,1}$ is obtained by sufficiently small changes to $\mathbf{x}[1] \in \mathcal{G}_{n,1}$, the chordal distance between $\mathbf{x}[1]$ and $\mathbf{x}[2]$ is

$$\begin{aligned} d(\mathbf{x}[1], \mathbf{x}[2]) &= \sqrt{1 - |\mathbf{x}^*[1]\mathbf{x}[2]|^2} \\ &= \sqrt{\sin^2(\angle(\mathbf{x}[1], \mathbf{x}[2]))} \quad (3.29) \\ &\approx \|\mathbf{x}[1] - \mathbf{x}[2]\|_2 \quad (3.30) \end{aligned}$$

where (3.29) follows from the subspace angle of vectors [25] and (3.30) follows from the small angle approximation. Thus, for a sufficiently small perturbation around $\mathbf{x}[1]$, the subspace distance between $\mathbf{x}[1]$ and $\mathbf{x}[2]$ is approximated by the usual Euclidean distance.

The current observed vector at time k, $\mathbf{x}[k]$, may be expressed in terms of the predicted vector and the error tangent vector as

$$\begin{aligned} \mathbf{x}[k] &= G(\tilde{\mathbf{x}}[k], \mathbf{e}[k], 1) \\ &\approx \tilde{\mathbf{x}}[k] + \vec{\mathbf{e}}[k]\|\mathbf{e}[k]\|_2 \\ &= \tilde{\mathbf{x}}[k] + \mathbf{e}[k] \quad (3.31) \end{aligned}$$

using the small angle approximation. Furthermore,

$$\begin{aligned} \mathbf{x}^*[k]\mathbf{x}[k] &\approx (\tilde{\mathbf{x}}[k] + \mathbf{e}[k])^*(\tilde{\mathbf{x}}[k] + \mathbf{e}[k]) \\ &= 1 + 2\|\mathbf{e}[k]\|_2 \Re(\vec{\mathbf{e}}^*[k]\tilde{\mathbf{x}}[k]) + \|\mathbf{e}[k]\|_2^2 \quad (3.32) \\ &\approx 1. \end{aligned}$$

The second term, $\vec{\mathbf{e}}^*[k]\tilde{\mathbf{x}}[k]$, in (3.32) is zero because the unit norm tangent vector $\vec{\mathbf{e}}[k]$ is orthogonal to $\tilde{\mathbf{x}}[k]$. Similarly, the estimated signal can be expanded as

$$\begin{aligned} \hat{\mathbf{x}}[k] &= G(\tilde{\mathbf{x}}[k], c_{m,\ell[k]}\mathbf{c}_{d,i[k]}, 1) \\ &\approx \tilde{\mathbf{x}}[k] + c_{m,\ell[k]}\mathbf{c}_{d,i[k]} \end{aligned} \tag{3.33}$$

where $\mathbf{c}_{d,i[k]}$ and $c_{m,\ell[k]}$ are the selected error tangent direction and error tangent magnitude codewords, respectively. Both (3.31) and (3.33) reveal that for a small enough change, both vectors are expressed as an additive correction to the predicted vector. Thanks to the locally Euclidean property and using the usual 2-norm for the local difference, the prediction error is then

$$\|\mathbf{x}[k] - \hat{\mathbf{x}}[k]\|_2 \approx \|\mathbf{b}[k] - \mathbf{c}_{d,i[k]}c_{m,\ell[k]}\|_2. \tag{3.34}$$

Therefore, the estimation error can be approximated as the normed difference between the actual tangent vector and the quantized tangent vector. Thus for small changes in the observed vector, the accuracy of tangent direction and tangent magnitude determines the accuracy of the estimate.

3.5.2 Distortion Bounds

Distortion induced by a quantizer is an important measure of performance. In what follows, an upper and lower bound on the distortion due to the proposed GPC algorithm are derived.

Recall that a metric ball $B_\delta(\mathbf{z})$ with radius δ centered at $\mathbf{z} \in \mathcal{G}_{n,1}$ on the Grassmann manifold is defined as

$$B_\delta(\mathbf{z}) = \{\mathbf{y} \in \mathcal{G}_{n,1} : d(\mathbf{y}, \mathbf{z}) \leq \delta\} \tag{3.35}$$

such that $B_\delta(\mathbf{z}) \subset \mathcal{G}_{n,1}$. A closed form volume formula for $B_\delta(\mathbf{z})$ is given as [16]

$$\mathrm{Vol}(B_\delta(\mathbf{z})) = \delta^{2(n-1)}. \tag{3.36}$$

Consider $B_\gamma(\mathbf{z}) \subset \mathcal{G}_{n,1}$ with $\delta \leq \gamma$ and volume of $B_\delta(\mathbf{z})$ given by (3.36). Let $(d\mathbf{y})$ denote the differential form of the Haar measure on $\mathcal{G}_{n,1}$. The distortion in the ball normalized by the volume of the ball was shown to be [74, Lemma 1]

$$\int_{B_\gamma(\mathbf{z})} \frac{d^2(\mathbf{y}, \mathbf{z})(d\mathbf{y})}{\mathrm{Vol}(B_\gamma(\mathbf{z}))} = \left(\frac{2(n-1)}{2n}\right)\gamma^2. \tag{3.37}$$

For memoryless quantization, the volume together with a point density and covering assumption over the entire $\mathcal{G}_{n,1}$ are used to characterize distortion. For the proposed GPC algorithm, the Voronoi region is determined by the tangent direction and tangent magnitude codebooks which makes the covering argument difficult. To overcome this difficulty, it is assumed that the tangent magnitude codebook provides concentric annular partitions of the sphere cap centered around the predicted vector and the tangent direction codebook partitioning each annulus into equiangle sectors. The bounds are obtained by considering the ball that is enclosed in the smallest annular sector and the ball that encloses the largest annular sector. Similarly, the distortion upper bound is given by the volume of the ball that covers the Voronoi cell.

Let $\gamma_d = \min_{\mathbf{c}_{d,i}, \mathbf{c}_{d,k} \in \mathcal{C}_d, i \neq k} d(\mathbf{c}_{d,i}, \mathbf{c}_{d,k})$ denote the minimum chordal distance between the tangent direction codewords and $\gamma_m = \min_{c_{m,i}, c_{m,k} \in \mathcal{C}_m, i \neq k} |c_{m,i} - c_{m,k}|$ denote the minimum Euclidean distance between the tangent magnitude codewords. Similarly, let $\lambda_d = \max_{\mathbf{c}_{d,i}, \mathbf{c}_{d,k} \in \mathcal{C}_d, i \neq k} d(\mathbf{c}_{d,i}, \mathbf{c}_{d,k})$ denote the maximum chordal distance between the tangent direction codewords and

$$\lambda_m = \max_{c_{m,i}, c_{m,k} \in \mathcal{C}_m, i \neq k} |c_{m,i} - c_{m,k}|$$

denote the maximum Euclidean distance between the tangent magnitude codewords. Then the following lemma provides the bounds on the distortion for GPC algorithm.

Lemma 11 (Distortion bounds)**.** *If* $\gamma_{lower} = \min\{\gamma_d, \gamma_m\}$ *and* $\lambda_{upper} = \max\{\lambda_d, \lambda_m\}$, *the lower and upper quantization distortion bound is given by*

$$\begin{aligned} D_{lower} &= \left(\frac{2(n-1)}{2n}\right)\left(\frac{\gamma_{lower}}{2}\right)^2 \\ D_{upper} &= \left(\frac{2(n-1)}{2n}\right)\left(\frac{\lambda_{upper}}{2}\right)^2. \end{aligned} \tag{3.38}$$

Proof. Assume that the tangent direction and magnitude codebooks maps uniformly to an equiangle sectors of concentric annulus centered at the predicted vector. The lower bound is given by the volume of a metric ball that has ball radius which is smaller of the half minimum chordal distance of tangent direction codebook and half minimum distance of tangent magnitude codebook. The upper bound is similarly obtained by considering the volume of a metric ball which covers a Voronoi region. The bounds are exact since the metric ball volume formula is accurate [74, Lemma 1]. □

Unfortunately, no claim is made on the tightness of the bound since an accurate description of the Voronoi region obtained by the proposed tangent codebook remains to be an open problem. In Section 3.6.1, numerical examples are provided comparing the bounds obtained with actual distortion using fixed codebooks.

Using the obtained lower bound, the reduction in distortion lower bound compared to memoryless quantization on the Grassmann manifold can be further quantified. For $\mathcal{G}_{n,1}$, the lower bound on the fixed rate quantizer on the Grassmann manifold was shown to be

$$D_{\mathcal{G}_{n,1}}(N) = \left(\frac{2(n-1)}{2n}\right) N^{-\frac{1}{n-1}} \tag{3.39}$$

where N is the size of the codebook with rate $\log_2(N)$ bits [15, 74]. Suppose that γ_{lower} is dominated by the tangent direction codebook such that $\gamma_{\text{lower}} = \gamma_d$ and that Grassmannian codebook is used for the tangent direction codebook. Then, the lower bound for the GPC algorithm can be expressed as

$$\begin{aligned} D_{\text{lower}} &= \left(\frac{2(n-1)}{2n}\right)\left(\frac{\gamma_{\text{lower}}^2}{4}\right) \\ &= \frac{1}{4}\left(\frac{2(n-1)}{2n}\right)^2 N_d^{-\frac{1}{n-1}} \\ &= \frac{1}{4}\left(\frac{2(n-1)}{2n}\right) D_{\mathcal{G}_{n,1}}(N_d) \end{aligned} \tag{3.40}$$

showing that the lower bound is smaller than $D_{\mathcal{G}_{n,1}}(N_d)$ when $\gamma_d < \gamma_m$. For the applications considered in this chapter, γ_m is typically smaller than γ_d. Hence the reduction in distortion is typically greater than (3.40) and it is dominated by the error tangent magnitude quantization.

3.5.3 Performance Measures

The closed loop prediction gain ratio is often used in vector quantization literature [22] as a measure of how well the predictor performs with respect to the changes in the input. The closed loop prediction gain is usually written as the ratio of mean squared norm of the observed signal over mean squared norm of the prediction error. For the proposed GPC algorithm on the Grassmann manifold, the closed loop prediction performance is measured by

$$\begin{aligned} G_{\text{clp}} &= \frac{\mathbb{E}[\|\mathbf{x}[k]\|^2]}{\mathbb{E}[d^2(\tilde{\mathbf{x}}[k], \mathbf{x}[k])]} \\ &= \frac{1}{\mathbb{E}[d^2(\tilde{\mathbf{x}}[k], \mathbf{x}[k])]} \end{aligned} \tag{3.41}$$

where $d^2(\tilde{\mathbf{x}}[k], \mathbf{x}[k])$ denotes the squared chordal prediction error. In fact, (3.41) can be further expressed as a function of the tangent vector assuming that the small angle approximation holds. Using (3.30) and (3.31), the distance function in the denominator can be approximated as $d(\tilde{\mathbf{x}}[k], \mathbf{x}[k]) \approx \|\mathbf{e}[k]\|_2$. Therefore, the closed loop prediction gain for GPC algorithm becomes

$$G_{\text{clp}} = \frac{1}{\mathbb{E}[\|\mathbf{e}[k]\|_2^2]} \tag{3.42}$$

which shows the dependence of the closed loop prediction gain performance on the tangent magnitude. The tangent magnitude is in turn dependent on the changes in the observed process. A closed form relationship between the observed process and the tangent magnitude is in general difficult to obtain. In Section 3.6.2, some empirical results of the closed loop prediction gain performance for the proposed GPC algorithm are shown.

3.6 Numerical Results

In this section, experimental results are provided to illustrate the performance of the proposed GPC algorithm. First, distortions for various codebook sizes from simulation are compared with the bounds obtained in Section 3.5.2. Then, numerical results for closed loop prediction gain and chordal distance error are obtained. Finally, simulation results are shown for two applications: single user limited feedback beamforming system and zero forcing limited feedback multiuser MIMO system.

3.6.1 Distortion Bounds

A numerical example is presented illustrating the operational distortion and compared with the upper and lower bounds given in Lemma 11. Correlated 3×1 vectors were generated according to a second order autoregressive model with memory coefficients $\alpha_1 = 0.9$ and $\alpha_2 = 0.75$ with additive noise distributed according to zero mean complex Gaussian with variance $(0.01)^2$. The normalized vectors were considered to be the samples on $\mathcal{G}_{3,1}$ to which the proposed GPC algorithm was applied. For this experiment, an $N_d = 2^4$ tangent direction codebook was used and the tangent magnitude codebook size was varied from $N_m = 2^2$ to 2^5. Fig. 3.7 shows the operational distortion with upper and lower distortion bounds obtained in Lemma 11 as a function of the tangent magnitude codebook size. The lower bound captures the distortion trend over the range of codebook sizes. The lower bound of a memoryless quantization using a Grassmannian codebook with codebook sizes of 6, 7, 8, and 9 bits are illustrated so that the total number of bits used for the codebook matches that of the proposed GPC algorithm. The proposed GPC algorithm provides significant improvement in distortion over the memoryless quantization technique. Unfortunately, the upper bound from Lemma 11 is dominated by the resolution of the 4-bit tangent direction codebook which has higher distortion than the memoryless quantization with adjusted number of codebook size. Nevertheless, the result shows that a significant reduction in distortion is achieved by the proposed GPC algorithm and the achievable distortion can be controlled by the tangent magnitude codebook which is a simple uniform scalar codebook.

3.6.2 Closed Loop Prediction Gain And Prediction Error

To illustrate the dependence on the tangent direction and tangent magnitude codebooks, Fig. 3.8 shows the closed loop prediction gains for various error tangent magnitude codebook sizes and fixed tangent direction codebook of size $N_d = 64$. For these numerical examples, correlated 4×1 vector sequence was generated according to a first order autoregressive model (or Gauss-Markov model [20]) with correlation coefficient $\alpha = J_0(2\pi\beta)$ where J_0 is Bessel function of zeroth order and β is the normalized Doppler frequency. The sequence was generated according to

$$\mathbf{h}[k] = \alpha\mathbf{h}[k-1] + \sqrt{1-\alpha^2}\mathbf{z}[k] \tag{3.43}$$

where k is the time index and $\mathbf{z}[k]$ is a 4×1 vector with each entry drawn from an independent identically distributed (i.i.d.) zero mean complex white Gaussian process. The normalized vectors $\mathbf{x}[k] = \mathbf{h}[k]/\|\mathbf{h}[k]\|$ are a correlated sequence on the Grassmann manifold. For the tangent direction codebook, an $N_d = 2^6$ Grass-

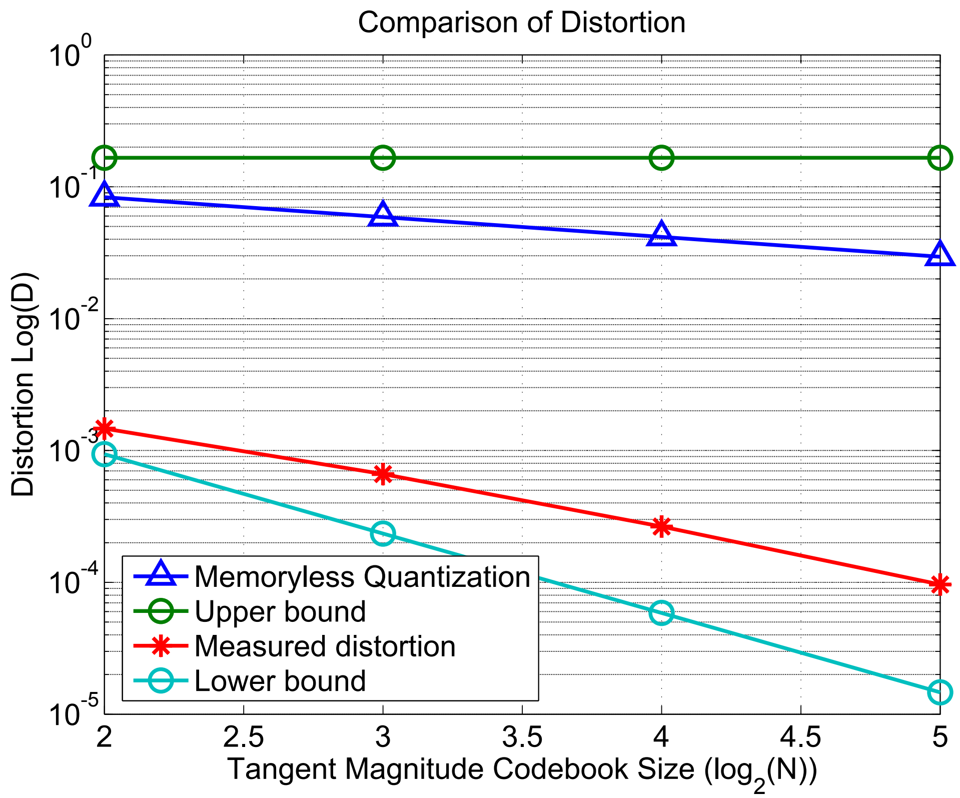

Figure 3.7: Comparison of operational distortion against the lower and upper bound for various tangent magnitude codebook size. Lower distortion bound for the memoryless quantizer using Grassmannian codebook is also shown.

mannian codebook [63] was used and the tangent magnitude codebooks were based on a uniform quantization between 0 and 1 using 2, 3, 4, and 5 bits. For an upper bound, closed loop prediction gain without quantizing the tangent magnitude is also shown. The result illustrates the dependence of closed loop prediction gain on tangent magnitude codebook size as a function of correlation parameter β. For highly correlated data, the tangent magnitude codebook resolution has higher impact on the closed loop prediction gain. This is because the smallest tangent magnitude

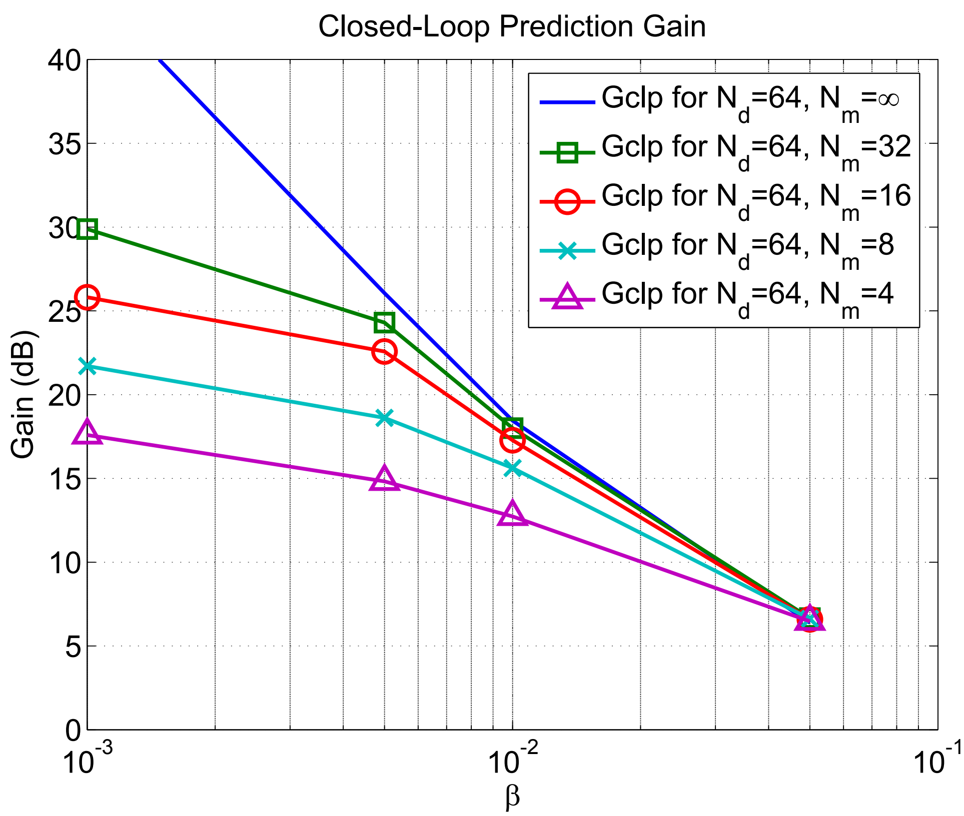

Figure 3.8: Closed loop prediction gain, G_{clp}, for $\mathcal{G}_{4,1}$ data with fixed tangent codebook ($N_d = 64$) and different tangent magnitude codebooks ($N_m = 2^2, 2^3, 2^4, 2^5$) over various correlation parameter β.

quantization level may be larger than the prediction error leading to an over estimation. If the tangent magnitude codebook is adjusted based on the correlation, e.g., quantize in the range of $[0, 0.1]$ instead of $[0, 1]$, this gap may be closed.

Another useful performance measure is the chordal distance error between the estimated vector $\hat{\mathbf{x}}[k]$ and the observed vector $\mathbf{x}[k]$. The chordal distance error $d(\hat{\mathbf{x}}[k], \mathbf{x}[k])$ shows how close the estimated vector is to the observed vector using the proposed GPC algorithm. In MIMO communication applications considered

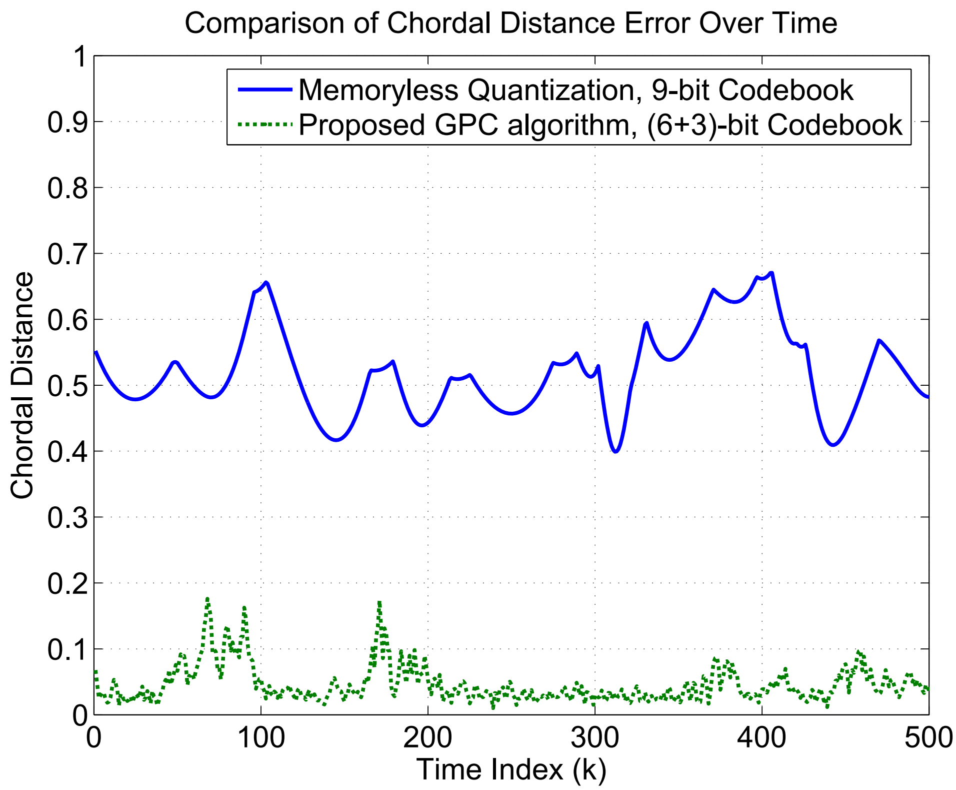

Figure 3.9: Chordal distance comparison over time between memoryless quantizer using 9-bit Grassmannian codebook and the proposed GPC algorithm using 6-bit Grassmannian tangent direction codebook and 3-bit tangent magnitude codebook.

in Section 3.6.3 and 3.6.4, the chordal distance error has a direct impact on the respective communication theoretic performance measures. In Fig. 3.9, the chordal distance between $\hat{\mathbf{x}}[k]$ and $\mathbf{x}[k]$ and the chordal distance between the quantized vector and the observed vector for memoryless quantization using Grassmannian codebook with $N = 2^6$ are shown. Fig. 3.9 illustrates the substantial improvement in the quantization accuracy compared with memoryless technique.

To further illustrate the quantizer accuracy, the operational mean squared

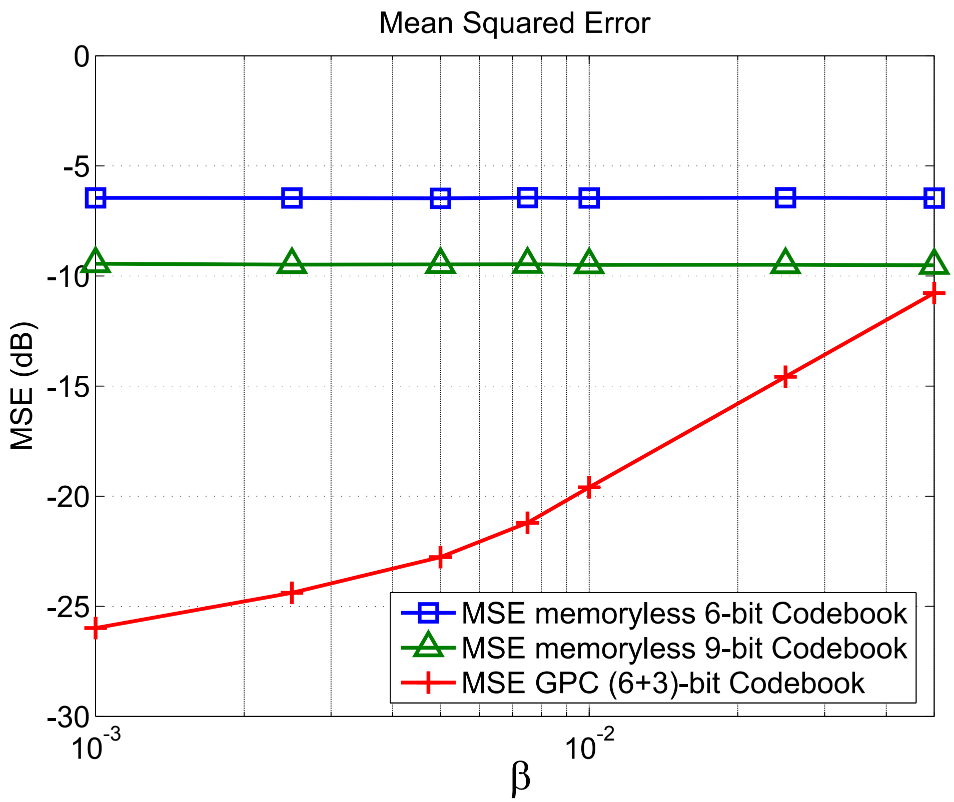

Figure 3.10: Mean squared error comparison between memoryless quantization using 6-bit and 9-bit Grassmannian codebook and the proposed GPC algorithm using 6-bit tangent direction and 3-bit tangent magnitude codebooks.

chordal distance error (MSE) is shown as a function of β for the proposed GPC algorithm and memoryless quantizer using Grassmannian codebook in Fig. 4.10. The memoryless quantizer provides approximately -7 dB of MSE whereas the proposed GPC algorithm provides as little as -26 dB of MSE which shows that significant accuracy can be obtained over memoryless quantization techniques. As the correlation decreases, the MSE approaches that of the memoryless quantization MSE.

Similarly, the operational MSE as a function of β are compared for delayed and non-delayed system [66]. Cases are shown for the memoryless 9-bit Grassmannian codebook with and without delay, GPC with naive prediction using delayed $\hat{\mathbf{x}}$ as output [48], the proposed algorithm with two types of step size adaptation, and the MSE using step size optimized prediction only in Fig. 3.11. First, observe that the memoryless codebook results in the worst MSE at -10dB. The delayed memoryless codebook degrading slightly for higher values of β. The naive GPC using delayed $\hat{\mathbf{x}}$ as the output at $R_f = 9$ bits still outperforms the memoryless case for most of β. Unfortunately, as the correlation increases, i.e., β decreases, the naive one step prediction tends to over-predict resulting in saturating MSE. The proposed adaptive step size GPC, however, achieves substantial MSE improvement across β. The proposed algorithm with optimized step size, marked with $*$, results in $R_f = 9 + 4 = 13$ bits whereas the proposed algorithm with step update every $12T_H$, marked with $\triangleleft$, results in $R_f = 9 + 4/12 = 9.33$ bits. Thus providing infrequent step size feedback can still yield substantial MSE improvement while providing flexibility to control R_f. Finally, the prediction only case uses past time-series samples on the Grassmann manifold to predict future vectors with step size prediction. The prediction only case provides the lower bound on the achievable MSE using the proposed adaptive step size prediction formulation.

3.6.3 Application to Limited Feedback Beamforming System

In this section, the application of the proposed GPC algorithm to single user limited feedback beamforming [67] system is illustrated. It is assumed that

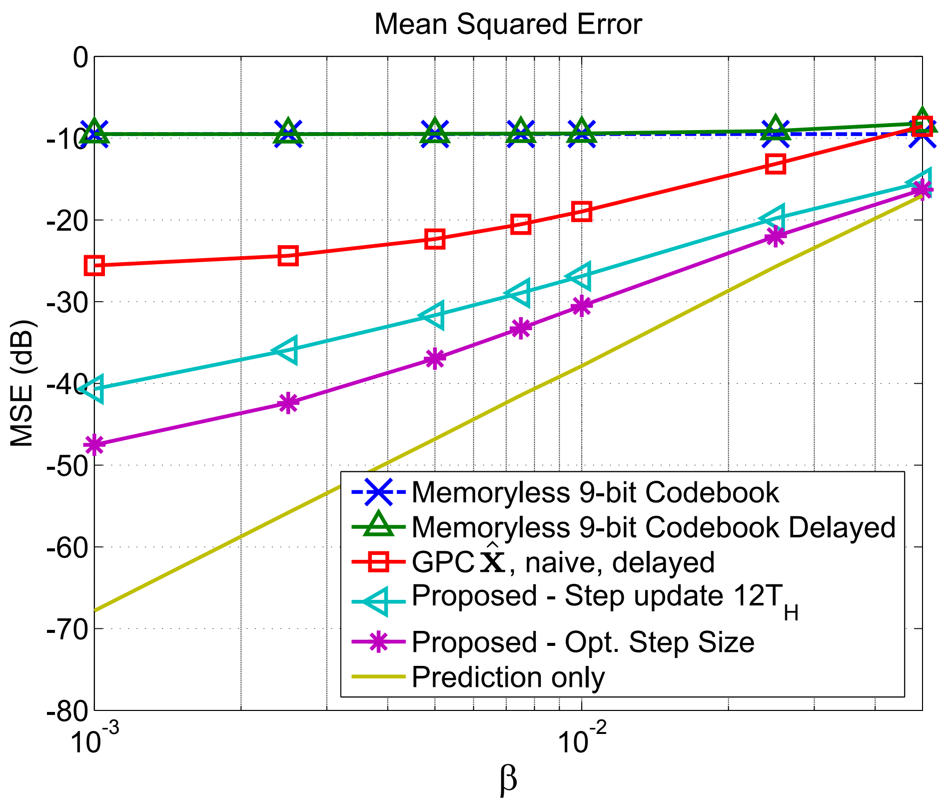

Figure 3.11: Mean squared chordal distance error comparison over correlation parameter β. Memoryless quantization using 9-bit Grassmannian codebook with delay(Δ) and without delay ($\times$) shows the deteriorating MSE as the correlation decreases. The proposed step size optimization and prediction with infrequent feedback $T_H = 12$ and $R_f = 9.33$ bits ($\lhd$) shows significant MSE improvement across the range of β.

the transmitter has $N_t = 4$ transmit antennas and the receiver has a single antenna ($N_r = 1$). The channel is assumed to be temporally correlated using the first order autoregressive model as in Section 3.6.2 where the normalized Doppler frequency is $\beta = f_D T_s$ with Doppler frequency f_D and symbol sampling interval T_s. The $N_t \times 1$ channel vectors are generated according to (3.43). The receiver is assumed to have a perfect estimate of the CSI and the feedback is assumed to be delay and error free. Assuming perfect synchronization and sampling, the baseband input-output relationship can be written as $\mathbf{y}[k] = \sqrt{P}\mathbf{h}^*[k]\mathbf{f}[k]\mathbf{s}[k] + \mathbf{n}[k]$ where P is the total transmit power, $\mathbf{f}[k]$ is the beamforming vector, $\mathbf{s}[k]$ is the normalized transmit symbols, and $\mathbf{n}[k]$ is an additive noise with each entry i.i.d. with distribution according to $\mathcal{CN}(0, N_0)$.

Four scenarios are considered for comparison. The first scenario is where the transmitter has perfect knowledge of the CSI. In this case, the beamformer is found as the normalized channel vector. The second scenario is the conventional memoryless limited feedback approach using the Grassmannian codebook [63]. The beamformer $\hat{\mathbf{f}}[k]$ is selected from the Grassmannian codebook $\mathcal{F}$ using $\hat{\mathbf{f}}[k] = \arg\max_{\mathbf{f}\in\mathcal{F}} \|\mathbf{h}^*[k]\mathbf{f}\|^2$ [67]. The third scenario uses the Householder based PVQ [62]. The codebook was trained using the closed loop approach for best performance and it was assumed to be known to both the transmitter and the receiver. The fourth scenario uses the proposed GPC algorithm. The initial beamformer information is assumed to be available at the transmitter. For every new channel estimate, the normalized channel vector is fed into the GPC algorithm and the indices of the tangent direction and magnitude codewords are fed back to the transmitter.

Fig. 3.12 shows the symbol error rate of $N_t = 4$ beamforming system over a temporally correlated channel ($f_D T_s = 0.005$) where 6 and 9-bit Grassmannian codebooks [63] were used for conventional memoryless limited feedback strategy. For the proposed GPC algorithm, a 6-bit tangent direction codebook and a 3-bit tangent magnitude codebook were used for a total of 9-bit feedback. The perfor-

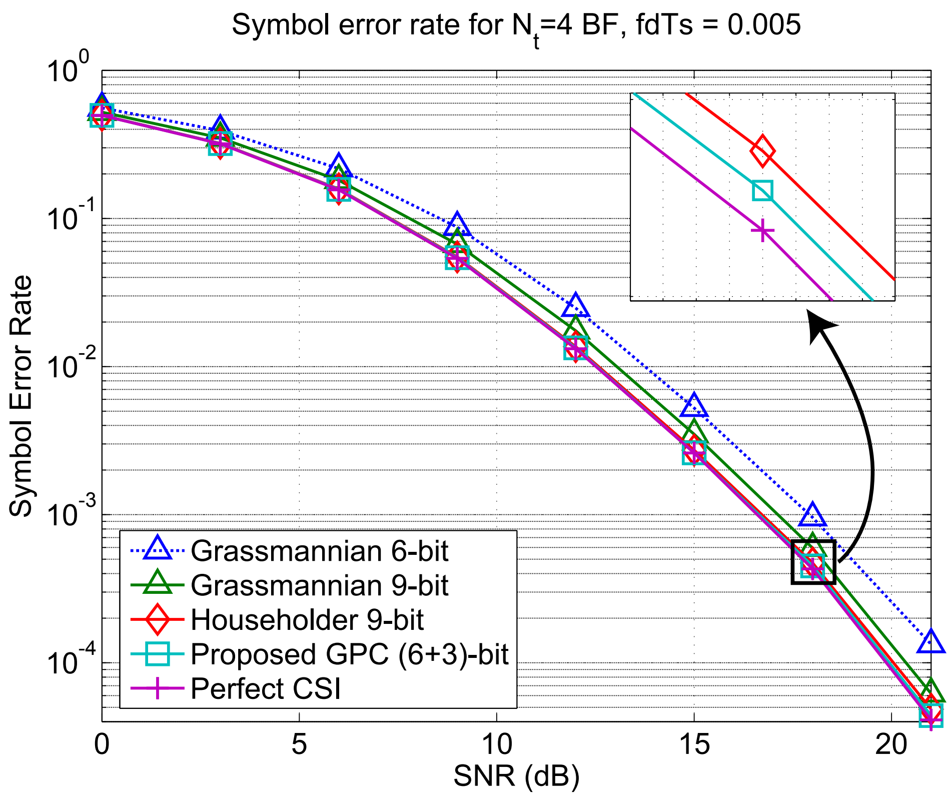

Figure 3.12: Vector symbol error rate plot for $N_t = 4$ beamforming system with perfect CSI, conventional limited feedback using 6 and 9-bit Grassmannian codebook, Householder based PVQ using 9-bit codebook, and proposed GPC algorithm using 6-bit Grassmannian tangent direction codebook and 3-bit tangent magnitude codebook.

mance of the complex Householder based PVQ using 9-bit codebook is also illustrated. At high SNR, the $(6 + 3)$-bit codebook with GPC algorithm outperforms the optimized Householder based PVQ and achieves performance very close to the perfect CSI case. The proposed GPC algorithm uses the same codebook as the memoryless limited feedback system and the codebook need not be optimized as in the Householder based PVQ. In Fig. 3.13, the achievable rate for limited feedback

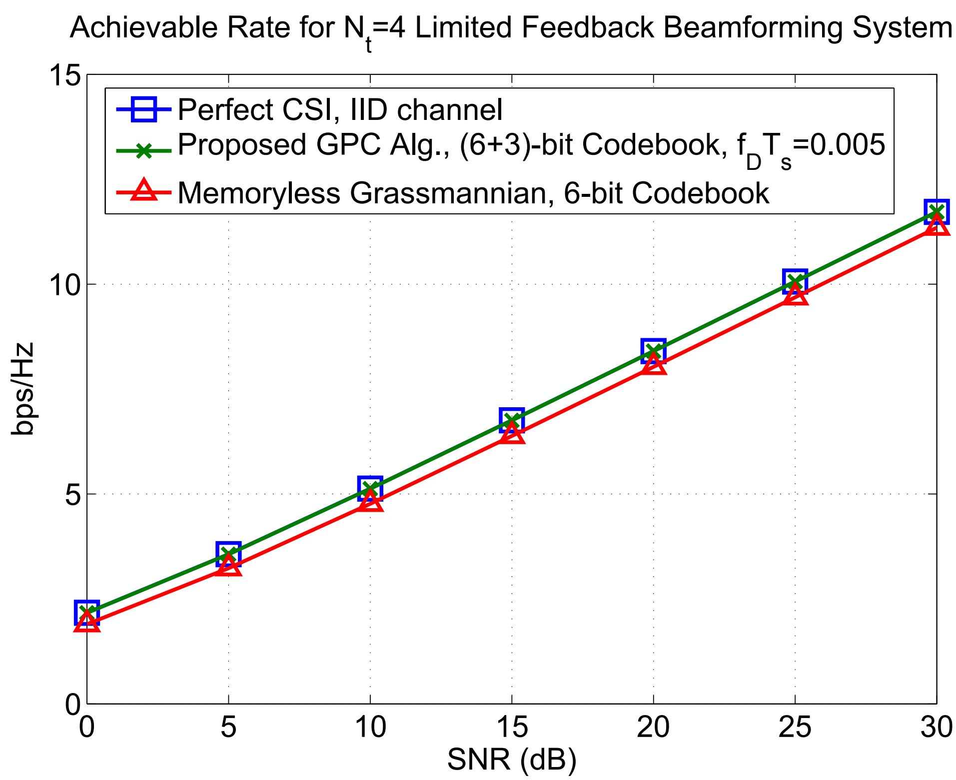

Figure 3.13: Achievable rate of $N_t = 4$ beamforming system using perfect CSI, conventional 6-bit Grassmannian codebook, and proposed GPC algorithm using 6-bit Grassmannian tangent direction codebook and 3-bit tangent magnitude codebook over temporally correlated channel with $f_D T_s = 0.005$.

beamforming system with perfect CSI, proposed GPC algorithm, and memoryless technique with Grassmannian codebook are shown. Due to significantly improved quantization error, the achievable rate of the proposed GPC algorithm essentially overlaps with that of the perfect CSI case.

Next, the same system with unit feedback delay is considered. Three scenarios are considered for comparison. The first scenario is where the transmitter has perfect knowledge of the CSI. In this case, the beamformer is found as the normalized channel vector. The second scenario is the conventional memoryless limited feedback approach using the Grassmannian codebook with unit feedback delay. The beamformer $\mathbf{f}[k]$ is selected from the Grassmannian codebook $\mathcal{F}$ using (2.1) and applied to the transmission one unit time later [67]. The final scenario uses the proposed adaptive step size GPC algorithm. We assume that the initial beamformer information is available at the transmitter. For every new channel estimate, the normalized channel vector is fed into the encoder and the indices of the tangent direction, magnitude, and predicted step size codewords are fed back to the transmitter.

Fig. 3.14 shows the symbol error rate of $N_t = 4$ beamforming system over a temporally correlated channel ($f_D T_H = 0.005$) where 9-bit Grassmannian codebook was used for delay free and delayed memoryless limited feedback strategies. For the proposed algorithm, a 6-bit tangent direction codebook and a 3-bit tangent magnitude codebook were used. Thus the feedback rate is constrained to be $R_f = 9$ bits. The proposed algorithm essentially overlaps with perfect CSIT case showing the high resolution feedback. Both the proposed and the memoryless approach begins to marginally deteriorate in high SNR regime for delayed feedback. The proposed algorithm thus provides near perfect CSIT performance under unit delay assumption.

In this section, it was shown that the proposed GPC algorithms can achieve symbol error rate very close to the case with perfect CSI at the transmitter. The achievable rate simulation has shown that, even in the presence of feedback delay,

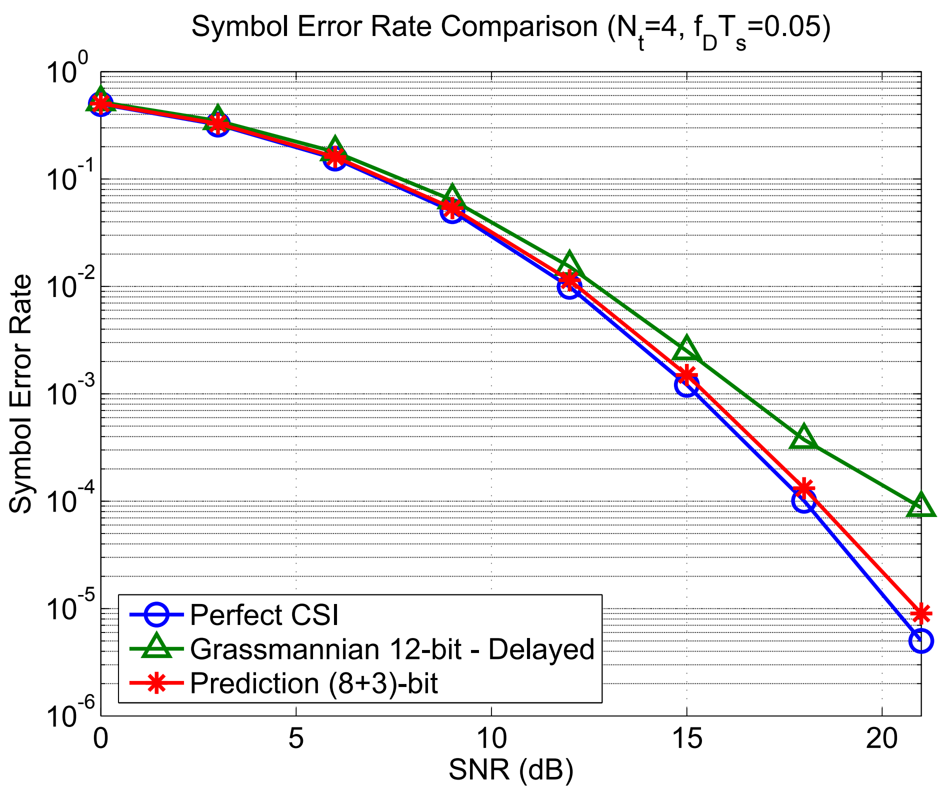

Figure 3.14: Vector symbol error rate plot for $N_t = 4$ beamforming system with perfect CSI, memoryless limited feedback using 12-bit Grassmannian codebook with unit delay, and the proposed algorithm using 8-bit Grassmannian tangent direction codebook, 3-bit tangent magnitude codebook, and 4-bit step size codebook.

essentially a perfect CSI performance can be obtained. Furthermore, a fixed Grassmannian codebook can be used for tangent direction codebook that does not require any training or overhead as in the Householder based PVQ method in [62].

3.6.4 Application to Zero Forcing Multiuser MIMO System

In this section, the application of proposed GPC algorithm to limited feedback multiuser MIMO system using zero forcing precoding is illustrated [51]. It is assumed that the base station has N_t transmit antennas and there are $U = N_t$ mobile users each equipped with single receive antenna. To isolate the impact of using the predictive coding for limited feedback, U users are already assumed scheduled from possibly large number of user pool and the problem of user scheduling is not considered in this dissertation. Let $s^{(u)}[k]$, $\mathbf{v}^{(u)}[k]$, and $\mathbf{h}^{(u)}[k]$ be the complex transmit symbol, $N_t \times 1$ unit norm beamforming vector, and $N_t \times 1$ channel vector for u-th user at time index k, respectively. Then, the input-output relationship for u-th user may be written as

$$y^{(u)}[k] = \mathbf{h}^{(u)*}[k]\mathbf{v}^{(u)}[k]s^{(u)}[k] + \mathbf{h}^{(u)*}[k] \sum_{n=1,n\neq u}^{U} \mathbf{v}^{(n)}[k]s^{(n)}[k] + n^{(u)}[k] \quad (3.44)$$

where $n[k]$ is an independent complex Gaussian noise with unit variance. The first term in (3.44) is the desired signal for u-th user while the second summation term is the interference signal. The signal to interference plus noise ratio (SINR) for the u-th user can be written as

$$\text{SINR}^{(u)} = \frac{\frac{P}{N_t}|\mathbf{h}^{(u)*}\mathbf{v}^{(u)}|^2}{1 + \sum_{n\neq u} \frac{P}{N_t}|\mathbf{h}^{(u)*}\mathbf{v}^{(n)}|^2}. \quad (3.45)$$

If the transmit signal $s^{(u)}$ is assumed to be Gaussian, the achievable rate for user u is given by

$$\mathcal{R}^{(u)} = \log_2(1 + \text{SINR}^{(u)}) \quad (3.46)$$

and the sum rate as $\mathcal{R} = \sum_{u=1}^{U} \mathcal{R}^{(u)}$.

The SINR expression (3.45) shows that the amount of interference depends on the design of the beamforming vectors. Zero forcing uses beamforming vectors such that they are orthogonal to other user's channel vectors, i.e., $\mathbf{h}^{(u)}[k]\mathbf{v}^{(n)}[k] = 0$ for $n \neq u$, to null the inter user interference. Let $\mathbf{H} = [\mathbf{h}^{(u)} \cdots \mathbf{h}^{(u)}]^*$ be the $U \times N_t$ composite channel matrix. With perfect CSI, the interference can be perfectly eliminated by choosing the unit norm beamforming vector as the normalized columns of pseudo inverse composite channel matrix, i.e., $\mathbf{v}^{(u)} = [\mathbf{H}^{\dagger}]_{:,u}/\|[\mathbf{H}^{\dagger}]_{:,u}\|$. Thus zero forcing technique creates interference free U parallel channels providing full multiplexing gain but with some power loss due to normalization [83].

In limited feedback multiuser MIMO systems, quantized channel vector information is fed back to the transmitter from each user [36, 51]. Assuming that perfect channel estimate $\mathbf{h}^{(u)}$ is obtained, the quantization of the channel shape $\mathbf{g}^{(u)} = \mathbf{h}^{(u)}/\|\mathbf{h}^{(u)}\|$ is considered and it is assumed that the scalar channel gain is known perfectly [36]. In this regime, the SINR can be rewritten as

$$\text{SINR}^{(u)} = \frac{\frac{P}{N_t}\|\mathbf{h}^{(u)}\|^2|\mathbf{g}^{(u)*}\mathbf{v}^{(u)}|^2}{1 + \sum_{n \neq u} \frac{P}{N_t}\|\mathbf{h}^{(u)}\|^2|\mathbf{g}^{(u)*}\mathbf{v}^{(n)}|^2}. \tag{3.47}$$

Two observation are noted from (3.47). First, if the channel vector $\mathbf{h}^{(u)}$ is an i.i.d. vector distributed according to $\mathcal{CN}(0, 1)$, $\mathbf{g}^{(u)}$ is isotropically distributed on the N_t-dimensional hyper-sphere. Second, due to the absolute value around $\mathbf{g}^{(u)*}\mathbf{v}^{(u)}$, the SINR is independent of arbitrary unitary rotations of the channel direction. That is, $|\mathbf{g}^{(u)*}\mathbf{v}^{(u)}|^2 = |e^{j\theta}\mathbf{g}^{(u)*}\mathbf{v}^{(u)}|^2$ for $\theta \in (0, 2\pi]$. Therefore, the space of channel shape may be identified as the Grassmannian manifold. Thus, the problem is to feedback channel shapes on the Grassmann manifold from each user u, and use the collected channel shape information at the transmitter to design the beamforming vectors by zero forcing.

In random codebook based limited feedback multiuser MIMO systems, each user has a normalized channel vector codebook of size N_{RC} which is shared with the transmitter [36, 51]. The transmitter maintains U tables of size N_{RC} codebooks. Each user selects the codeword with minimum chordal distance from the normalized channel vector estimate. The index of the selected codeword using $\log_2(N_{\mathrm{RC}})$ bits is fed back to the transmitter. The transmitter collects the decoded channel vectors $\hat{\mathbf{h}}^{(u)}$ to form the composite channel matrix $\hat{\mathbf{H}} = [\hat{\mathbf{h}}^{(u)} \cdots \hat{\mathbf{h}}^{(u)}]^*$. The beamforming vectors are computed as $\hat{\mathbf{v}}^{(u)} = [\hat{\mathbf{H}}^{\dagger}]_{:,u}/\|[\hat{\mathbf{H}}^{\dagger}]_{:,u}\|$. It was shown in [51] that sum rate performance becomes interference limited as SNR increases and that codebook size needs to be increased linearly as a function of SNR to maintain multiplexing gain. Herein, lies two practical limitation of random codebook technique. First, coordination of random codebooks among multiple users may be difficult in practice. Second, the codebook size that approaches the achievable sum rate becomes impractical even for moderate SNR. The proposed GPC algorithm solves both problems.

For the limited feedback multiuser MIMO system using the proposed GPC algorithm, each user is assumed to have perfect channel vector estimate. Without loss of generality, it is assumed that encoder and decoder are initialized. Then, each user performs the prediction as described in Section 3.4 and feedback the indices of quantized tangent direction and tangent magnitude codewords. The transmitter uses the received indices and performs the prediction as depicted in Fig. 3.3. Same

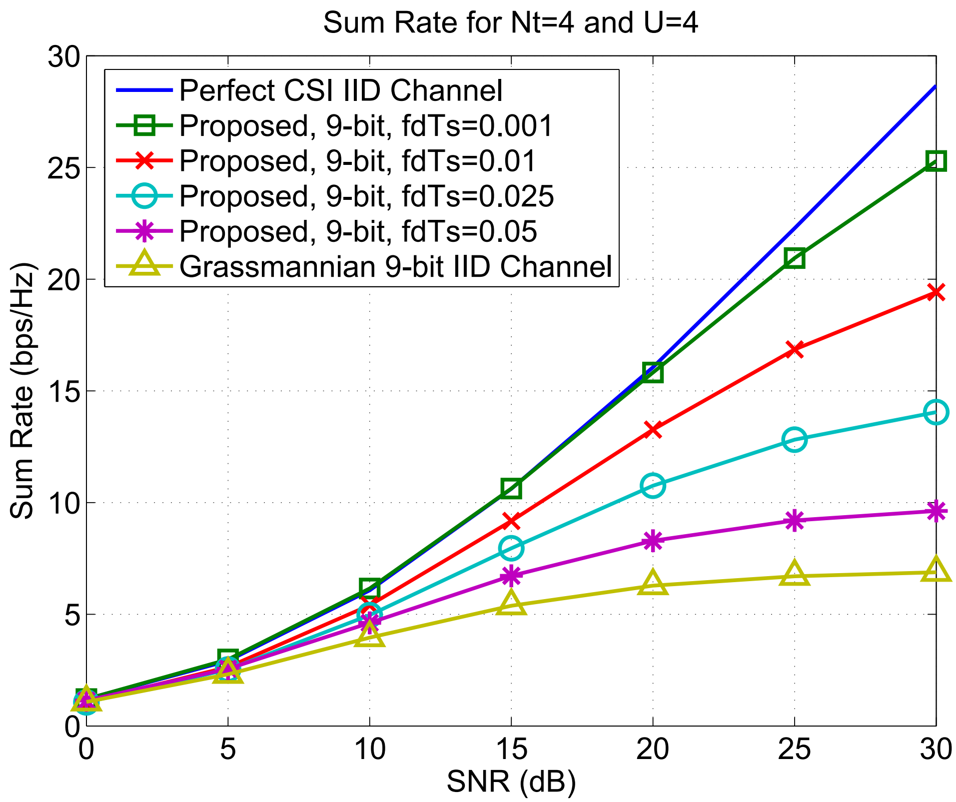

Figure 3.15: Sum rate for $N_t = U = 4$ i.i.d. channel with perfect CSI, i.i.d. channel with 9-bit Grassmannian codebook, and the proposed GPC algorithm with 6-bit Grassmannian tangent direction codebook and 3-bit tangent magnitude codebook for various normalized Doppler frequencies.

procedure is used to compute the beamforming vectors using the output of the GPC decoder.

To compare the random codebook approach and the proposed GPC algorithm, each user's vector channel is assumed to be temporally correlated with correlation according to $J_0(2\pi f_D T_s)$ [98]. Fig. 3.15 illustrates the achievable sum rate estimate obtained with i.i.d. Gaussian channel with perfect CSI at the transmitter, 9-bit random codebook limited feedback approach, and the proposed GPC limited feedback using 6-bit Grassmannian codebook and 3-bit tangent magnitude codebook for various normalized Doppler frequency $f_D T_s$. Contrary to the random codebook strategy, the proposed GPC algorithm provides significant sum rate gain. In fact, for $f_D T_s = 0.001$, the system starts to become interference limited above SNR of 20dB illustrating the superior CSI accuracy when the channel is highly correlated. Furthermore, each user is equipped with the same codebooks which eliminates the need to store multiple codebooks at the transmitter, thus reducing the overhead for practical applications.

Next, the sum rate performance of limited feedback zero forcing multiuser MIMO system using the proposed algorithm in the presence of delay is considered. Fig. 3.16 and Fig. 3.17 illustrates the achievable sum rate estimates for various approaches with $\beta = f_D T_H = 0.04$ and 0.02, respectively. The upper limit is the case with i.i.d. Gaussian channel with perfect CSI at the transmitter. For all the prediction results, baseline $R_f = 10$ bits were used where 6 bits were allocated for tangent direction codeword and 4 bits were allocated for tangent magnitude codeword. Additional 4 bits are allocated for step size feedback. The standard LMS prediction approach uses 10 bits to match the baseline feedback rate R_f of other methods presented.

First, the standard LMS based prediction provides the lowest sum rate which improves by 3 bits at higher SNR when the normalized Doppler increases to 0.02. The naive predictive coding using $\tilde{\mathbf{x}}$ is marked with $\triangleleft$ showing significant sum rate improvement [48]. The proposed adaptive step size predictive coding, however,

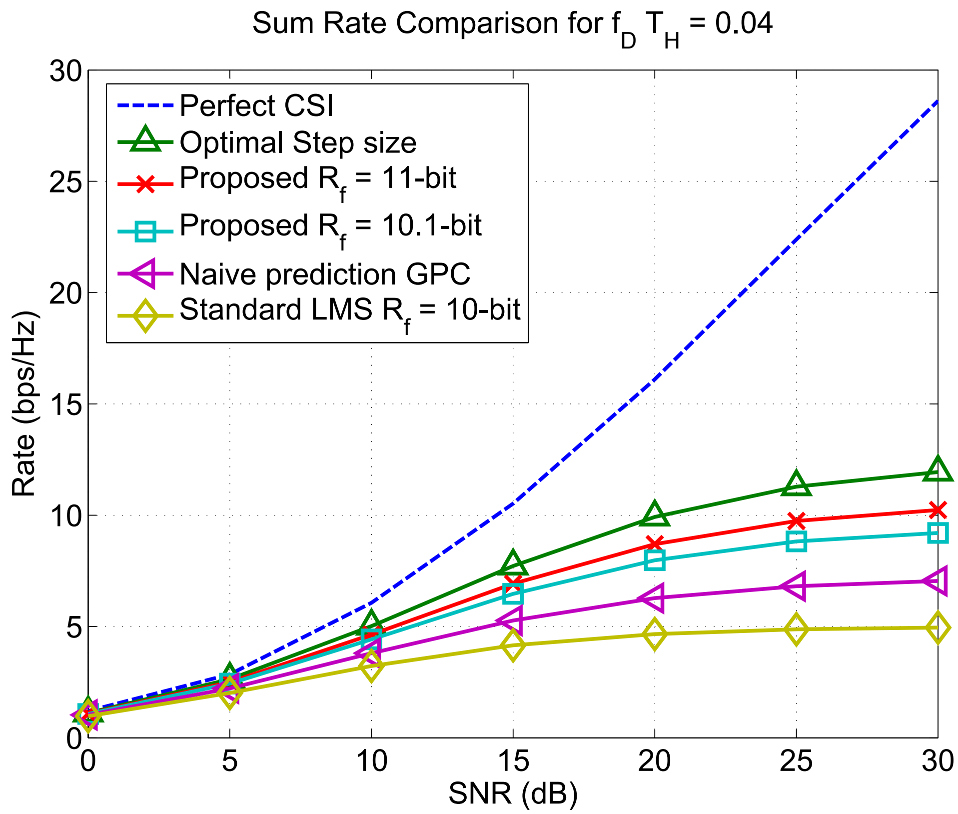

Figure 3.16: Sum rate comparison for $N_t = U = 4$ and $f_D T_H = 0.04$. Plots are shown for perfect CSI, step size optimized Grassmannian predictive coding (GPC), proposed adaptive step size GPC with step size feedback every $4T_H$ and $40T_H$, naive one step GPC using $\tilde{\mathbf{x}}$ as output, and first order linear prediction.

provides even greater sum rate improvement across SNR and normalized Doppler. The step size optimized strategy with $R_f = 14$ bits yields the highest sum rate showing the upper bound on the achievable sum rate using the proposed prediction framework. Two alternative step size feedback strategies are also shown. The case for $R_f = 11$ bits, marked by $\times$, illustrates the case where step size is fed back every $4T_H$. The case for $R_f = 10.1$ bits, marked by $\square$, illustrates the case where

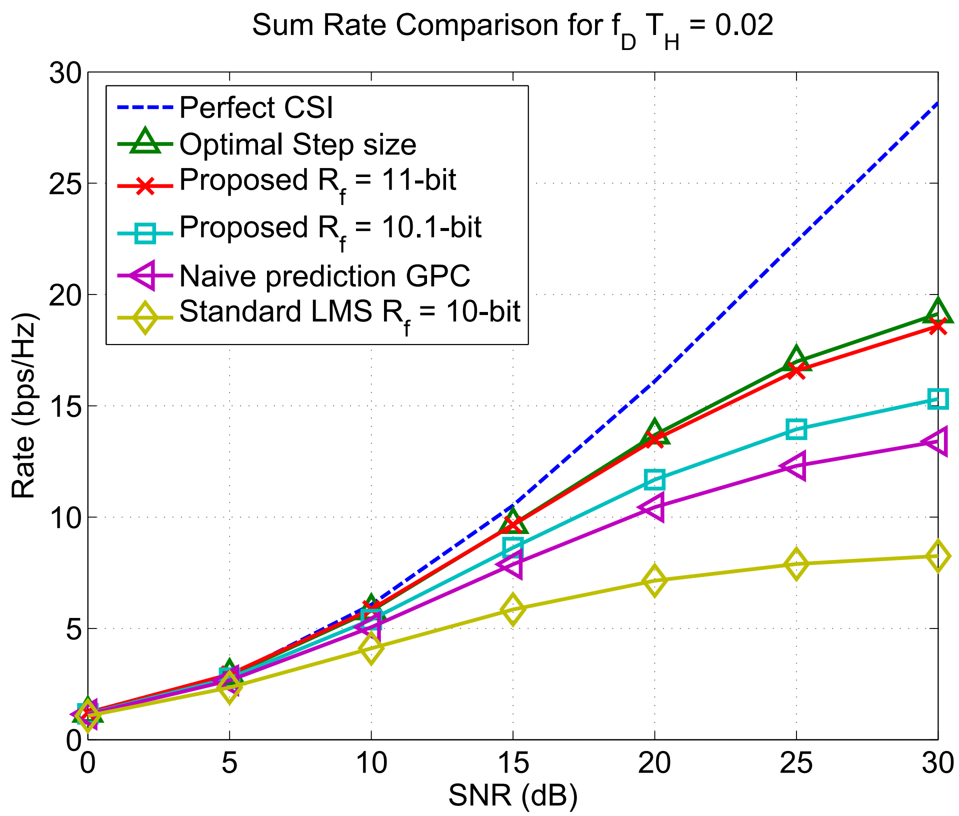

Figure 3.17: Sum rate comparison for $N_t = U = 4$ and $f_D T_H = 0.02$. Plots are shown for perfect CSI, step size optimized Grassmannian predictive coding (GPC), proposed adaptive step size GPC with step size feedback every $4T_H$ and $40T_H$, naive one step GPC using $\tilde{\mathrm{x}}$ as output, and first order linear prediction.

the step size is fed back every $40T_H$. Despite infrequent feedback of the step size, the achievable sum rate shows the importance of step size adaptation. Therefore, these numerical results illustrate the effectiveness of the proposed adaptive step size prediction formulation and that significant sum rate gain is achievable in the presence of delayed limited feedback.

Next, the sum rate performance for the proposed adaptive step size GPC using different codebook sizes (16, 14, 12, and 10 bits) and the conventional memoryless approach using 10 bits are compared in Fig. 3.18. The plot illustrates the dependence of the performance on codebook size. Reducing the codebook size has significant effect on the sum rate performance due to coarse representation of the tangent error vector. Fortunately, at 10 bits of feedback, the proposed adaptive step size GPC still outperforms the conventional memoryless approach.

Finally, mode switching point for single user MIMO and multiuser MIMO with the proposed predictive coding in delayed feedback system is considered. In Fig. 3.19, the achievable throughputs are compared for single user MIMO system using $N_t = 4$ beamforming and multiuser MIMO system with $N_t = 4$ and 4 users using zero forcing precoding. Carrier frequency is assumed to be 2GHz with $T_H = 5$ms. The proposed adaptive step size GPC algorithm is used for both single user and multiuser mode with 16 bits of feedback. The lower SNR crossing point remains at approximately 10dB while the higher SNR switching point increases from 35.4dB to 71.4dB for slower mobile speed. Fig. 3.19 illustrates that the range of SNR where multiuser MIMO gives higher throughput can be significantly extended compared to memoryless techniques used in [103].

3.7 Summary

In this chapter, a new predictive coding algorithm on the Grassmann manifold was proposed generalizing the classical predictive vector quantization. In addition, adaptive step size Grassmannian predictive coding was proposed for MIMO systems with feedback delay. The geometric properties of the Grassmann manifold

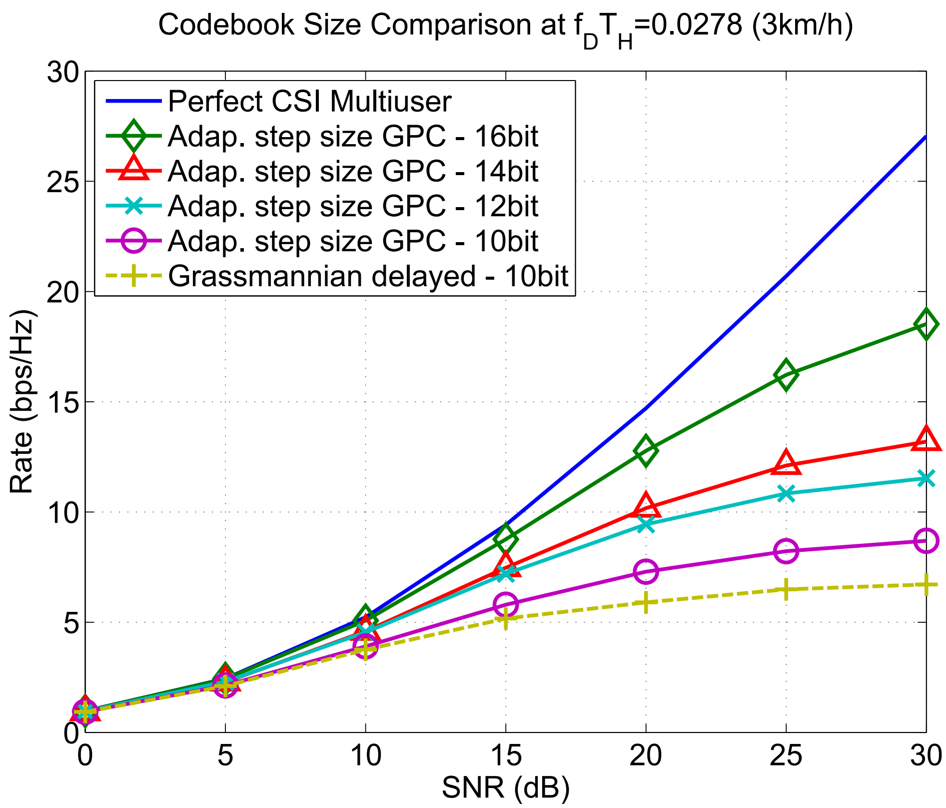

Figure 3.18: Sum rate comparison for $N_t = U = 4$ at $f_D T_H = 0.0278$ (3km/h) for the proposed adaptive step size GPC using codebook sizes 16, 14, 12, and 10 bits and conventional memoryless codebook approach using 10 bits. The plot shows that using 10-bit codebook with the proposed algorithm still provides sum rate improvement.

were exploited to derive a prediction function, an error tangent vector, a quantizer, and a step size optimization framework. Distortion bounds were obtained showing significant distortion improvement over memoryless quantization techniques. Two immediate applications in limited feedback beamforming system and limited feedback multiuser MIMO system were simulated. In particular, the proposed GPC algorithms were shown to provide significant sum rate improvement, even under

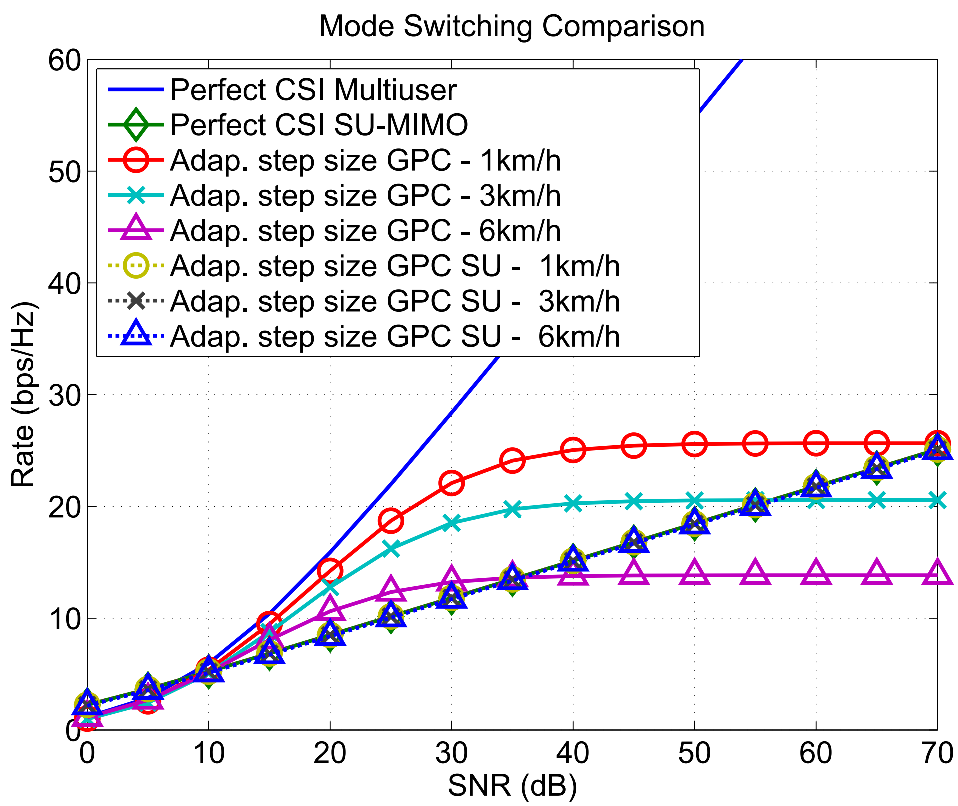

Figure 3.19: Single user and multiuser sum rate comparison for $N_t = U = 4$ with $f_D T_H = 0.0556$ (6km/h), 0.0278 (3km/h), and 0.0092 (1km/h) assuming carrier frequency of 2GHz and $T_H = 5$ms. The plot shows that, at lower SNR region and high SNR region depending on mobile speed, single user (SU) MIMO mode can outperform multiuser MIMO system.

feedback delay, for multiuser MIMO system using practical codebook sizes.

3.8 Appendix

3.8.1 Proof of Theorem 6

Proof. It was shown in [44] that the tangent vector between $\mathbf{x}[1]$ and $\mathbf{x}[2]$ in $\mathcal{G}_{n,1}$ can be written as

$$\mathbf{e} = \tan^{-1}\left(\left\|\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\right\|\right)\frac{\mathbf{x}[2]/\rho - \mathbf{x}[1]}{\|\mathbf{x}[2]/\rho - \mathbf{x}[1]\|_2}. \tag{3.48}$$

The normed term can be simplified as

$$\begin{aligned}\left\|\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\right\|_2^2 &= \left(\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\right)^*\left(\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\right) \\ &= \frac{1}{|\rho|^2} - 1. \end{aligned} \tag{3.49}$$

Therefore,

$$\left\|\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\right\|_2 = \sqrt{\frac{1}{|\rho|^2} - 1} = \frac{d}{|\rho|}$$

where $d = \sqrt{1-|\rho|^2}$ is the chordal distance between $\mathbf{x}[1]$ and $\mathbf{x}[2]$. Clearly, $\|\mathbf{e}\|_2 = \tan^{-1}(d/\|\rho|) \geq 0$ and $\vec{\mathbf{e}} = (\mathbf{x}[2]/\rho - \mathbf{x}[1])/(d/|\rho|)$ such that $\mathbf{e} = \|\mathbf{e}\|_2\vec{\mathbf{e}}$.

Using the exponential form of trigonometric identities

$$\tan^{-1}(x) = (j/2)\ln\{(1-jx)/(1+jx)\}$$

and

$$\cos^{-1}(x) = -j\ln(x + \sqrt{x^2-1})$$

, we have

$$\begin{aligned}
\tan^{-1}\left(\frac{d}{|\rho|}\right) &= \frac{j}{2}\ln\left(\frac{1-j\left(\frac{d}{|\rho|^2}\right)}{1+j\left(\frac{d}{|\rho|^2}\right)}\right) \\
&= -j\ln(|\rho|+\sqrt{|\rho|^2-1}) \\
&= \cos^{-1}|\rho|. \qquad (3.50)
\end{aligned}$$

Since $|\rho|$ is the cosine of the subspace angle between $\mathbf{x}[1]$ and $\mathbf{x}[2]$, this shows that the norm of the tangent vector is equal to the arc length, i.e., $|\theta|$ with subspace angle θ [25, p. 603]. $\square$

3.8.2 Proof of Theorem 7

Proof. For the general case where $\mathbf{x}[1],\ \mathbf{x}[2] \in \mathcal{G}_{n,p}$ with $n > p > 0$, the geodesic between $\mathbf{x}[1]$ and $\mathbf{x}[2]$ was shown to be [18]

$$\mathbf{X}(t) \quad = \quad \mathbf{x}[1]\mathbf{V}\cos(\Sigma t)\mathbf{V}^* + \mathbf{U}\sin(\Sigma t)\mathbf{V}^*$$

where $\mathbf{U}\Sigma\mathbf{V}^*$ is the compact singular value decomposition of the tangent emanating from $\mathbf{x}[1]$ to $\mathbf{x}[2]$. For the case $\mathbf{x}[1],\ \mathbf{x}[2] \in \mathcal{G}_{n,1}$, let $\mathbf{e}$ be the tangent vector emanating from $\mathbf{x}[1]$ to $\mathbf{x}[2]$. Then, we may assume $\mathbf{V} = 1$ without loss of generality and identify $\mathbf{U}$ with $\vec{\mathbf{e}}$ and Σ with $\|\mathbf{e}\|_2$ to obtain

$$G(\mathbf{x}[1], \mathbf{e}, t) = \mathbf{x}[1]\cos(\|\mathbf{e}\|_2 t) + \vec{\mathbf{e}}\sin(\|\mathbf{e}\|_2 t). \qquad (3.51)$$

It is clear that $G(\mathbf{x}[1], \mathbf{e}, 0) = \mathbf{x}[1]$. At $t = 1$, we have

$$
\begin{aligned}
G(\mathbf{x}[1], \mathbf{e}, 1) &= \frac{\mathbf{x}[1]}{\sqrt{1 + d^2/|\rho|^2}} \\
&\quad + \frac{\mathbf{x}[2]/\rho - \mathbf{x}[1]}{d/|\rho|} \frac{d/|\rho|}{\sqrt{1 + d^2/|\rho|^2}} && (3.52) \\
&= \frac{\mathbf{x}[2]}{\rho\sqrt{1 + d^2/|\rho|^2}} && (3.53) \\
&= \mathbf{x}[2]
\end{aligned}
$$

where we have used the identities

$$
\begin{aligned}
\sin(x) &= \frac{x}{\sqrt{1 + x^2}} \\
\cos(x) &= \frac{1}{\sqrt{1 + x^2}} && (3.54)
\end{aligned}
$$

in (3.52) and the fact that $\rho\sqrt{1 + d^2/|\rho|^2} = 1$ in (3.53).

To verify that $G(\mathbf{x}[1], \mathbf{e}, t)$ for $t \in [0, 1]$ is a valid point on the Grassmann manifold, taking the inner product of $G(\mathbf{x}[1], \mathbf{e}, t)$ with itself yields 1 for $t \in [0, 1]$ by using the fact that $\mathbf{x}[1] \perp \mathbf{e}$. $\square$

3.8.3 Proof of Theorem 8

Proof. For the general case where $\mathbf{x}[1],\ \mathbf{x}[2] \in \mathcal{G}_{n,p}$, $n > p > 0$, the parallel transport of tangent $\mathbf{E}$ emanating from $\mathbf{x}[1]$ along the geodesic direction Δ with compact singular value decomposition, $\mathbf{U}\Sigma\mathbf{V}^*$, was shown to be [18]

$$\hat{\mathbf{E}} = [-\mathbf{x}[1]\mathbf{V}\sin(\Sigma t)\mathbf{U}^* + \mathbf{U}\cos(\Sigma t)\mathbf{U}^* + (\mathbf{I} - \mathbf{U}\mathbf{U}^*)]\,\mathbf{E}. \quad (3.55)$$

It needs to be shown the parallel transport of the tangent vector $\mathbf{e}$ emanating from $\mathbf{x}[1]$ to $\mathbf{x}[2]$ in the geodesic direction $\mathbf{e}$ for the case $\mathbf{x}[1],\ \mathbf{x}[2] \in \mathcal{G}_{n,1}$. Without loss of generality, we may assume that the singular value decomposition of $\mathbf{e}$ is given with $\vec{\mathbf{e}}$ as the left singular vector, $\|\mathbf{e}\|_2$ as the singular value, and 1 for the right singular vector. Then

$$\begin{aligned}\vec{\mathbf{e}}(t) &= \left[-\mathbf{x}[1]\vec{\mathbf{e}}^* \sin(\|\mathbf{e}\|_2 t) + \vec{\mathbf{e}}\vec{\mathbf{e}}^* \cos(\|\mathbf{e}\|_2 t) + (\mathbf{I} - \vec{\mathbf{e}}\vec{\mathbf{e}}^*)\right]\mathbf{e} \\ &= -\mathbf{x}[1]\|\mathbf{e}\|_2 \sin(\|\mathbf{e}\|_2 t) + \mathbf{e}\cos(\|\mathbf{e}\|_2 t). \end{aligned} \tag{3.56}$$

Since $G(\mathbf{x}[1], \mathbf{e}, 1) = \mathbf{x}[2]$, the parallel transported tangent vector emanating from $\mathbf{x}[2]$ is found by evaluating (3.56) for $t = 1$. Using (4.12) and (3.54), we have

$$\begin{aligned}\hat{\mathbf{e}} &= -\mathbf{x}[1]\|\mathbf{e}\|_2 \sin(\|\mathbf{e}\|_2) + \mathbf{e}\cos(\|\mathbf{e}\|_2) \\ &= \frac{-\mathbf{x}[1]\tan^{-1}(d/|\rho|)(d/|\rho|)}{\sqrt{1 + d^2/|\rho|^2}} \\ &\quad + \frac{\tan^{-1}(d/|\rho|)(\mathbf{x}[2]/\rho - \mathbf{x}[1])}{d/|\rho|}\frac{1}{\sqrt{1 + d^2/|\rho|^2}} \\ &= \frac{\tan^{-1}(d/|\rho|)}{(d/|\rho|)\sqrt{1 + d^2/|\rho|^2}}\left(\frac{\mathbf{x}[2]}{\rho} - \mathbf{x}[1]\left(1 + \frac{d^2}{|\rho|^2}\right)\right) \\ &= \tan^{-1}\left(\frac{d}{|\rho|}\right)\frac{\mathbf{x}[2]\rho^* - \mathbf{x}[1]}{d} \end{aligned} \tag{3.57}$$

which is the desired result. $\square$

3.8.4 Proof of Proposition 9

Proof. The predicted vector is obtained as

$$\begin{aligned}\tilde{\mathbf{x}}[k+1] &= G(\mathbf{x}[k], \hat{\mathbf{e}}, t) \\ &= \mathbf{x}[k]\cos(\|\hat{\mathbf{e}}\|t) + \frac{\hat{\mathbf{e}}}{\|\hat{\mathbf{e}}\|}\sin(\|\hat{\mathbf{e}}\|t)\end{aligned}$$

Since

$$\frac{\hat{\mathbf{e}}}{\|\hat{\mathbf{e}}\|} = \frac{\mathbf{x}[k]\rho^* - \mathbf{x}[k-1]}{d} \tag{3.58}$$

with $\rho = \mathbf{x}^*[k-1]\mathbf{x}[k]$ and $d = \sqrt{1-|\rho|^2}$ by (3.4), we obtain the desired formula.

$\square$

3.8.5 Proof of Theorem 10

Proof. Recall that the parallel transported tangent vector $\hat{\mathbf{e}}$ emanating from $\mathbf{x}[2]$ is given in (3.4). Computing the geodesic with from $\mathbf{x}[2]$ along $\hat{\mathbf{e}}$ at $t = 1$ gives

$$\begin{aligned}\hat{\mathbf{x}} &= G(\mathbf{x}[2], \hat{\mathbf{e}}, 1) \\ &= \frac{\mathbf{x}[2]}{\sqrt{1+d^2/|\rho|^2}} + \frac{\mathbf{x}[2]\rho^* - \mathbf{x}[1]}{\sqrt{1+d^2/|\rho|^2}} \\ &= |\rho|\mathbf{x}[2] + \rho^*\mathbf{x}[2] - \mathbf{x}[1]. \end{aligned} \tag{3.59}$$

To see that $\hat{\mathbf{x}} \in \mathcal{G}_{n,1}$, we have

$$\begin{aligned}\hat{\mathbf{x}}^*\hat{\mathbf{x}} &= (|\rho|\mathbf{x}[2] + \rho^*\mathbf{x}[2] - \mathbf{x}[1])^*(|\rho|\mathbf{x}[2] + \rho^*\mathbf{x}[2] - \mathbf{x}[1]) \\ &= 1 \end{aligned} \tag{3.60}$$

where we have used the fact that $\rho = \mathbf{x}^*[1]\mathbf{x}[2]$. To see that the prediction is distance preserving, the inner product of $\mathbf{x}[2]$ and $\hat{\mathbf{x}}$ gives

$$\begin{aligned}\mathbf{x}^*[2]\hat{\mathbf{x}} &= \mathbf{x}^*[2]\mathbf{x}[2]|\rho| + \mathbf{x}^*[2]\mathbf{x}[2]\rho^* - \mathbf{x}^*[2]\mathbf{x}[1] \\ &= |\rho|. \end{aligned} \tag{3.61}$$

Therefore,

$$d(\mathbf{x}[2], \hat{\mathbf{x}}) = \sqrt{1-|\rho|^2} = d(\mathbf{x}[1], \mathbf{x}[2]). \tag{3.62}$$

$\square$

Chapter 4

Signal Processing on the Manifold of Unitary Matrices

4.1 Prior Work

Differential and predictive coding play a prominent role in areas such as speech, image, and video coding as well as in various applications of data quantization and compression [17, 19, 22, 55]. Differential coding exploits the correlation among the data, or equivalently, the memory in the underlying process in which the data arise. Instead of independently encoding sample by sample, encoding the *difference* between samples often reduces the dynamic range for quantization, hence improving the accuracy of the coded information for a given number of bits. The exhibited correlation may be due to speech production mechanisms (temporal), objects in an image with similar adjacent pixel colors (spatial), or moving objects in a video (spatio-temporal). Similarly, predictive coding also exploits correlation in the data by encoding the difference between the observed and the predicted data. Linear predictive coding techniques have been successfully used to obtain high resolution source coding in speech, image and video applications. Both differential and predictive coding are well understood for signals represented in linear vector space [50, 100]. Unfortunately, differential and predictive coding of time series evolving on the space of unitary matrices have not appeared before.

For MIMO applications, most prior research on limited feedback designs have focused on the quantization of *fixed rank* channel state information, hence the focus on Grassmann manifold of p-dimensional subspaces in n-dimensional space [66]. Chapter 3 dealt with the predictive coding of special case of Grassmann manifold consisting of 1-dimensional subspaces. Prior work was discussed in Section 3.1. The rank of the precoder defines the number of spatial streams transmitted simultaneously in spatial multiplexing systems with beamforming being the special case with only one stream of data being sent. This approach has led to numerous codebook design as discussed in previous chapters. In particular, a differential feedback approaches were proposed in [4, 58]. The idea there was to consider the change in the precoder matrix over time as a unitary transformation using the group theoretic approach. In both cases, codebook of unitary matrices that represent incremental transformation, or rotation, of the precoder were proposed. While this approach is attractive for its simple implementation, the design of unitary matrix codebook is challenging due to the large number of variables even for moderate number of antennas. Furthermore, the rank of the precoder is fixed and it is smaller than the unitary matrices representing the change resulting in more parameters than the precoders themselves. Unfortunately, fixing the rank of the precoder limits the available spatial multiplexing gain thus resulting in limited throughput. In fact, it has been show that an adaptive technique, called multimode precoding, which adapts the rank of the precoder based on tradeoff between throughput and error rate is capacity achieving [65]. Furthermore, the cellular standard such as 3GPP has adopted a nested limited feedback codebook structure such that the transmitter can override the rank and reduce the rate for higher reliability [2]. This motivates a new avenue of limited feedback approach that is independent of the rank. Fortunately, the optimal unitary precoder was shown to be the p dominant right singular vectors of the channel matrix [64, 94]. Thus, we see that the unitary matrix of right singular vectors of a given channel contains the necessary spatial information for every rank unitary precoding.

4.2 Contributions

In this chapter, a differential and predictive coding algorithms are proposed for correlated data on the manifold of unitary matrices that arise from the right singular matrix of the MIMO channel. The proposed algorithms are motivated by a new avenue of limited feedback approach that is rank independent for higher flexibility at the transmitter and the need for higher resolution feedback strategy. The proposed algorithms are derived using group and differential geometric properties of Lie group of special unitary matrices. The main contributions of this paper are as follows.

- *Differential and prediction framework for unitary matrix time series:* A new differential and prediction technique are proposed for time series of unitary matrices. A transformation of unitary matrix into an equivalent special unitary matrix under the signal to noise ratio metric is used. The key idea is to use the Lie group structure of the special unitary matrix to derive a tangent space difference between two points. The group symmetry of the Lie group is exploited to introduce a notion of one step prediction. Furthermore, a closed form step size optimization for prediction is obtained to minimize the tangent space error between the predicted and observed unitary matrix. An adaptive step size prediction filter is proposed to eliminate the dependence of step size optimization on the current observation for practicality.

- *Differential coding of unitary matrix time series:* A framework for differential coding of correlated unitary matrix time series is proposed. The key idea is to transform the unitary matrix into a special unitary matrix and compute the tangential error between two successive observations. A strategy for quantizing the tangential error is proposed by exploiting the skew-Hermitian symmetry of the tangential error.

- *Predictive coding of unitary matrix time series:* Using the adaptive step size predictor, a predictive coding strategy is proposed. The main idea is to encode the tangential difference between the predicted matrix and the observed matrix exploiting the group and differential geometric structure of the Lie group of special unitary matrices. Simulation results for mean square error performance of the proposed algorithms are shown.

- *Application to limited feedback MIMO Systems:* The proposed differential and predictive coding algorithms are applied to single user limited feedback MIMO systems. The proposed differential and predictive coding algorithms are used to encode and feedback the right singular matrix of the MIMO channel. Based on the feedback information, achievable throughputs and symbol error rates for rank 1 (beamforming) and 2 (spatial multiplexing) strategies are compared with memoryless limited feedback approaches. The proposed approaches are shown to yield throughput and error rate performance close to perfect CSI case that is independent of the chosen rank. Furthermore, using the predicted output at the transmitter, the value of predictive coding strategy for systems with feedback delay is illustrated.

4.3 System Overview

4.3.1 Discrete-time System Model

For single user limited feedback unitary precoded MIMO wireless system, the system model in Section 2.3 is considered.

A limited feedback-based block diagonalization for multiuser MIMO wireless system is also considered. The base station is assumed to have N_t transmit antennas and there are $U = N_t$ mobile users each equipped with N_r antennas. To isolate the impact of predictive coding for limited feedback, U users are assumed to be scheduled from possibly large number of user pool and the problem of user scheduling is not considered in this dissertation. The transmitter sends $N_s = N_r$ streams of data to each user. Although $N_r \geq N_s$ is possible with appropriate combiner at the receiver, the design of combiner is not considered in this dissertation. Let $\mathbf{s}^{(u)}[k]$, $\mathbf{F}^{(u)}[k]$, and $\mathbf{H}^{(u)}[k]$ be the $N_s \times 1$ transmit vector, $N_t \times N_s$ precoding vector satisfying $\mathbf{F}^{(u)*}[k]\mathbf{F}^{(u)}[k] = \mathbf{I}_{N_s}$, and $N_t \times N_r$ channel matrix for u-th user at time index k, respectively. Then, the input-output relationship for u-th user may be written as

$$\mathbf{y}^{(u)}[k] = \mathbf{H}^{(u)*}[k]\mathbf{F}^{(u)}[k]\mathbf{s}^{(u)}[k] + \mathbf{H}^{(u)*}[k] \sum_{n=1, n\neq u}^{U} \mathbf{F}^{(n)}[k]\mathbf{s}^{(n)}[k] + \mathbf{n}^{(u)}[k] \quad (4.1)$$

where $\mathbf{n}[k]$ is an independent complex Gaussian noise vector with unit variance. The first term in (4.1) is the desired signal for u-th user while the second summation term is the interference signal. The signal to interference plus noise ratio (SINR) for the u-th user can be written as

$$\text{SINR}^{(u)} = \frac{\frac{P}{N_t}\|\mathbf{H}^{(u)*}\mathbf{F}^{(u)}\|_2^2}{1 + \sum_{n\neq u} \frac{P}{N_t}\|\mathbf{H}^{(u)*}\mathbf{F}^{(n)}\|_2^2}. \quad (4.2)$$

If the transmit signal $\mathbf{s}^{(u)}$ is assumed to be Gaussian, the achievable rate for user u is given by

$$\mathcal{R}^{(u)} = \log_2(1 + \text{SINR}^{(u)}) \quad (4.3)$$

and the sum rate as $\mathcal{R} = \sum_{u=1}^{U} \mathcal{R}^{(u)}$.

The block diagonalization strategy involves linear precoding that suppress inter user interference when perfect CSI is available. Following the block diagonalization procedure [101], each $\mathbf{F}^{(u)}$ is chosen such that $\mathbf{H}^{(n)*}\mathbf{F}^{(u)} = \mathbf{0}$ for $n \neq u$. The precoding matrices can be designed by stacking the channel matrices $\mathbf{H}^{(u*)}$ for $u \neq n$ and finding the basis corresponding to the null space of the stacked channel matrix. For limited feedback-based block diagonalization, quantized version of the subspace spanned by the rows of $\mathbf{H}^{(u)*}$ is used [86]. In this dissertation, limited feedback of right singular matrix associated with $\mathbf{H}^{(u)}$ is considered.

Dropping the user index for notational brevity, the singular value decomposition of the channel matrix at time k is given by

$$\mathbf{H}[k] = \mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{V}^*[k] \tag{4.4}$$

where $\mathbf{U}[k]$ is the $N_r \times N_r$ unitary matrix of left singular vectors, $\mathbf{\Sigma}[k]$ is the $N_r \times N_t$ diagonal matrix of singular values, and $\mathbf{V}[k]$ is the $N_t \times N_t$ unitary matrix of right singular vectors. Note that the $\mathbf{V}[k]$ is a square unitary matrix that only depends on the number of transmit antennas. It has been shown that for N_s-stream transmission with $N_s \leq \min\{N_t, N_r\}$ that the optimal unitary precoder is given by the N_s dominant columns of $\mathbf{V}[k]$ corresponding to the N_s largest singular values [64, 94]. Therefore, the unitary matrix $\mathbf{V}[k]$ contains the spatial dimensions available in a given channel.

To develop a feedback strategy for $\mathbf{V}[k]$, a unitary matrix decomposition from group theory is used. A unitary matrix $\mathbf{V}$ can be represented as a direct product of elements in special unitary group and unitary group of dimension 1 [57]. Let $\mathcal{U}_n = \{\mathbf{X} \in \mathbb{C}^{n\times n} : \mathbf{X}^*\mathbf{X} = \mathbf{I}_n\}$ denote the space of $n \times n$ unitary matrices and $\mathcal{SU}_n = \{\mathbf{X} \in \mathcal{U}_n : \det(\mathbf{X}) = 1\}$ denote the space of special unitary matrices. Then, $\mathbf{V}[k] \in \mathcal{U}$ can be decomposed as $\mathbf{V}[k] = \mathbf{S}[k]\mathbf{Q}[k]$ where $\mathbf{S}[k] \in \mathcal{SU}_n$ and $\mathbf{Q}[k] = \text{diag}\begin{bmatrix}\det(\mathbf{V}[k]) & 1 & \cdots & 1\end{bmatrix}$. Thus any unitary matrix can be transformed into a special unitary matrix by right multiplication $\mathbf{Q}^*[k]$. Since $\mathbf{Q}[k]$ is unitary and the performance metrics makes it invariant to unitary transformations, $\mathbf{S}[k] \in \mathcal{SU}_n$ is considered as the equivalent channel state information to be fed back to the transmitter. As it will be shown in Section 4.4, there is an advantage in using the differential geometric structure of $\mathcal{SU}_n$ as opposed to $\mathcal{U}_n$. Based on this, the limited feedback of derived channel state information $\mathbf{S}[k] \in \mathcal{SU}_n$ which is equivalent under the performance metric to feedback of $\mathbf{V}[k] \in \mathcal{U}_n$ is considered.

4.3.2 Performance Metrics

Two types of performance metrics will be used to evaluate the proposed algorithms. The first type of performance metric is based on signal processing theoretic performance measures. Specifically, two distance measures are used to evaluate the distance between the optimal precoder and the precoder obtained by the proposed algorithms. The first of this type is the chordal distance metric (3.1). The chordal distance measure is used for rank 1 limited feedback or limited feedback beamforming systems. For rank 2 and higher, generalization of chordal distance to $N_t \times N_s$ unitary matrices with $N_s < N_t$ is used. The projection Frobenius norm is

$$\begin{aligned} d_{pF}(\mathbf{F}_1, \mathbf{F}_2) &= 2^{-1/2} \|\mathbf{F}_1\mathbf{F}_1^* - \mathbf{F}_2\mathbf{F}_2^*\|_F \\ &= \|\sin\theta\|_2. \end{aligned} \tag{4.5}$$

Both chordal distance and projection Frobenius norm are used to evaluate mean squared error performance and per sample errors over time.

The second type of performance metric is based on communication theoretic performance measures; achievable throughput, symbol error rate, and *feedback rate*. The feedback rate is defined to be the number of feedback bits per update interval. The feedback rate is used to compare the required feedback throughput for the proposed algorithms.

4.4 Lie Groups and Lie Algebras

First, an overview of the Lie group theory is given which provides the group theoretic and differential geometric concepts used to derive the proposed algorithms. Then, some key operators are constructed to perform differencing and projections onto the special unitary matrices. Finally, the notion of prediction on Lie groups of special unitary matrices is developed and prediction optimization strategies are shown.

4.4.1 Preliminaries

Recall that a matrix group structure is a set endowed with the usual matrix multiplication, inverse operator, and the identity element satisfying the closure and associativity [57]. An $n \times n$ unitary group, $\mathcal{U}_n$, is a set of matrices satisfying $\mathbf{U}^*\mathbf{U} = \mathbf{I}_n$ for $\mathbf{U} \in \mathbb{C}^{n\times n}$. An $n \times n$ special unitary group, denoted $\mathcal{SU}_n$, is a subset of $\mathcal{U}_n$ with condition that $\det(\mathbf{U}) = 1$ for $\mathbf{U} \in \mathcal{U}_n$. For the special unitary group, the inverse is $\mathbf{X}^{-1} = \mathbf{X}^*$ and the usual identity matrix is given by $\mathbf{I}_n$. The right translation of a point $\mathbf{Y} \in \mathcal{SU}_n$ about an element $\mathbf{X} \in \mathcal{SU}_n$, $R_{\mathbf{X}} : \mathcal{SU} \to \mathcal{SU}$, can be defined as

$$R_{\mathbf{X}}(\mathbf{Y}) = \mathbf{Y}\mathbf{X}^*. \tag{4.6}$$

The inverse right translation operator $\ell_X : \mathcal{SU} \to \mathcal{SU}$ is defined as

$$R_{\mathbf{X}}^{-1}(\mathbf{Y}) = \mathbf{Y}\mathbf{X}. \tag{4.7}$$

The right translation and inverse can be used to translate two points, say $\mathbf{X}, \mathbf{Y} \in \mathcal{SU}_n$, such that $\mathbf{X}$ is translated to the identity by $R_{\mathbf{X}}(\mathbf{X})$ and $\mathbf{Y}$ is translated to another point, say $\mathbf{Z} = R_{\mathbf{X}}(\mathbf{Y})$. Thus the translation operation enables us to consider the relationship between points with respect to the identity rather than considering absolute local relationships. A left translation can be similarly defined and used to follow the developments to follow but only the right translation will be used in this dissertation for conciseness.

In addition to the usual group structure, a Lie group is a group that also has a smooth manifold structure compatible with the underlying group structure. Formally, a Lie group is defined as follows [27].

Definition 12 (Lie Group)**.** *A Lie group G is a differentiable manifold which is also a group and such that the group product and the inverse map are differentiable.*

A differentiable, or smooth, manifold is a space that is locally Euclidean and on which calculus can be performed [61]. The unitary group and the special unitary group are well known matrix Lie groups[1]. In fact, smooth manifolds such as Stiefel and Grassmann manifolds that frequently arise in recent literatures on limited feedback MIMO systems [8,48,64,76] and optimization [6,18,71] are cases of special Lie group with quotient group structure [89].

Since Lie group is a differentiable manifold, a tangent space can be defined at the identity, denoted $\mathfrak{su}_n$ with lower case Gothic letters, of all matrices $\mathbf{E}$ for which there is a differentiable function $f(\tau)$, $\tau \in \mathbb{R}$, that lies in $\mathcal{SU}_n$ and satisfies

[1] All matrix Lie groups are Lie groups, but the converse is not true [27]. The terminology between the matrix Lie group and the Lie group is not distinguished in this dissertation since the discussion is limited to matrix groups.

$f(0) = \mathbf{I}_n$ and $f'(0) = \mathbf{E}$ where $'$ denotes the derivative of the function with respect to τ. The tangent space at the identity of a Lie group is known as Lie algebra which is a real vector space. For $\mathcal{SU}_n$, the corresponding Lie algebra is defined as $\mathfrak{su}_n = \{\mathbf{E} \in \mathbb{C}^{n\times n} : \mathbf{E}^* = -\mathbf{E} \quad \text{and} \quad \text{trace}(\mathbf{E}) = 0\}$ [27]. Therefore, the Lie group of $\mathcal{SU}_n$ has a tangent space of real vector space consisting of skew-Hermitian matrices with trace zero. Furthermore, the tangent space at various points on the group can be characterized at the identity using the translation operator on the group. This is the key distinguishing feature from tangent space of Grassmann manifold [48] where the tangent space at each point on the Grassmann manifold is different.

The exponential map plays a fundamental role in passing the information from the Lie algebra to the Lie group [89].

Definition 13. *For an $n \times n$ matrix $\mathbf{E}$, define the matrix exponential $e^{\mathbf{E}}$ by*

$$e^{\mathbf{E}} = \sum_{m=0}^{\infty} \frac{\mathbf{E}^m}{m!}. \tag{4.8}$$

Letting $f(\tau) = e^{\tau\mathbf{E}}$, it is easy to verify that f satisfies the initial condition and the derivative constraint. For the Lie group of $\mathcal{SU}_n$, $e^{\mathbf{E}}$ must be unitary with determinant one. Therefore, $\mathbf{E}$ must satisfy $(e^{\mathbf{E}})^* = (e^{\mathbf{E}})^{-1}$ for every $\mathbf{E}$ by the unitary condition. By taking a term by term expansion of (4.8), the matrix $\mathbf{E}$ must satisfy the skew-Hermitian property, i.e., $\mathbf{E}^* = -\mathbf{E}$, for e^E to be in $\mathcal{SU}_n$. Furthermore, since matrices in $\mathcal{SU}_n$ have determinant one, $\det(e^{\mathbf{E}}) = e^{\text{trace}(\mathbf{E})}$ yields the property that trace($\mathbf{E}$) $= 0$. For $\tau \in [0,1]$, $f(\tau)$ gives the so called one-parameter map of geodesic path, the shortest distance path on $\mathcal{SU}_n$, between two points given by $\mathbf{I}_n$ and $e^{\mathbf{E}}$, both in $\mathcal{SU}_n$.

Given a point near the identity, the difference is computed by find the tangent matrix $\mathbf{E}$ representing the differential between the identity and a point, say, $\mathbf{S} \in \mathcal{SU}_n$. This is accomplished by the matrix logarithm that lifts a point onto the tangent space [27].

Definition 14. *For an $n \times n$ special unitary matrix $\mathbf{S}$, define the matrix logarithm $\log(\mathbf{S})$ by*

$$\log(\mathbf{S}) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(\mathbf{S} - \mathbf{I}_N)^n}{n}. \tag{4.9}$$

The direct form of matrix exponential and logarithm are not computationally attractive. For the special unitary group, however, simplified computation may be obtained by noting that any special unitary matrix $\mathbf{S}$ is diagonalizable as $\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ where $\mathbf{U}$ is another unitary matrix and $\mathbf{D} = \mathrm{d}iag \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots \lambda_N \end{bmatrix}$ is the diagonal matrix with eigenvalues of the form $e^{j\theta_n}$ along the diagonal. Thus, the matrix logarithm is written as

$$\mathbf{E} = \log(\mathbf{S}) = \mathbf{U} \begin{bmatrix} \log(\lambda_1) & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \log(\lambda_n) \end{bmatrix} \mathbf{U}^* \tag{4.10}$$

and the matrix exponential as

$$\mathbf{S} = e^{\mathbf{E}} = \mathbf{U} \begin{bmatrix} e^{\log(\lambda_1)} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & e^{\log(\lambda_n)} \end{bmatrix} \mathbf{U}^* \tag{4.11}$$

by expanding the summation terms to factor out $\mathbf{U}$. Computational aspects of approximating matrix exponential in general has been considered in [49].

Based on these tools, the basic operators for the proposed algorithms are developed next. A time series $\mathbf{S}[k]$ evolving on $\mathcal{SU}_n$, e.g., time evolution of CSI encountered in MIMO applications, is considered. Given two successive points, the differential is represented in the tangent space.

Tangent: For $\mathbf{S}[k-1], \mathbf{S}[k] \in \mathcal{SU}_n$, the tangent matrix emanating from $\mathbf{S}[k-1]$ to $\mathbf{S}[k]$ is

$$\mathbf{E}[k] = \log(R_{\mathbf{S}[k-1]}(\mathbf{S}[k])) \tag{4.12}$$

with the property that $\mathbf{E}[k] = -\mathbf{E}^*[k]$ and $\mathrm{trace}(\mathbf{E}[k]) = 0$. For notational brevity, the tangent operation is denoted by $\mathbf{E}[k] = L(\mathbf{S}[k-1], \mathbf{S}[k])$. This operation combines the right translation of $\mathbf{S}[k]$ with respect to $\mathbf{S}[k-1]$ such that $\mathbf{S}[k-1]$ corresponds to the identity element. Using the group structure only, the relationship between $\mathbf{S}[k-1]$ and $\mathbf{S}[k]$ is given by an $n \times n$ unitary transformation involving $\mathbb{C}^{n^2}$ or $\mathbb{R}^{2n^2}$ parameters. Thanks to the skew-Hermitian symmetry, the tangent representation results in $\mathbb{R}^{n^2}$, or half, the number of real parameters. This symmetry is exploited for quantization of the tangent matrices.

The tangent matrix describes the shortest distance path between $\mathbf{S}[k-1]$ and $\mathbf{S}[k]$, called the geodesic. The geodesic can be parameterized by one real parameter t as follows.

Geodesic: If $\mathbf{S}[k-1], \mathbf{S}[k] \in \mathcal{SU}_n$ and $\mathbf{E}[k] \in \mathfrak{su}_n$ is the tangent matrix emanating from $\mathbf{S}[k-1]$ to $\mathbf{S}[k]$, the the geodesic path between $\mathbf{S}[k-1]$ and $\mathbf{S}[k]$ is

$$\begin{aligned} G(\mathbf{S}[k-1], \mathbf{E}[k], t) &= R^{-1}_{\mathbf{S}[k-1]}(e^{\mathbf{E}[k]t}) \\ &= e^{\mathbf{E}[k]t}\mathbf{S}[k-1] \end{aligned} \tag{4.13}$$

for $t \in [0, 1]$ such that $G(\mathbf{S}[k-1], \mathbf{E}[k], 0) = \mathbf{S}[k-1]$ and $G(\mathbf{S}[k-1], \mathbf{E}[k], 1) = \mathbf{S}[k]$. The geodesic formula provides a convenient formula to map the tangent matrix back to $\mathcal{SU}_n$. Thus the tangent operator and the geodesic operator provides the mechanism for mapping between group $\mathcal{SU}_n$ and tangent space $\mathfrak{su}_n$.

Since the tangent space will be exploited to develop the proposed algorithms, a product of geodesic expression is frequently encountered. Let $\mathbf{E}_1$ and $\mathbf{E}_2$ be two tangent matrices. Then, the tangent expression $\mathbf{E}$ which describes the product is

$$\mathbf{E} = \log(e^{\mathbf{E}_1} e^{\mathbf{E}_2}). \tag{4.14}$$

Since the skew-Hermitian matrices in $\mathfrak{su}_n$ do not commute, the solution is given by an infinite series called the *Baker-Campbell-Hausdorff* formula [27]

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \frac{1}{2}[\mathbf{E}_1, \mathbf{E}_2] + \frac{1}{12}[\mathbf{E}_1, [\mathbf{E}_1, \mathbf{E}_2]] - \frac{1}{12}[\mathbf{E}_2, [\mathbf{E}_1, \mathbf{E}_2]] + \cdots \tag{4.15}$$

where $[\cdot, \cdot]$ denotes the *bracket* operator defined as $[\mathbf{E}_1, \mathbf{E}_2] = \mathbf{E}_1\mathbf{E}_2 - \mathbf{E}_2\mathbf{E}_1$ which is a skew-Hermitian preserving operator. In fact, the Lie algebra is usually defined as a real vector space together with the bracket operator which is bilinear, skew symmetric, and satisfies the Jacobi identity [27]. The interpretation of Baker-Campbell-Hausdorff formula is that the tangent matrices cannot simply be added, as one might expect, and that the higher order terms represents the deviation from tangent space additivity. For the proposed algorithms and its analysis, the first order approximation by $\mathbf{E} \approx \mathbf{E}_1 + \mathbf{E}_2$ is shown to give good results. Finally, based on the first order approximation, the Frobenius norm is used to measure the tangent space distance as

$$d(\mathbf{E}_1, \mathbf{E}_2) = \|\mathbf{E}_1 - \mathbf{E}_2\|_F. \tag{4.16}$$

4.4.2 Prediction on Lie Groups

To introduce the notion of *prediction* based on the known tangent information, one method is to extend the geodesic path past $\mathbf{S}[k+1]$ in the direction of $\mathbf{E}$. This is accomplished by the inverse right translation with respect to the new desired base of the tangent. This is in contrast to the parallel transport needed in the Grassmann manifold case in Chapter 3. The reason for this is that Grassmann manifold does not admit a universal tangent space that can be translated by the group operation. Therefore, a different tangent needs to be computed at each point. The case with Lie groups is much simpler thanks to the symmetry of the space. Therefore, a prediction function on $\mathcal{SU}_n$ is proposed.

General Prediction Formula: Let $\mathbf{S}[k-1], \mathbf{S}[k] \in \mathcal{SU}_n$ and $\mathbf{E}[k] \in \mathfrak{su}_n$ be the tangent matrix emanating from $\mathbf{S}[k-1]$ to $\mathbf{S}[k]$. Then, the one step prediction is

$$\begin{aligned} \tilde{\mathbf{S}}[k+1] &= P(\mathbf{S}[k], \mathbf{E}[k], t) \\ &= R^{-1}_{\mathbf{S}[k]}(e^{\mathbf{E}[k]t}) \\ &= e^{\mathbf{E}[k]t}\mathbf{S}[k] \end{aligned} \tag{4.17}$$

where the *step size* parameter $t \in [0, 1]$ can be used to control the prediction step.

The simplest method to perform prediction is to assume the step size parameter $t = 1$. But there is no reason a full step should be taken. In what follows, it is shown that the optimal step size can be learned from the past observations to perform prediction of the step size parameter in such a way to minimize prediction error.

Let $\mathbf{S}[k+1]$ be the matrix which is to be predicted and $\tilde{\mathbf{S}}[k+1]$ be the predicted matrix based on (4.17). The tangent matrix emanating from $\mathbf{S}[k]$ to $\mathbf{S}[k+1]$ is

$$\begin{aligned} \mathbf{E}[k+1] &= L(\mathbf{S}[k], \mathbf{S}[k+1]) \\ &= \log(R_{\mathbf{S}[k]}(\mathbf{S}[k+1])). \end{aligned} \tag{4.18}$$

Then, using the geodesic formula, $\mathbf{S}[k+1]$ can be expressed as

$$\begin{aligned} \mathbf{S}[k+1] &= G(\mathbf{S}[k], \mathbf{E}[k+1], 1) \\ &= e^{\mathbf{E}[k+1]}\mathbf{S}[k]. \end{aligned} \tag{4.19}$$

Then, the prediction error tangent matrix $\mathbf{E}_e[k+1]$ representing the error between $\tilde{\mathbf{S}}[k+1]$ and $\mathbf{S}[k+1]$ is

$$\begin{aligned} \mathbf{E}_e[k+1] &= L(\tilde{\mathbf{S}}[k+1], \mathbf{S}[k+1]) \\ &= \log(\mathbf{S}[k+1]\tilde{\mathbf{S}}^*[k+1]) \\ &= \log(e^{\mathbf{E}[k+1]}\mathbf{S}[k]\mathbf{S}^*[k]e^{-\mathbf{E}[k]t}) \qquad (4.20) \\ &= \log(e^{\mathbf{E}[k+1]}e^{-\mathbf{E}[k]t}) \\ &\approx \mathbf{E}[k+1] - \mathbf{E}[k]t \qquad (4.21) \end{aligned}$$

where the last approximation is due to (4.15). Therefore, the prediction error is minimized by minimizing the norm of $\mathbf{E}_e[k+1]$ as the following theorem shows.

Theorem 15 (Step Size Optimization)**.** *The optimal instantaneous step size selection to predict* $\mathbf{S}[k+1]$ *is*

$$t_{opt} = \arg\min_{t \in [0,1]} \|\mathbf{E}[k+1] - \mathbf{E}[k]t\|_F^2 \tag{4.22}$$

where the closed form solution is

$$t_{opt} = \frac{\Re(tr(\mathbf{E}^*[k+1]\mathbf{E}[k]))}{tr(\mathbf{E}^*[k]\mathbf{E}[k])}. \tag{4.23}$$

Proof. Please see Appendix 4.9.1. □

Unfortunately, from a computational point of view, the step size optimization can only be performed *after* observing $\mathbf{S}[k+1]$. This is not practical for differential and predictive coding because as accurate a prediction as possible of $\mathbf{S}[k+1]$ is needed before $\mathbf{S}[k+1]$ becomes available. To overcome this problem, a least mean square (LMS) based predictor is proposed using step size t based on past history of t_{opt} computed after $\mathbf{S}[k+1]$ becomes available.

Let $t_{\text{opt}}[k]$ denote the optimal step size to predict $\mathbf{S}[k]$. The main idea is to compute $t_{\text{opt}}[k]$ after observing $\mathbf{S}[k]$ and then predict the next step size $\tilde{t}[k+1]$ based on the past knowledge of $t_{\text{opt}}[k], t_{\text{opt}}[k-1], \ldots, t_{\text{opt}}[k-M+1]$ where M denotes the depth of the memory used for prediction of the step size. Then, the step size prediction error is

$$e_t[k+1] = t_{\text{opt}}[k+1] - \tilde{t}[k+1]. \tag{4.24}$$

The mean squared prediction error $\mathbb{E}\left[|e_t|^2\right]$ criterion used to design a linear predictor to minimize the mean squared prediction error. Since $t_{\text{opt}}[k]$ is a scalar process, an M-th order linear predictor

$$\tilde{t}[k+1] = \sum_{n=0}^{M-1} a_n t_{\text{opt}}[k-n] \tag{4.25}$$

with filter coefficients $\{a_n\}_{n=0}^{M-1}$ is used. If $\mathbf{R}_t$ is an $M \times M$ autocorrelation matrix for $t_{\text{opt}}[k]$, $\mathbf{a} = \begin{bmatrix} a_0 & \cdots a_{M-1} \end{bmatrix}^T$ is a vector of filter coefficients, and $\mathbf{r}$ is an $M \times 1$ vector of cross correlation between the desired optimized step size t_{opt} and the predicted step size $\tilde{t}$, the Wiener-Hopf equation is [31]

$$\mathbf{R}_t \mathbf{a} = \mathbf{r} \tag{4.26}$$

and the optimal filter coefficients can be found by computing $\mathbf{R}_t^{-1}\mathbf{r}$. Since sufficient sample history of $t_{\text{opt}}[k]$ may not be available to compute the correlation matrices, the LMS algorithm is used to adapt the filter coefficients $\{a_n\}_{n=0}^{M-1}$ based on instantaneous correlation estimates [31]. The pseudo code is shown in Algorithm 7.

Algorithm 7 Adaptive Step Size Prediction

Input: $\mathbf{t}_{\text{opt}}[k+1] = \begin{bmatrix} t_{\text{opt}}[k] & \cdots & t_{\text{opt}}[k-M-1] \end{bmatrix}^T$
1: Initialize $\mathbf{a}[1] = \begin{bmatrix} a_0[1] & \cdots & a_{M-1}[1] \end{bmatrix}^T$
2: **for all** k=1,2,... **do**
3: $\quad \tilde{t}[k+1] = \mathbf{a}[k]^T \mathbf{t}_{\text{opt}}$
4: $\quad e_t[k+1] = t_{\text{opt}}[k+1] - \tilde{t}[k+1]$
5: $\quad \mathbf{a}[k+2] = \mathbf{a}[k+1] + \mu e_t[k+1] \mathbf{t}_{\text{opt}}[k+1]$
6: **end for**
Output: $\tilde{t}[k+1]$

Therefore, an adaptive step size unitary predictor which consists of step size optimization, LMS algorithm to predict the step size, and the general prediction in

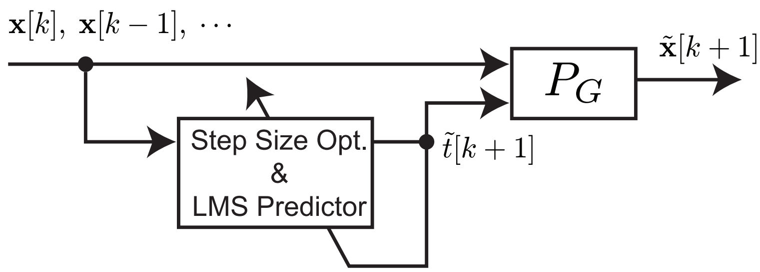

Figure 4.1: Block diagram of proposed adaptive step size predictor with step size optimization, least mean square (LMS) step size predictor, and the general prediction.

(4.17) are obtained. A high level block diagram for the adaptive step size predictor is shown in Fig. 4.1.

4.5 Differential Coding Algorithm

In this section, the proposed Lie theoretic differential coding technique is described for time series evolving on the special unitary group. Unlike prior techniques using rotation based differential coding [4, 58], the propose approach computes the tangent space differential between successive measurement on the special unitary group and perform the quantization on the tangent space. The main benefit of the proposed approach is that the number of real parameters to be quantized are reduced by a half thanks to the skew-Hermitian symmetry of the Lie algebra. Unfortunately, the benefit comes with increase in computational complexity.

Let $\{\mathbf{S}[k]\}_{k\in\mathbb{N}} \in \mathcal{SU}_n$ be the correlated input with time index k on the $n \times n$ special unitary group. The main idea of differential coding is to compute the tangential difference $\mathbf{E}[k]$ between the previous estimate $\hat{\mathbf{S}}[k-1]$ and the current

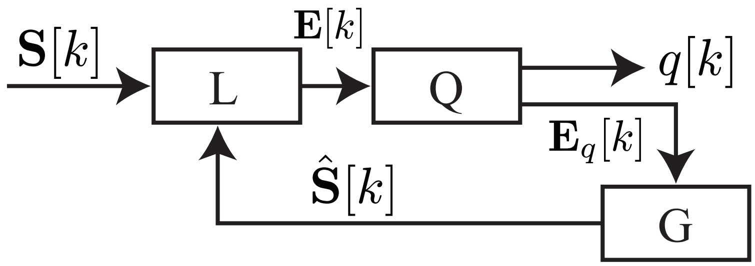

Figure 4.2: Block diagram of differential encoder

observation $\mathbf{S}[k]$ at the encoder. The quantized tangential difference is fed to the decoder to update the previous estimate $\hat{\mathbf{S}}[k-1]$ to obtain the current estimate $\hat{\mathbf{S}}[k]$.

Algorithm 8 Differential encoding algorithm

Input: $\mathbf{S}[k]$ and $\hat{\mathbf{S}}[k-1]$
for all k=1,2,... **do**
  $\mathbf{E}[k] = L(\hat{\mathbf{S}}[k-1], \mathbf{S}[k])$
  $q[k] = Q(\mathbf{E}[k])$
  $\hat{\mathbf{S}}[k] = G(\hat{\mathbf{S}}[k-1], \mathbf{E}_q[k], 1)$
end for
Output: $q[k]$

Fig. 4.2 illustrates the block diagram; the pseudo code is provided in Algorithm 8. At time k, an error tangent matrix $\mathbf{E}[k]$ is computed between the input signal $\mathbf{S}[k]$ and the previous estimate $\hat{\mathbf{S}}[k-1]$ using (4.12). The quantization of the tangent error matrix is performed in two steps. The tangent error matrix is decomposed into a magnitude component and unit directional component as

$$\mathbf{E}[k] = \|\mathbf{E}[k]\| \frac{\mathbf{E}[k]}{\|\mathbf{E}[k]\|}. \tag{4.27}$$

Note that the unit directional component preserves the skew-Hermitian structure. Let $\mathcal{C}_m = \{c_{m,\ell}\}_{\ell=1}^{N_m}$ with N_m codewords denote the codebook of error tangent magnitudes in nonnegative reals. Also, let $\mathcal{C}_d = \{\mathbf{C}_{d,i}\}_{i=1}^{Nd}$ with N_d codewords denote the codebook of unit norm directional tangent matrices in $\mathfrak{su}_n$. The codeword selection criterion is to minimize the tangent error between the estimate $\hat{\mathbf{S}}[k]$ and the observed matrix $\mathbf{S}[k]$. That is, the codeword indices $\ell[k]$ and $i[k]$ are selected according to

$$(\ell[k], i[k]) = \underset{\ell\in\{1,2,\ldots,N_m\}, i\in\{1,2,\ldots,N_d\}}{\arg\min} d(G(\hat{\mathbf{S}}[k-1], c_{m,\ell[k]}\mathbf{C}_{d,i}, 1), \mathbf{S}[k]) \tag{4.28}$$

Recall that

$$G(\hat{\mathbf{S}}[k-1], c_{m,\ell[k]}\mathbf{C}_{d,i}, 1) = e^{c_{m,\ell[k]}\mathbf{C}_{d,i}}\hat{\mathbf{S}}[k-1] \tag{4.29}$$

and

$$\mathbf{S}[k] = e^{\mathbf{E}[k]}\hat{\mathbf{S}}[k-1]. \tag{4.30}$$

Therefore, the quantization error tangent $\hat{\mathbf{E}}[k]$ becomes

$$\begin{aligned}\hat{\mathbf{E}}[k] &= \log(\mathbf{S}[k]\hat{\mathbf{S}}^*[k-1]e^{-c_{m,\ell[k]}\mathbf{C}_{d,i}}) \\ &= \log(e^{\mathbf{E}[k]}\hat{\mathbf{S}}[k-1]\hat{\mathbf{S}}^*[k-1]e^{-c_{m,\ell[k]}\mathbf{C}_{d,i}}) \\ &\approx \mathbf{E}[k] - c_{m,\ell[k]}\mathbf{C}_{d,i}\end{aligned} \tag{4.31}$$

where the last approximation is due to (4.15). Thus the codeword selection becomes

$$(\ell[k], i[k]) = \underset{ell\in\{1,2,\ldots,N_m\}, i\in\{1,2,\ldots,N_d\}}{\arg\min} \|\mathbf{E}[k] - c_{m,\ell[k]}\mathbf{C}_{d,i}\|_F \tag{4.32}$$

which can be computed by searching over $N_m \times N_d$ codeword combinations.

Unfortunately, despite the skew-Hermitian structure of $\mathbf{C}_d$, constructing a good codebook is a difficult task, especially as the dimension n increases. For smaller n, the Lloyd algorithm may be employed to construct the error tangent direction codebook [22]. Unfortunately, it is well known that codebooks with larger dimensions requires large number of iterations. For example, consider the case $n = 2$. Thanks to the skew-Hermitian symmetry and trace zero property, $\mathbf{E}[k]$ can be written as

$$\mathbf{E}[k] = \begin{bmatrix} ja & b + jc \\ -b + jc & -ja \end{bmatrix} \tag{4.33}$$

where $a,\ b,$ and c are real numbers. Then, there are three real numbers representing the tangent space of $\mathcal{SU}_2$. By vectorizing $a,\ b,$ and c, Lloyd algorithm may be performed to obtain a reasonable codebook. Therefore, for smaller dimensions where the codebook for $\mathcal{C}_d$ is available, the selection criterion in (4.32) is used. For larger dimensions, the error tangent magnitude is quantized first according to

$$\ell[k] \quad = \quad \underset{i\in\{1,2,\ldots,N_m\}}{\arg\min}\ |\|\mathbf{E}[k]\|_2 - c_{m,i}| \tag{4.34}$$

where $\ell[k]$ is the index of the selected error tangent magnitude codeword in $\mathcal{C}_m$. Then, the normalized error tangent direction $\mathbf{E}[k]/\|\mathbf{E}[k]$ is quantized element by element to obtain $\mathbf{C}_{d,q}$ using N_q bits per dimension where the total number of bits used to encode $\mathbf{E}[k]/\|\mathbf{E}[k]$ is $N_d = n^2 N_q$ bits. Note that due to the symmetry of the skew-Hermitian matrix, only the upper off-diagonal and the imaginary part of the diagonal needs to be quantized. For notational brevity, the quantization is denoted by $Q : \mathbb{C}^{n^2} \to \mathbb{N}$ such that $q[k] = Q(\mathbf{E}[k])$ consists of $N_m + N_d$ bits representing the quantized $\mathbf{E}[k]$. Finally, the quantized error tangent matrix is used at the encoder to obtain the estimate $\hat{\mathbf{S}}[k]$ to be used at the next time interval. The closed loop approach, i.e., using the difference between the quantized $\hat{\mathbf{S}}[k]$ and observed $\mathbf{S}[k]$, as opposed to taking the difference between successive observed signal $\mathbf{S}[k-1]$

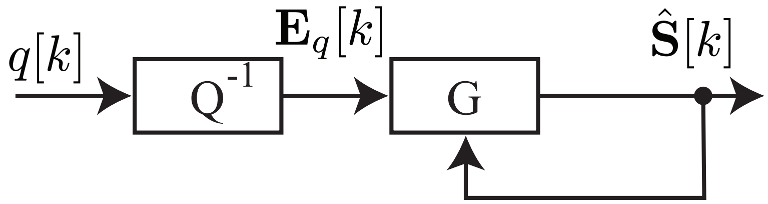

Figure 4.3: Block diagram of differential decoder

and $\mathbf{S}[k]$. This is because unlike source coding applications where the encoding is performed in blocks of signals, limited feedback MIMO applications requires per sample update. Therefore, the closed loop approach provides a better estimate at the decoder side.

Algorithm 9 Differential decoder algorithm

Input: $q[k]$
1: Initialize $\hat{\mathbf{S}}[0]$
2: **for all** k=1,2,… **do**
3: $\mathbf{E}_q[k] = Q^{-1}(q[k])$
4: $\hat{\mathbf{S}}[k] = G(\hat{\mathbf{S}}[k-1], \mathbf{E}_q[k], 1)$
5: **end for**
Output: $\hat{\mathbf{S}}[k]$

Fig. 4.3 illustrates the proposed differential decoder; the pseudo code is provided in Algorithm 9. The quantized error tangent index $q[k]$ is the input to the decoder. Based on the predefined quantization method, i.e., codebook based or per-element quantization, the received index $q[k]$ is used to reconstruct the quantized error tangent matrix $\mathbf{E}_q[k]$. Then, $\mathbf{E}_q[k]$ is used in (4.13) with $t = 1$ to obtain the estimate $\hat{\mathbf{S}}[k]$.

Due to the closed loop approach, initial value of $\hat{\mathbf{S}}[0]$ must be agreed upon between the encoder and the decoder. Fortunately, this is easily accomplished by setting the initial value $\hat{\mathbf{S}}[0] = \mathbf{I}$.

4.6 Predictive Coding Algorithm

In this section, tools developed thus far are used to develop the proposed predictive coding framework for time series on $\mathcal{SU}_n$.

Let $\{\mathbf{S}[k]\}_{k\in\mathbb{N}} \in \mathcal{SU}_n$ be the correlated input as before. The general operation of the predictive coding follows closely to that of the conventional predictive coding algorithm [22]. Linear operations such as difference and addition are replaced by the tangent and geodesic operations. The main idea of predictive coding is to use a predicted matrix based on past estimates and quantize the difference between the predicted matrix and the observed matrix. Given a good predictor, the benefit of predictive coding is that the prediction error usually has smaller dynamic range compared to differential coding technique.

Algorithm 10 Predictive encoder algorithm

Input: $\mathbf{S}[k],\ t_{\text{opt}}[k]$

Initialize $\tilde{\mathbf{S}}[1]$
for all k=1,2,... **do**
  $\mathbf{E}[k] = L(\tilde{\mathbf{S}}[k], \mathbf{S}[k])$
  $q[k] = Q(\mathbf{E}_k)$
  $\hat{\mathbf{S}}[k] = G(\tilde{\mathbf{S}}[k], \mathbf{E}_q[k], 1)$
  $\tilde{\mathbf{S}}[k+1] = P_{\text{opt}}(\hat{\mathbf{S}}[k-1], \hat{\mathbf{S}}[k], t_{\text{opt}}[k])$
end for

Output: $q[k]$

Fig. 4.4 illustrates the block diagram of the proposed predictive coding en-

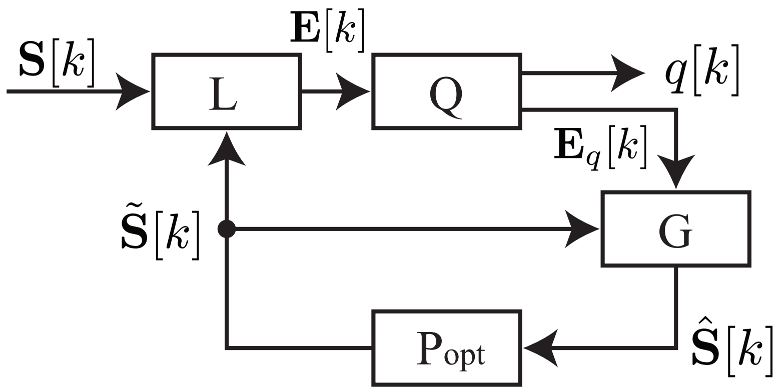

Figure 4.4: Block diagram of predictive encoder

coder; the pseudo code is provided in Algorithm 10. At time k, a prediction error tangent matrix $\mathbf{E}[k]$ from the predicted matrix $\tilde{\mathbf{S}}[k]$ to $\mathbf{S}[k]$ is computed using (4.12). The prediction error is quantized using the same quantization process described in Section 4.5. Thus $q[k]$ is obtained with $N_m + N_d$ bits representing the quantized $\mathbf{E}[k]$. Then $q[k]$ is transmitted to the decoder via a finite rate feedback channel. Continuing at the encoder, the quantized prediction error $\mathbf{E}_q[k]$ is used to obtain the quantized estimate $\hat{\mathbf{S}}[k]$ using (4.13). Finally, $\hat{\mathbf{S}}[k]$ together with $\hat{\mathbf{S}}[k-1]$ is used to obtain the predicted matrix $\tilde{\mathbf{S}}[k+1]$ using the adaptive step size prediction described in Section 4.4.2. The adapted step size is also quantized and transmitted to the decoder. Note that prediction based on the past estimates rather than the observed matrices is important because the decoder only has the estimated matrices.

Fig. 4.5 illustrates the block diagram of the proposed predictive coding decoder; the pseudo code is provided in Algorithm 11. The method for decoding the quantized prediction error, i.e., codebook based or element by element quantiza-

Algorithm 11 Predictive decoder algorithm

Input: $q[k],\ t_{\text{opt}}[k]$

Initialize $\tilde{\mathbf{S}}[1]$

for all k=1,2,... **do**

  $\mathbf{E}_q[k] = Q^{-1}(q[k])$

  $\hat{\mathbf{S}}[k] = G(\tilde{\mathbf{S}}[k], \mathbf{E}_q[k], 1)$

  $\tilde{\mathbf{S}}[k+1] = P_{\text{opt}}(\hat{\mathbf{S}}[k-1], \hat{\mathbf{S}}[k], t_{\text{opt}}[k])$

end for

Output: $\hat{\mathbf{S}}[k]$ or $\tilde{\mathbf{S}}[k]$

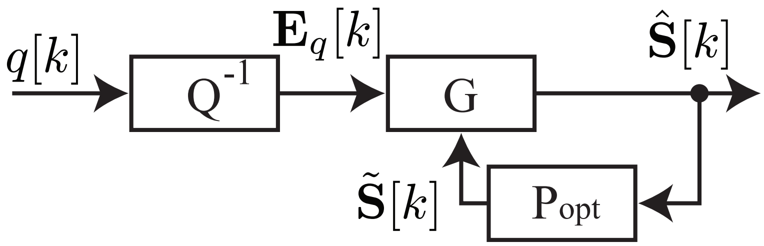

Figure 4.5: Block diagram of predictive decoder

tion, is assumed to be known a priori. The decoder receives $q[k]$ from which the quantized prediction error $\mathbf{E}_q[k]$ is constructed by $Q^{-1}(q[k])$. Then, the quantized prediction error is used together with previous estimate $\hat{\mathbf{S}}[k-1]$ to obtain the estimated matrix $\hat{\mathbf{S}}[k]$ using (4.13). The output of the decoder is $\hat{\mathbf{S}}[k]$. Finally, using the step size received from the encoder, prediction is performed as described in Section 4.4.2 to obtain $\tilde{\mathbf{S}}[k+1]$ for the next time period. For initialization, the identity can be used again as the initial predicted matrix, e.g., $\tilde{\mathbf{S}}[0] = \mathbf{I}$, so that both the encoder and decoder have the same starting point.

4.7 Simulation Results

In this section, simulation results of the proposed differential and predictive coding algorithm for limited feedback MIMO communication systems are presented. First, the channel model used throughout the section is described. Second, the proposed differential coding strategy is applied to single user limited feedback MIMO systems. Third, application of predictive coding to limited feedback MIMO systems and mean squared error performance are illustrated. Finally, sum rate performance of the limited feedback-based block diagonalization in multiuser MIMO using the proposed predictive coding is shown.

4.7.1 Channel Model

It is assumed that the $N_r \times N_t$ channel matrix $\mathbf{H}[k]$ is temporally correlated according to a first order autoregressive model (or Gauss-Markov model [20,38,58]) with correlation coefficient $\alpha = J_0(2\pi\beta)$ where J_0 is the Bessel function of zeroth order and β is the normalized Doppler frequency. The channel matrix at time k is generated according to

$$\mathbf{H}[k] = \alpha\mathbf{H}[k-1] + \sqrt{1-\alpha^2}\mathbf{Z}[k] \tag{4.35}$$

where $\mathbf{Z}[k]$ is a random matrix drawn from an i.i.d. zero mean complex white Gaussian process. The receiver is assumed to have perfect knowledge of the channel matrix. From the channel matrix, singular value decomposition $\mathbf{H}[k] = \mathbf{U}[k]\mathbf{S}[k]\mathbf{V}^*[k]$ is performed where $\mathbf{U}[k]$ is $N_r \times N_r$ unitary matrix of left singular vectors as columns, $\mathbf{S}[k]$ is $N_r \times N_t$ diagonal matrix of singular values, and $\mathbf{V}[k]$ is $N_t \times N_t$

unitary matrix of right singular vectors as columns. The time series of right singular matrix $\mathbf{V}[k]$ is considered to be the evolution of unitary matrices from which the special unitary matrices are derived as described in Section 4.3.1.

4.7.2 Differential Coding

In this section, $N_t = N_r = 4$ limited feedback MIMO system using $N_s = 1$ and 2 streams with differential coding feedback is considered. The limited feedback channel is assumed to be error and delay free. Three scenarios are considered for comparison. The first scenario is where the transmitter has perfect knowledge of the CSI. In this case, the precoding matrix, or vector for beamforming, is obtained as N_s dominant columns of $\mathbf{V}[k]$. The second scenario is the Grassmannian memoryless limited feedback approach [64,67]. Finally, the third scenario is the proposed differential coding strategy. The proposed differential coding uses $N_m + N_d$ bits of feedback from which a precoder for any N_s can be obtained by selecting the first N_s columns of $\hat{\mathbf{S}}[k]$. All the simulation results are performed for $\beta = 0.001$.

Fig. 4.6 and Fig. 4.7 shows the symbol error rate and vector symbol error rate performance for $N_s = 1$ and 2, respectively. Note that for each N_s, the memoryless limited feedback strategy uses different codebooks where as the results across both figures for the proposed differential coding strategy uses the precoder from single feedback. Thanks to the improved resolution obtained by the proposed differential feedback, the symbol error rate curves essentially overlaps with the perfect CSI case. Thus the proposed strategy provides improved CSI accuracy with additional freedom for the transmitter to select the rank of the precoder which may

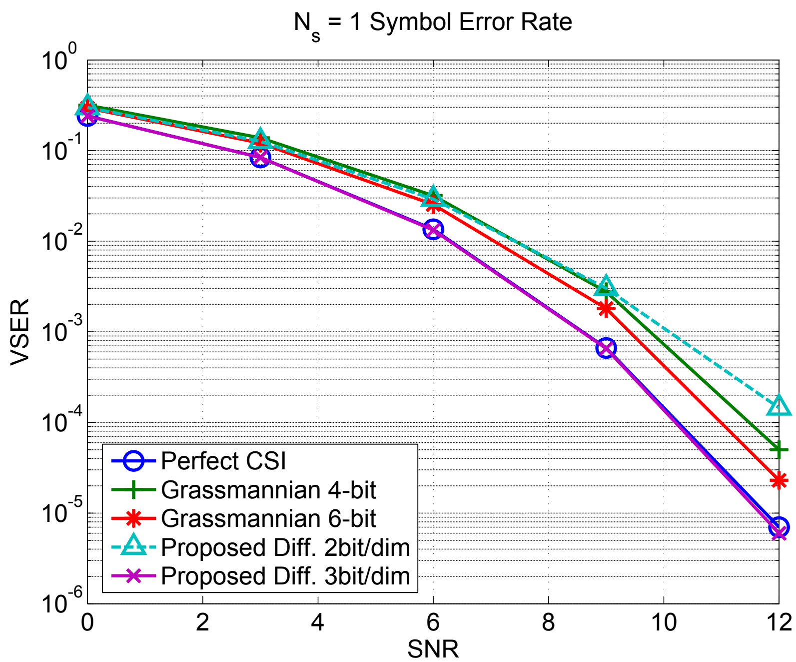

Figure 4.6: Symbol error rate comparison for $N_t = N_r = 4$ limited feedback MIMO system using $N_s = 1$ with memoryless limited feedback approach (4 and 6-bit codebooks) and the proposed differential coding strategy using 2 and 3 bits per dimension with 3 bit tangent magnitude codebook.

be useful, for example, in multimode precoding [65].

4.7.3 Predictive Coding

In this section, the proposed predictive coding for $N_t = N_r = 4$ limited feedback MIMO system using $N_s = 1$ and 2 streams is considered. First, the case with error-free and delay free limited feedback channel is considered. Again, three scenarios are considered for comparison; perfect CSI, Grassmannian memoryless limited feedback, and the proposed predictive coding strategy. All the simulation

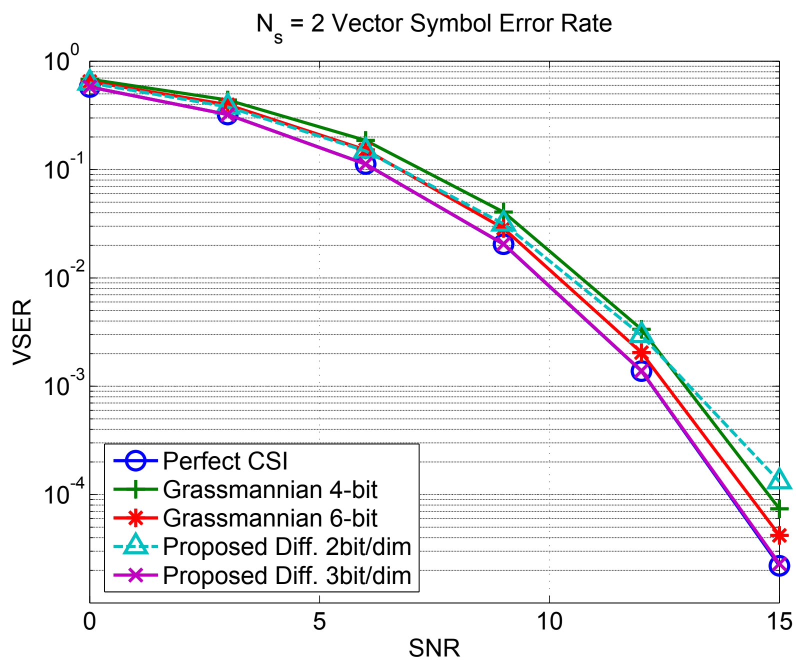

Figure 4.7: Vector Symbol error rate comparison for $N_t = N_r = 4$ limited feedback MIMO system using $N_s = 2$ with memoryless limited feedback approach (4 and 6-bit codebooks) and the proposed differential coding strategy using 2 and 3 bits per dimension with 3 bit tangent magnitude codebook.

results are performed for $\beta = 0.001$.

Fig. 4.8 shows the achievable throughput comparison for cases $N_s = 1,\ 2$ and 3 using perfect CSI, memoryless limited feedback with 4 and 6 bits of feedback and the proposed predictive approach using 3 bits per dimension and 3 bits for error tangent magnitude codebook. The proposed approach essentially overlaps with the perfect CSI case showing its superior resolution and achievable throughput performance. Note that for memoryless limited feedback, each N_s case uses its specific codebook where as the proposed approach can achieve any one of the rates shown

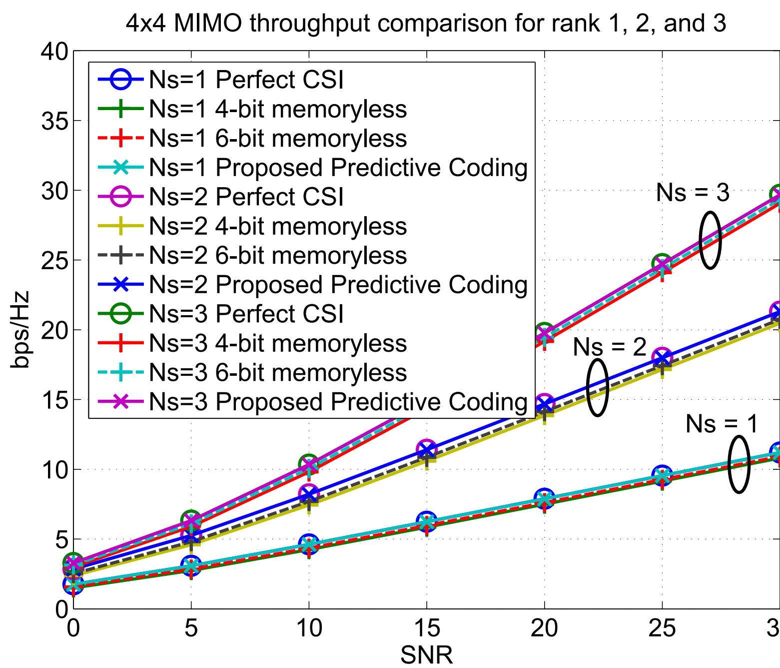

Figure 4.8: Achievable throughput comparison for $N_t = N_r = 4$ and $N_s = 1,\ 2$ and 3 limited feedback MIMO system using perfect CSI, memoryless limited feedback approach (4 and 6-bit codebooks), and the proposed predictive coding strategy using 3 bits per dimension with 3 bit tangent magnitude codebook. The achievable throughput for the proposed approach essentially overlaps with the perfect CSI case.

with one instance of the feedback. For comparison, if 6 bits were used for memoryless feedback for all N_s, the total feedback rate is 24 bits to allow different ranks. Note that corresponding codebooks needs to be stored at both the transmitter and the receiver. The proposed predictive coding results in 16 real elements quantized at 3 bits and an additional 3 bits for error tangent magnitude resulting in 51 bits of feedback. Unfortunately, the feedback rate is larger than the memoryless approach but there is essentially no codebook storage overhead with significant performance benefit.

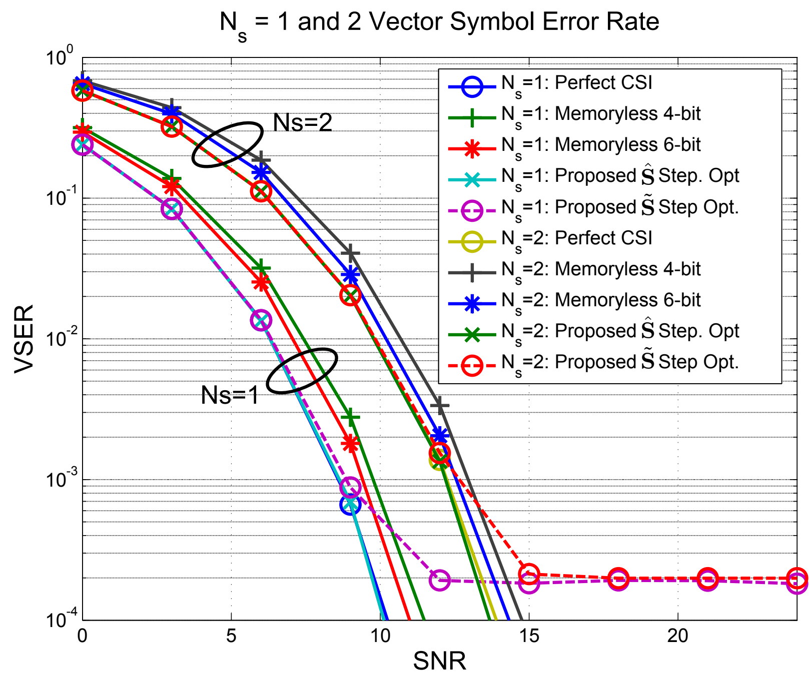

Figure 4.9: Symbol error rate comparison for $N_t = N_r = 4$ and $N_s = 1$ and 2 limited feedback MIMO system using perfect CSI, memoryless limited feedback approach (4 and 6-bit codebooks), and the proposed predictive coding strategy using 3 bits per dimension with 3 bit tangent magnitude codebook.

Fig. 4.9 shows the symbol error rate and vector symbol error rate performance for $N_s = 1$ and 2, respectively. The perfect CSI and memoryless limited feedback results are identical to the results presented for differential coding strategy in Section 4.7.2. The results for predictive coding across both figures were generated based on the same feedback information. Thanks to the improved resolution obtained by the predictive coding, the symbol error rate curves essentially overlaps with the perfect CSI case. Thus the proposed strategy provides improved CSI accuracy with additional freedom for the transmitter to select the rank of the precoder which may be useful, for example, in multimode precoding [65].

Now suppose that there is a unit time delay in the limited feedback channel. Then similar to the predictive coding methods proposed in [45, 46], the predicted matrix $\tilde{\mathbf{S}}[k]$ may be used as the output of the decoder as the *predicted estimate*. The symbol error rate and vector symbol error rate using $\tilde{\mathbf{S}}[k]$ as the output of the decoder are shown with the marker $\triangle$. The results show that for low SNR, the predicted output performs well following the perfect CSI curve. Unfortunately, as the SNR increases, the error rates worsens and shows an error floor above 15dB. It is conjectured that the prediction based on quantized estimate causes the error rate to floor at higher SNR. This subject will be investigated in the future work.

To evaluate the performance of the proposed predictive coding algorithm, Fig. 4.10 illustrates the mean squared error performance of the proposed approaches over normalized Doppler frequency β. For $N_s = 1$, the mean squared chordal distance (3.1) was evaluated. For $N_s = 2$ and 3, the projection Frobenius norm (4.5) was used. Fig. 4.10 shows that by using $\tilde{S}$ with fixed step size as the output of the decoder results in the worst performance. Furthermore, at lower values of β, i.e., input more correlated, the MSE starts to saturate. Using $\hat{\mathbf{S}}$ as the output of the decoder with fixed step size prediction, the overall MSE performance is improved by approximately 5dB. Again, saturation is observed when the input is highly correlated. Finally, using $\tilde{\mathbf{S}}$ with adaptive step size optimization, the MSE performance is improved approximated another 5dB without the saturation effect at lower values of β. This shows that the adaptive step size strategy is especially effective for more correlated input.

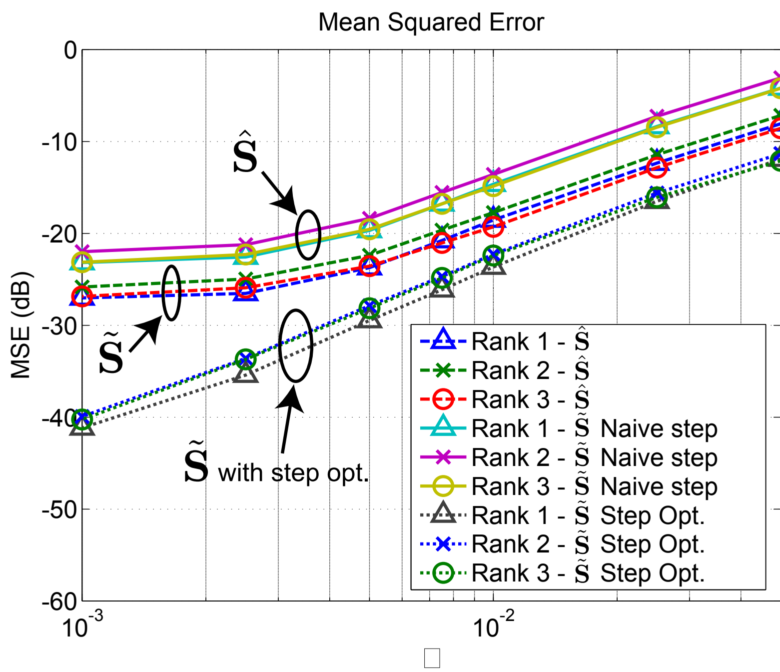

Figure 4.10: Mean squared error (MSE) performance of proposed approaches over normalized Doppler frequency β. The naive one step prediction using $\tilde{\mathbf{S}}$ results in the worst MSE performance. Using the estimated output $\hat{\mathbf{S}}$ improves the MSE by approximately 5dB for all the ranks. The best MSE performance is obtained with step size optimization which improves the MSE by approximately 5dB for higher β (less correlated) and it continues to improve for lower β (more correlated).

Finally, the sum rate performance of the limited feedback-based block diagonalization in multiuser MIMO system is simulated using the proposed predictive coding algorithm. For this simulation, a system with $N_t = U = 4$ and $N_r = N_s = 2$ is considered. At each mobile, the right singular matrix corresponding to the observed $N_r \times N_t$ channel matrix is computed from which the $N_t \times N_t$ special unitary matrix $\mathbf{S}[k]$ is derived. Then, $\mathbf{S}[k]$ is fed into the proposed predictive coding algorithm. For this simulation, normalized Doppler frequency of $\beta = 0.0278$ and

0.0556 are considered. In practical systems, $\beta = 0.0278$ corresponds to a system operating at carrier frequency of 2GHz, mobile speed of 3km/h, and channel update interval of 5milliseconds. Similarly, $\beta = 0.0556$ corresponds to a system with mobile moving at 6km/h. For quantization, 3 bits per dimension is used for the proposed algorithms. Fig. 4.11 shows the sum rate performance with perfect CSI at the transmitter, proposed predictive coding with $t = 1$ and adaptive step size with and without delay, and the Grassmannian codebook approach [86]. The figure illustrates the superior sum rate performance of the proposed predictive coding algorithm, especially for lower mobile speed. For each mobile speed, the figure also illustrates the improvement obtained by using the adaptive step size as opposed to fixed step size $t = 1$.

4.8 Summary

In this chapter, a new differential and predictive coding framework for time-series of unitary matrices were proposed. Lie theory was used to derive a tangent difference, mapping back to the special unitary group, and most importantly a prediction framework with step size optimization exploiting the group and differential geometric structure. Furthermore, an adaptive step size algorithm using LMS was proposed for practicality. Applications of the differential and predictive coding in limited feedback MIMO system was shown to provide near perfect CSI performance and additional flexibility for the transmitter to override the rank of the transmission strategy. The main drawback of the proposed algorithms is the increase in the number of feedback bits. Future work will consider exploiting the structure of

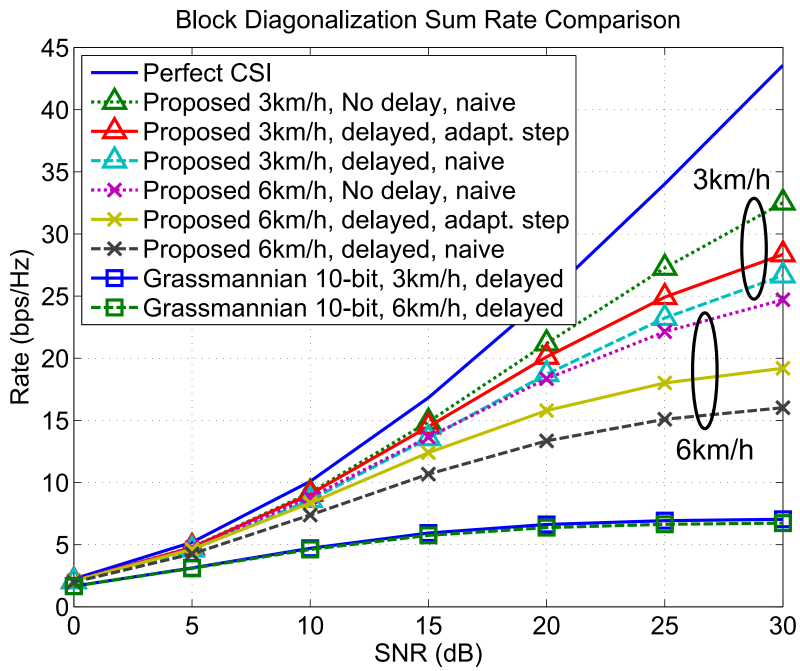

Figure 4.11: Sum rate performance of limited feedback-based block diagonalization in multiuser MIMO system at mobile speeds of 3km/h and 6km/h. Perfect CSI at the transmitter, proposed predictive coding algorithms, and memoryless Grassmannian codebook approach from [86] are shown.

Lie algebra to obtain a more efficient quantization strategy.

4.9 Appendix

4.9.1 Proof of Theorem 15

Proof. First, expand the Frobenius norm expression on the right hand side of (4.22)

$$\begin{aligned}
& \|\mathbf{E}[k+1] - \mathbf{E}[k]t\|_F^2 \\
= \;& \mathrm{tr}\left[(\mathbf{E}[k+1] - \mathbf{E}[k]t)^*(\mathbf{E}[k+1] - \mathbf{E}[k]t)\right] \\
= \;& \mathrm{tr}(\mathbf{E}^*[k+1]\mathbf{E}[k+1]) - 2\Re(\mathrm{tr}(\mathbf{E}^*[k+1]\mathbf{E}[k]))t \\
& + \mathrm{tr}(\mathbf{E}^*[k]\mathbf{E}[k])t^2. \qquad (4.36)
\end{aligned}$$

This expression is quadratic function of t, so taking the derivative of (4.36)

$$\begin{aligned}
& \frac{d\|\mathbf{E}[k+1] - \mathbf{E}[k]t\|_F^2}{dt} \\
= \;& 2\mathrm{tr}(\mathbf{E}^*[k]\mathbf{E}[k])t - 2\Re(\mathrm{tr}(\mathbf{E}^*[k+1]\mathbf{E}[k]))
\end{aligned}$$

and setting it equal to zero yields the optimum step size

$$t_{\text{opt}} = \frac{\Re(\mathrm{tr}(\mathbf{E}^*[k+1]\mathbf{E}[k]))}{\mathrm{tr}(\mathbf{E}^*[k]\mathbf{E}[k])}.$$

□

Chapter 5

Conclusion

5.1 Summary

In this dissertation, a new codebook design and new signal processing foundations, namely, prediction, predictive coding, and prediction step size optimization techniques, on Grassmann manifold and manifold of unitary matrices were proposed for applications to single user and multiuser MIMO wireless systems.

In Chapter 2, a new avenue of structured codebook design, called the Kerdock codebook, for limited feedback unitary precoded MIMO systems was proposed. Ideas from mutually unbiased bases and coding theory were used to arrive at a systematic codebook construction and codebook generation techniques. The proposed Kerdock codebook has elements drawn from a quaternary alphabet, $\{\pm 1,\ \pm j\}$, resulting in reduced storage, reduced computational complexity, and search complexity. Furthermore, it was shown that both beamforming and spatial multiplexing codebooks can be derived from a single codebook. Subspace distance properties of the codebook were proven showing that the codebook possess a favorable structure similar to Grassmannian codebook. Simulation results for beamforming and spatial multiplexing systems showed that despite the finite alphabet representation, symbol error rate and achievable throughput performance comparable to the same size Grassmannian and Fourier based codebooks can be obtained.

In Chapter 3, a new signal processing foundation was developed for signals on the Grassmann manifold. Using the differential geometric structure of the Grassmann manifold, a simple formulas for tangential difference, mapping onto the Grassmann manifold, parallel transport, prediction, and optimization of prediction were developed. Using these building blocks, the Grassmannian predictive coding algorithm was developed. The general algorithmic construction follows the classical predictive coding algorithm with blocks replaced by Grassmann manifold counterparts. A new tangent space quantization strategy was proposed which decomposed the error tangent vector into its magnitude and directional components. Furthermore, motivated by the delayed limited feedback MIMO system applications, a modified form Grassmannian predictive coding algorithm with adaptive step size predictor was proposed in which the predictor output is used as the output of the decoder. Using this modified Grassmannian predictive coding algorithm, a unit sample delay between the encoder and the decoder can be compensated. Distortion bound for the Grassmannian predictive coding was proven showing significant improvement over memoryless quantization techniques. Two immediate applications in limited feedback beamforming system and multiuser MIMO system were shown. Simulation results showed that under feedback rate constraint, the proposed prediction algorithm provides substantial symbol error rate improvement and sum rate improvement over memoryless feedback approach. For mild temporal channel correlation, the proposed algorithms can attain near perfect CSI performance under feedback constraint. For limited feedback multiuser MIMO systems, the proposed method overcomes the severe sum rate limitation with memoryless feedback approaches.

In Chapter 4, a new signal processing foundation was developed for signals on the manifold of unitary matrices. A new decomposition of unitary matrix was used to obtain a special unitary matrix. Using the group and differential geometric operation from Lie group theory, basic operations such as differencing, mapping onto the manifold, and prediction were proposed. In contrast to Grassmannian prediction, the symmetry in Lie groups of special unitary matrices eliminates the need to perform parallel transport. The interpretation is that tangent space at any point on the manifold can be translated to the identity where the tangent space is characterized by the Lie algebra. Using the Baker-Campbell-Hausdorff formula, a first order prediction step size optimization formula was derived. A tangent space quantization strategy was proposed where the skew-Hermitian symmetry was exploited to reduce the number of parameters by a half. Exploiting the Lie group structure, a differential coding and predictive coding strategies were proposed. The differential coding was shown to provide high resolution feedbacks using simple element by element quantization. Unfortunately, differential coding is only suitable when the channel between the encoder and the decoder is error and delay free. A unitary predictive coding algorithm with step size optimization and prediction was proposed. The mean squared error performance with and without step size showed that step size optimization improves the mean squared error by approximately 12dB. Applications of the proposed differential and predictive coding algorithms were shown for single user limited feedback MIMO systems with transmission rank 1, 2, and

3. The main benefit of the proposed unitary feedback approach is that one instance of feedback can be used to obtain high resolution CSI as well as the freedom for the transmitter to override the transmission rank. The proposed predictive coding algorithm was also applied to limited feedback-based multiuser MIMO system. The proposed algorithms were shown to provide significantly improved sum rate in mildly correlated channel.

5.2 Future Work

Several future research directions are identified below.

Extension of Kerdock Codebook: Recently, multi-cell and cooperative MIMO systems where multiple base stations collaborate to form a large multiuser system have been receiving significant interest both from the research community as well the commercial wireless standards [3, 82, 95–97]. Limited feedback strategies and codebook designs for cooperative MIMO systems are still largely an open problem. In cooperative MIMO systems, it is desirable to have a codebook with unitary constraint for each cooperating base stations as this would separate the beamforming from power allocation for maximum flexibility. The Kerdock codebook, upon inspection, has Hadamard matrix-like structure within each mutually unbiased basis. It is possible to take size 2 sub-vector from size 4 codebook. With appropriate scaling, a unitary vector is obtained. Thus if two $N_t = 2$ base stations are cooperating, it is possible to use partitions of the Kerdock codebook for each base station, yet the combined beamformer can result in desired $N_t = 4$ beamforming vector. It would be interesting to investigate the structure of the Kerdock codebook further and consider its application in cooperative MIMO systems.

Extensions of Predictive Coding on the Manifold: In Chapters 3 and 4, a fundamental form of predictive coding algorithms on Grassmann manifold and manifold of unitary matrices were derived. A direct extension of this work may consider different quantization approaches, more rigorous quantization bounds, feedback compression techniques, and alternative prediction strategies. The tangent space of the manifolds considered in this dissertation deserves further investigation to gain an understanding of its structure. Exploiting the structure of the tangent space will likely result in a more efficient quantization techniques, as was done with the manifolds in this dissertation. For example, what is the smallest codebook that can be used maintain certain communication theoretic performance criterion? A further extension may consider the fundamental compressibility of signals arising on the manifolds. An information theoretic study of signals on manifolds has not appeared before thus this area appears to be wide open.

Optimal Signal Processing on the Manifold: In Chapters 3 and 4, although the chordal distance metric was used for optimization of prediction functions, the operations used were based on deterministic geometric formulations without a connection to underlying statistics. In order to generalize the basic operations identified in this dissertation, statistical characterizations as well as optimality criteria will aid in analyzing the impact of these algorithms to applications in MIMO wireless systems. For example, what is the minimum feedback information rate needed to achieve certain rate guarantee? A detailed study is needed to understand the statistical distribution and temporal behavior of CSI derived from channels with known statistics. This aspect is complicated by the singular value decomposition to obtain the CSI. Once the connection can be made between the actual channel statistics with manifold valued signal, a rich field of statistical signal processing on a manifold may emerge. Some literatures do exists on directional statistics on manifolds but connecting them with communication theoretic channel statistics may be challenging [10, 72].

Higher Order Prediction and Interpolation: In Chapters 3 and 4, only the first order prediction was proposed. In MIMO channels with longer memory, higher order prediction may prove useful. This will entail defining a smooth function on the manifold based on past measurements and constructing a prediction procedure such that some smoothness and optimality criteria are met. Interpolation on the manifold will require similar tools to construct a smooth function on the manifold based on measurement samples. The difficulty with interpolation is that the formulas derived in this dissertation relied on the shortest distance geodesics. The function to be interpolated may not always be on the geodesic as evidenced in [11, 81]. A detailed study on differential geometry and Lie group is needed to quantify deviations of a function on the manifold. Inevitably, statistical characterization of the function on the manifold is expected to play a role in defining an optimal interpolation.

Application to Interference Alignment: Recently, a fully connected K-user MIMO interference channel with interference alignment has been proposed [9]. The main idea is to design a beamforming direction from each transmitter such that each user sees a desired signal subspace and interference subspace where all the interferences are to fall in a common subspace at each receiver terminal. So far, the problem of designing beamformers has been solved assuming global CSI and reciprocal channels using iterative algorithms and closed form solutions for a limited case. Solutions for each beamformer appear to fall in the Grassmann manifold but the problem is made difficult by the fact that 1) there are K beamformers on Grassmann manifold, 2) k^2 interconnecting channels, and 3) K receiver Grassmann manifolds in which the interference must be aligned. Unlike the MIMO wireless systems presented in this dissertation, K-user interference channel bring a complex interaction of manifolds. Some promising related result has recently appeared which formulate a consensus optimization problem on the manifold [92]. This problem identifies one of many pertinent future research direction to extend manifold-constrained signal processing techniques developed in this dissertation.

Bibliography

[1] 3GPP. Physical layer aspects of UTRA high speed downlink packet access. *Technical Report TR25.814*, 2006.

[2] 3GPP. 3GPP TS 36.211 V8.5.0 E-UTRA Physical Channels and Modulation (Release 8). 2008.

[3] 3GPP. Evolved universal terrestrial radio access (E-UTRA); physical layer procedures. *Technical Report TS 36.213 v.8.7.0*, May 2009.

[4] T. Abe and G. Bauch. Differential codebook MIMO precoding technique. In *Proc. of IEEE Global Telecom. Conf.*, pages 3963–3968, 2007.

[5] P.-A. Absil, Mahony R. E., and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[6] P.-A. Absil, R. E. Mahony, and R. Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80:199–220(22), 2004.

[7] A. Ashikhmin and R. Gopalan. Grassmannian packings for efficient quantization in MIMO broadcast systems. In *Proc. of IEEE Int. Symp. on Info. Theory*, pages 1811–1815, 2007.

[8] A. Barg and D. Nogin. Bounds on packings of spheres in the grassmann manifold. *IEEE Trans. Inf. Theory*, 48(9):2450–2454, 2002.

[9] V. R. Cadambe, S. A. Jafar, and S. Shamai. Interference alignment on the deterministic channel and application to fully connected Gaussian interference networks. *IEEE Trans. Inf. Theory*, 55(1):269–274, 2009.

[10] Y. Chikuse. *Statistics on Special Manifolds*. Lecture notes in statistics (Springer-Verlag) 174. Springer, New York, 2003.

[11] J. Choi and R. W. Heath Jr. Interpolation based transmit beamforming for MIMO-OFDM with limited feedback. *IEEE Trans. Signal Process.*, 53(11):4125–4135, 2005.

[12] A. M. Cipriano, I. Kammoun, and J.-C. Belfiore. Simplified decoding for some non-coherent codes over the grassmannian. In *Proc. of IEEE Int. Conf. on Commun.*, volume 2, pages 757–761, 2005.

[13] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes, etc.: packings in grassmannian spaces. *Experimental Mathematics*, 5:139–159, 1996.

[14] M. Costa. Writing on dirty paper (corresp.). *IEEE Trans. Inf. Theory*, 29(3):439–441, 1983.

[15] W. Dai, Y. Liu, and B. Rider. Quantization bounds on grassmann manifolds of arbitrary dimensions and MIMO communications with feedback. In *Proc. of IEEE Global Telecom. Conf.*, volume 3, pages 1456–1460, 2005.

[16] W. Dai, Y. Liu, and B. Rider. Quantization bounds on grassmann manifolds and applications to MIMO communications. *IEEE Trans. Inf. Theory*, 54(3):1108–1123, 2008.

[17] J. R. Deller Jr., J. G. Proakis, and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, 1993.

[18] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Analysis and Applications*, 20(2):303–353, 1998.

[19] O. Egger, P. Fleury, T. Ebrahimi, and M. Kunt. High-performance compression of visual information-a tutorial review. i. still pictures. *Proc. IEEE*, 87(6):976–1013, 1999.

[20] R. H. Etkin and D. N. C. Tse. Degrees of freedom in some underspread MIMO fading channels. *IEEE Trans. Inf. Theory*, 52(4):1576–1608, 2006.

[21] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6:311–335, 1998.

[22] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic, 1991.

[23] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Salzer. Shifting the MIMO paradigm. *IEEE Signal Process. Mag.*, 24(5):36–46, 2007.

[24] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE J. Sel. Areas Commun.*, 21(5):684–702, 2003.

[25] G. H. Golub and C. H. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[26] R. Gow. Generation of mutually unbiased bases as powers of a unitary matrix in 2-power dimensions. 2007.

[27] B. C. Hall. *Lie groups, Lie algebras, and Representations : An Elementary Introduction*. Graduate texts in mathematics. 222. Springer-Verlag, New York, 2003.

[28] A. R. Hammons Jr., P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Sole. The Z4-linearity of Kerdock, Preparata, Goethals, and related codes. *IEEE Trans. Inf. Theory*, 40(2):301–319, 1994.

[29] H.-M. Hang and J. Woods. Predictive vector quantization of images. *IEEE Trans. Commun.*, 33(11):1208–1219, 1985.

[30] A. Haoui and D. Messerschmitt. Predictive vector quantization. In *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, volume 9, pages 420–423, 1984.

[31] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[32] R. W. Heath Jr., S. Sandhu, and A. J. Paulraj. Antenna selection for spatial multiplexing systems with linear receivers. *IEEE Commun. Lett.*, 5(4):142–144, 2001.

[33] R. W. Heath Jr., T. Strohmer, and A. J. Paulraj. On quasi-orthogonal signatures for CDMA systems. *IEEE Trans. Inf. Theory*, 52(3):1217–1226, 2006.

[34] R. W. Heath Jr., T. Wu, and A. C. K. Soong. Progressive refinement for high resolution limited feedback multiuser MIMO beamforming. In *Proc. of Asilomar Conf. on Signals, Systems and Computers*, pages 743–747, 2008.

[35] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke. Systematic design of unitary space-time constellations. *IEEE Trans. Inf. Theory*, 46(6):1962–1973, 2000.

[36] K. Huang, J. G. Andrews, and R. W. Heath Jr. Orthogonal beamforming for SDMA downlink with limited feedback. In *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, volume 3, pages 97–100, 2007.

[37] K. Huang, J. G. Andrews, and R. W. Heath Jr. Performance of orthogonal beamforming for SDMA with limited feedback. *IEEE Trans. Veh. Technol.*, 58(1):152–164, 2009.

[38] K. Huang, R. W. Heath, Jr., and J. G. Andrews. Limited feedback beamforming over temporally-correlated channels. *IEEE Trans. Signal Process.*, 57(5):1959–1975, 2009.

[39] K. Huang, R. W. Heath Jr., and J. G. Andrews. Multiuser limited feedback for wireless multi-antenna communication. In *Proc. of IEEE Int. Symp. on Info. Theory*, pages 2036–2040, 2007.

[40] K. Huang, B. Mondal, R. W. Heath Jr., and J. G. Andrews. Markov models for limited feedback MIMO systems. In *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, volume 4, pages 9–12, 2006.

[41] IEEE. IEEE 802.16e-2005 IEEE Standard for Local and metropolitan area networks, Part 16: Air interface for fixed broadcast wireless access systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in License Bands and Corrigendum 1. Dec. 2005.

[42] IEEE. IEEE P802.11n-2006 part 11: Wireless LAN medium access control (MAC) and physical layer (phy) specifications: High-speed physical layer in the 5 GHz band. *IEEE Standard 802.11n*, 2006.

[43] IEEE. IEEE 802.16m System Description Document 09/0034r1. *IEEE 802.16 Broadband Wireless Access Working Group*, Sep 2009.

[44] T. Inoue and R. W. Heath Jr. Geodesic prediction for limited feedback multiuser MIMO systems in temporally correlated channels. In *Proc. of IEEE Radio and Wireless Symposium*, pages 167–170, 2009.

[45] T. Inoue and R. W. Heath Jr. Grassmannian prediction and coding for delayed limited feedback MIMO systems. *To be submitted to IEEE Trans. on*

Signal Processing, 2009.

[46] T. Inoue and R. W. Heath Jr. Grassmannian predictive coding for delayed limited feedback MIMO systems. In *To appear in Proc. of the Allerton Conf. on Comm. Control and Comp.*, 2009.

[47] T. Inoue and R. W. Heath Jr. Kerdock codes for limited feedback precoded MIMO systems. *IEEE Trans. Signal Process.*, 57(9):3711–3716, 2009.

[48] T. Inoue and R. W. Heath Jr. Predictive coding on the Grassmann manifold. *Submitted to IEEE Trans. Signal Process.*, Aug 2009.

[49] A. Iserles and A. Zanna. Efficient computation of the matrix exponential by generalized polar decompositions. *SIAM J. Numer. Anal.*, 42(5):2218–2256, 2005.

[50] A. K. Jain. Image data compression: A review. *Proc. IEEE*, 69(3):349–389, 1981.

[51] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Trans. Inf. Theory*, 52(11):5045–5060, 2006.

[52] G. Jongren and M. Skoglund. Quantized feedback information in orthogonal space-time block coding. *IEEE Trans. Inf. Theory*, 50(10):2473–2486, 2004.

[53] I. Kammoun and J.-C. Belfiore. A new family of grassmann space-time codes for non-coherent MIMO systems. *IEEE Commun. Lett.*, 7(11):528–530, 2003.

[54] A. Kerdock. Studies of low-rate binary codes (ph.d. thesis abstr.). *IEEE Trans. Inf. Theory*, 18(2):316–316, 1972.

[55] H. Khalil, K. Rose, and S. L. Regunathan. The asymptotic closed-loop approach to predictive vector quantizer design with application in video coding. *IEEE Trans. Image Process.*, 10(1):15–23, 2001.

[56] J. H. Kim, W. Zirwas, and M. Haardt. Efficient feedback via subspace-based channel quantization for distributed cooperative antenna systems with temporally correlated channels. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008.

[57] S. K. Kim. *Group Theoretical Methods and Applications to Molecules and Crystals*. Cambridge University Press, 1999.

[58] T. Kim, D. J. Love, B. Clerckx, and S. J. Kim. Differential rotation feedback MIMO system for temporally correlated channels. In *Proc. of IEEE Global Telecom. Conf.*, pages 1–5, 2008.

[59] A. Klappenecker and M. Roetteler. Constructions of mutually unbiased bases. *Finite Fields and Applications*, pages 137–144, 2004.

[60] A. Klappenecker and M. Roetteler. Mutually unbiased bases, spherical designs, and frames. In *Proc. of SPIE*, volume 5914, pages 1–13, 2005.

[61] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate texts in mathematics; 218. Springer, 2003.

[62] L. Liu and H. Jafarkhani. Novel transmit beamforming schemes for time-selective fading multiantenna systems. *IEEE Trans. Signal Processing*, 54(12):4767–4781, 2006.

[63] D. J. Love. Grassmannian subspace packing webpage.

[64] D. J. Love and R. W. Heath Jr. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Trans. Inf. Theory*, 51(8):2967–2976, 2005.

[65] D. J. Love and R. W. Heath Jr. Multimode precoding for MIMO wireless systems. *IEEE Trans. Signal Process.*, 53(10):3674–3687, 2005.

[66] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE J. Sel. Areas Commun.*, 26(8):1341–1365, 2008.

[67] D. J. Love, R. W. Heath Jr., and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Trans. Inf. Theory*, 49(10):2735–2747, 2003.

[68] R. E. Mahony. *Optimization algorithms on homogeneous spaces: with applications in linear system theory*. PhD thesis, Australian National University, 1994.

[69] J. H. Manton. Modified steepest descent and newton algorithms for orthogonally constrained optimisation. part i. the complex Stiefel manifold. In

Proc. of Int. Symp. on Signal Process. and its Appl., volume 1, pages 80–83, 2001.

[70] J. H. Manton. Modified steepest descent and newton algorithms for orthogonally constrained optimisation. part ii. the complex Grassmann manifold. In *Proc. of Int. Symp. on Signal Process. and its Appl.*, volume 1, pages 84–87, 2001.

[71] J. H. Manton. On the role of differential geometry in signal processing. In *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, volume 5, pages 1021–1024, 2005.

[72] K. V. Mardia. Distributions on stiefel and grassmann manifolds, and their applications. *Advances in Applied Probability*, 9(3):435–436, September 1977.

[73] B. Mielczarek and W.A. Krzymien. Influence of CSI feedback delay on capacity of linear multi-user MIMO systems. In *Proc. of IEEE Wireless Commun. and Networking Conf.*, pages 1188–1192, 2007.

[74] B. Mondal, S. Dutta, and R. W. Heath Jr. Quantization on the Grassmann manifold. *IEEE Trans. Signal Process.*, 55(8):4208–4216, 2007.

[75] B. Mondal and R. W. Heath Jr. Adaptive feedback for MIMO beamforming systems. In *Proc. of IEEE Workshop on Signal Process. Adv. in Wireless Commun.*, pages 213–217, 2004.

[76] B. Mondal, R. W. Heath Jr., and L. W. Hanlen. Quantization on the grassmann manifold: applications to precoded MIMO wireless systems. In *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, volume 5, pages 1025–1028, 2005.

[77] B. Mondal, T. A. Thomas, and M. Harrison. Rank-independent codebook design from a quaternary alphabet. In *Proc. of Asilomar Conf. on Signals, Systems and Computers*, pages 297–301, 2007.

[78] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Trans. Inf. Theory*, 49(10):2562–2579, 2003.

[79] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE J. Sel. Areas Commun.*, 16(8):1423–1436, 1998.

[80] A. Nilsson, E. Tell, D. Wiklund, and D. Liu. Design methodology for memory-efficient multi-standard baseband processors. In *Proc. of Asia-Pacific Conf. on Commun.*, pages 28–32, 2005.

[81] T. Pande, D. J. Love, and J. V. Krogmeier. Reduced feedback MIMO-OFDM precoding and antenna selection. *IEEE Trans. Signal Process.*, 55(5):2284–2293, 2007.

[82] A. Papadogiannis, D. Gesbert, and E. Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. In *Proc.*

of IEEE Int. Conf. on Commun., pages 4033–4037, 2008.

[83] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst. A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. *IEEE Trans. Commun.*, 53(1):195–202, 2005.

[84] M. Planat, H. Rosu, S. Perrine, and M. Saniga. A survey of finite algebraic geometrical structures underlying mutually unbiased quantum measurements. *Foundations of Physics*, 36(11):1662–1680, 2006.

[85] V. Raghavan, R. W. Heath Jr., and A. M. Sayeed. Systematic codebook designs for quantized beamforming in correlated MIMO channels. *IEEE J. Sel. Areas Commun.*, 25(7):1298–1310, 2007.

[86] N. Ravindran and N. Jindal. Limited feedback-based block diagonalization for the MIMO broadcast channel. *IEEE J. Sel. Areas Commun.*, 26(8):1473–1482, 2008.

[87] J. C. Roh and B. D. Rao. Design and analysis of MIMO spatial multiplexing systems with quantized feedback. *IEEE Trans. Signal Process.*, 54(8):2874–2886, 2006.

[88] J. C. Roh and B. D. Rao. Transmit beamforming in multiple-antenna systems with finite rate feedback: a VQ-based approach. *IEEE Trans. Inf. Theory*, 52(3):1101–1112, 2006.

[89] W. Rossmann. *Lie Groups: An Introduction Through Linear Groups*. Number 5 in Oxford Graduate Texts in Mathematics. Oxford University Press, 2002.

[90] D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig. QAM codebooks for low-complexity limited feedback MIMO beamforming. In *Proc. of IEEE Int. Conf. on Commun.*, pages 4162–4167, 2007.

[91] Samsung. Downlink MIMO for EUTRA. *3GPP TSG WG1 Meeting 44, R1-060335*, 2006.

[92] A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM J. Control Optim.*, 48(1):56–76, Jan. 2009.

[93] S. Savazzi and U. Spagnolini. Optimizing training lengths and training intervals in time-varying fading channels. *IEEE Trans. Signal Process.*, 57(3):1098–1112, 2009.

[94] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath. Optimal designs for space-time linear precoders and decoders. *IEEE Trans. Signal Process.*, 50(5):1051–1064, 2002.

[95] S. Shamai, O. Simeone, O. Somekh, and H.V. Poor. Joint multi-cell processing for downlink channels with limited-capacity backhaul. In *Proc. of Info. Theory and Appl. Workshop*, pages 345–349, 2008.

[96] O. Simeone, O. Somekh, G. Kramer, H.V. Poor, and S. Shamai. Cellular systems with multicell processing and conferencing links between mobile

stations. In *Proc. of Info. Theory and Appl. Workshop*, pages 361–365, 2008.

[97] O. Simeone, O. Somekh, H.V. Poor, and S. Shamai. Distributed MIMO in multi-cell wireless systems via finite-capacity links. In *Proc. of Int. Symp. on Commun., Control and Signal Process.*, pages 203–206, 2008.

[98] C. Simon, R. de Francisco, D. T. M. Slock, and G. Leus. Feedback compression for correlated broadcast channels. In *Proc. of IEEE Symp. on Commun. and Veh. Technol. in the Benelux*, pages 1–4, 2007.

[99] B. Song and M. Haardt. Effects of imperfect channel state information on achievable rates of precoded multi-user MIMO broadcast channels with limited feedback. In *Proc. of IEEE Int. Conf. on Commun.*, pages 1–5, 2009.

[100] A. S. Spanias. Speech coding: a tutorial review. *Proc. IEEE*, 82(10):1541–1582, 1994.

[101] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt. Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Trans. Signal Process.*, 52(2):461–471, 2004.

[102] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer. Designing structured tight frames via an alternating projection method. *IEEE Trans. Inf. Theory*, 51(1):188–209, 2005.

[103] J. Zhang, R. W. Heath Jr., M. Kountouris, and J. G. Andrews. Mode switching for the multi-antenna broadcast channel based on delay and channel quantization. *EURASIP Journal on Advances in Signal Processing*, 2009(802548):15, 2009.

[104] L. Zheng and D. N. C. Tse. Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inf. Theory*, 48(2):359–383, 2002.

[105] L. Zheng and D. N. C. Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inf. Theory*, 49(5):1073–1096, 2003.

[106] S. Zhou and B. Li. BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers. *IEEE Trans. Signal Process.*, 54(5):1653–1665, 2006.

Vita

Takao Inoue received his B.S. and M.S. degrees in electrical engineering from Oregon State University, Corvallis, OR, in 1996 and 1998 respectively. In 1998, he joined Motorola Inc. where he was involved in ADSL modem development and ITU standardization for G.dmt. From 2004, he was a consultant to Cybernetics Technology Inc., Tokyo, Japan, to develop an FPGA based signal processing accelerator. From 2005 to 2007, he was the co-founder and chief technology officer of Fish Technologies Inc., Tokyo, Japan, specializing in next generation wireless system prototyping and consulting.

He was the recipient of 2004 IEEE MTT-S Meritorious Service Award for his contributions to MTT-S sponsored conferences. He received the best student paper award at the 2009 IEEE Radio and Wireless Symposium.

Permanent address: 3-1-26 Tamai-cho Toyonaka-city
Osaka, 560-0026, Japan

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.