AN INCREMENTAL GRASSMANNIAN FEEDBACK SCHEME FOR LINEARLY PRECODED SPATIAL MULTIPLEXING MIMO SYSTEMS

Ahmed Medra and Timothy N. Davidson

Department of Electrical and Computer Engineering
McMaster University, Hamilton, Ontario, Canada

ABSTRACT

An effective strategy for transmitter adaptation on slow block-fading multiple-input multiple-output (MIMO) links is for the receiver to inform the transmitter of the subspace over which transmission should take place, and for the transmitter to allocate power uniformly over that subspace. The design of a feedback scheme to implement this strategy can be viewed as a (lossy) source compression problem on a Grassmannian manifold. Memoryless vector quantization on each fading block is one approach to that problem, but it neglects any correlation between blocks. In some recent work, several approaches have been proposed to take advantage of this correlation. In this paper we propose an alternative technique that leverages existing Grassmannian codebooks from memoryless schemes and employs a quantized form of geodesic interpolation. Distinguishing features of the proposed technique include the fact that it only requires a single codebook, and the fact that it enables the step length of the geodesic interpolation to be adapted to the channel realization, rather than the channel statistics. In some straightforward simulation experiments, the proposed approach provides better performance than an existing scheme.

1. INTRODUCTION

In the design of systems for communicating over slow block-fading channels, the provision of a feedback channel offers the opportunity to enhance the quality-of-service by enabling adaptation of the transmitter to the channel state [1]. In such systems, the receiver typically performs (lossy) source compression on the information to be fed back to the transmitter so that the communication resources allocated to the feedback channel can be suitably constrained [2]. In multiple-input multiple-output (MIMO) point-to-point links, a simplified strategy that has proven to be effective is for the receiver to inform the transmitter of the directions in which transmission should take place, and for the transmitter to allocate power uniformly over these directions. This is consistent with results for sets of parallel subchannels, where the gap to capacity of a scheme that employs constant power loading over the subchannels with significant gain is small [3]. In the case of uniform power loading over given directions, many communication objectives depend on the subspace spanned by the directions of transmission, rather than the directions themselves. Since subspaces of a given dimension can be represented by points on the Grassmannian manifold [4], this manifold is the natural setting for the source compression problem at the receiver [2]. In particular, memoryless vector quantization schemes on the Grassmannian manifold form the basis of a number of proposals for including limited feedback in standards [2, 5].

This work was supported in part by NSERC and the Canada Research Chairs program.

Although limited feedback schemes based on memoryless vector quantization on the Grassmannian manifold offers substantial performance gains at reasonably low feedback rates, in the case of block fading channels with significant correlation between the blocks, the introduction of memory into the source compression scheme offers the potential for improved fidelity and lower feedback rates [6]. The goal of this paper is to describe an intuitively-motivated incremental Grassmannian quantization technique for exploiting temporal correlation. Like several recently proposed schemes [7–9], some of the intuition behind the proposed technique is derived from the structure of iterative algorithms for optimization over the Grassmannian manifold [4, 11]. In particular, in the proposed scheme the incremental information fed back to the transmitter implicitly informs the transmitter of the direction and the length of the geodesic update that it should perform on its current transmission subspace. The direction is captured by the of an element in a conventional Grassmannian codebook for a memoryless scheme. As such, the proposed method requires only a single codebook, and this codebook can be one designed using existing approaches to codebook design for the memoryless case.

Geodesic updates for beamforming-based systems have been considered in [9]. Geodesic updates for spatial multiplexing were considered in the concurrent work reported in [10], but the nature of the geodesic update proposed herein is somewhat different from that in [10]. In particular, in the proposed approach a fraction of the bit budget is reserved for the length of the geodesic update, and hence the length of this step can be adapted to the channel realization. The incremental Grassmannian feedback scheme for spatial multiplexing in [8] is somewhat different in the sense that the increments don't necessarily remain on the manifold and are projected back to the manifold using a QR decomposition. In our straightforward simulation experiments, the proposed scheme offers some performance advantages over that in [8], even when it uses a smaller codebook. This is in addition to the advantage that it requires only a single codebook. Those results need to be balanced, however, against the different structure of the computational requirements of the proposed algorithm.

The incremental Grassmannian feedback scheme proposed in this paper is formulated in the context of spatial multiplexing, but its simple structure suggests that the principles of the proposed method are amenable to application in other contexts, including the MIMO downlink.

2. SPATIAL MULTIPLEXING WITH LINEAR UNITARY PRECODING UNDER LIMITED FEEDBACK

Consider a narrowband slow block-fading MIMO link with M transmit antennas and N receive antennas, and let $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ denote the channel matrix for the current block. The principles that underlie the considered class of limited feedback schemes are that the receiver can use its knowledge of the channel to select the number of symbols that the transmitter should transmit per channel use, which we denote by $K \leq \min\{M, N\}$, a matrix $\boldsymbol{P} \in \mathbb{C}^{M\times K}$ with orthonormal columns with which the transmitter is to linearly precode its symbols, and an appropriate coding and modulation scheme for the symbols. The receiver informs the transmitter of its choice of K, $\boldsymbol{P}$ and the modulation and coding scheme using a limited number of feedback bits. Then, for each channel use in the given fading block, the transmitter synthesizes a transmitted signal vector of the form $\boldsymbol{x} = \sqrt{E_s/K}\,\boldsymbol{P}\boldsymbol{s}$, where $\boldsymbol{s}$ denotes a vector of K symbols with $E\{\boldsymbol{s}\boldsymbol{s}^H\} = \boldsymbol{I}$, and E_s denotes the total transmitted energy per channel use. The received signal vector then takes the form

$$\boldsymbol{y} = \sqrt{\frac{E_s}{K}}\,\boldsymbol{H}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{v}$ denotes the additive noise, which be assumed to be zero-mean, circular complex Gaussian, with covariance $E\{\boldsymbol{v}\boldsymbol{v}^H\} = \sigma^2\boldsymbol{I}$. The constraint that the precoder $\boldsymbol{P}$ has orthonormal columns implicitly restricts the precoder to uniform power loading, but the gap to capacity is typically small; cf. [3].

Since $\boldsymbol{P}$ is a matrix, a critical step in the development of a limited feedback scheme is to devise a way in which the receiver's choice of $\boldsymbol{P}$ can be reconstructed at the transmitter using only a small amount of feedback. A key observation in devising such a scheme is to recognize that for a system of the form in (1) in which $\boldsymbol{P} \in \mathbb{C}^{M\times K}$ has orthonormal columns, a number of common communication objectives, including the Gaussian mutual information, depend on the subspace spanned by the columns of $\boldsymbol{P}$, rather than the columns themselves. Since subspaces of dimension $K < M$ in $\mathbb{C}^{M\times K}$ can be represented by points on a Grassmannian manifold [4], that manifold is the natural setting in which to devise a feedback scheme for $\boldsymbol{P}$. In the remainder of the paper we will consider the simplified case in which K is fixed. Although there are natural ways to extend the methodology, our focus on this case will enable an efficient exposition of the principles of the approach.

3. MEMORYLESS GRASSMANNIAN LIMITED FEEDBACK

As pointed out above, the problem of devising a low-rate feedback scheme for the unitary precoder $\boldsymbol{P}$ can often be viewed as a source compression problem on a Grassmannian manifold [12]. In scenarios in which the fading blocks are largely uncorrelated, a simple approach is to neglect the correlation between blocks and employ the principles of memoryless vector quantization [6] on the manifold. In such systems, for a given $K < M$, both the receiver and the transmitter are provided with a codebook $\mathcal{F} = \{\boldsymbol{F}_i\}_{i=1}^{L}$ containing $L = 2^B$ (tall) matrices of size $M \times K$ with orthonormal columns, each of which is a Grassmannian representative of the subspace it spans. The receiver selects the precoder $\boldsymbol{P}$ from the elements of the codebook, and sends the B-bit index of this precoder to the transmitter. As such, there are two key design aspects: the off-line process of constructing of the codebook, and the on-line process that is used to select the most appropriate precoder from the codebook [2].

An advantage of the technique that will be proposed herein for incorporating memory into Grassmannian limited feedback schemes is that the underlying principles are not dependent on these choices. However, for simplicity in our discussion we will focus on the unstructured case in which the elements of the channel matrix $\boldsymbol{H}$ are assumed to be i.i.d. zero-mean complex Gaussian random variables. In that case, it can be shown that the elements of the codebook should be well separated on the manifold [2, 12]. The distance metric on the manifold that is appropriate for quantifying the notion of separation is dependent on the performance metric for the limited feedback scheme. In the case in which the performance metric is the Gaussian mutual information, the Fubini-Study distance, $d_{\text{FS}}(\boldsymbol{F}_i, \boldsymbol{F}_j) = \arccos\bigl|\det\bigl(\boldsymbol{F}_j^H\boldsymbol{F}_i\bigr)\bigr|$, is an appropriate distance metric for codebook construction. Given the availability of a codebook $\mathcal{F}$ at both the transmitter and receiver, the on-line precoder selection process performed by the receiver is simply to select the precoder in the codebook that maximizes the chosen performance metric. In the case that the performance metric is the Gaussian mutual information, that problem can be written as

$$\boldsymbol{P} = \arg\max_{\boldsymbol{F}_i\in\mathcal{F}} \log\det\bigl(\boldsymbol{I}_K + \rho\boldsymbol{F}_i^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{F}_i\bigr), \tag{2}$$

where $\rho = E_s/(M\sigma^2)$. As mentioned above, what is actually fed back to the transmitter is the index of the optimal $\boldsymbol{F}_i$ in (2). The transmitter obtains $\boldsymbol{P}$ by consulting its local codebook.

4. TOPOLOGY OF GRASSMANNIAN MANIFOLD

The intuition that underlies the proposed technique for incorporating memory into Grassmannian feedback schemes resides in the topology of the Grassmannian manifold. In particular, given a point on the manifold represented by $\boldsymbol{F}_i$, one can move along the manifold to a point represented by $\boldsymbol{F}_j$ by following the geodesic, which is analogous to a great circle on a sphere. To write an expression for that geodesic, we note that the tangent to the manifold at the point $\boldsymbol{F}_i$ in the direction of the point $\boldsymbol{F}_j$ is [13]

$$\boldsymbol{\Delta}(\boldsymbol{F}_i, \boldsymbol{F}_j) = \bigl(\boldsymbol{I}_M - \boldsymbol{F}_i\boldsymbol{F}_i^H\bigr)\boldsymbol{F}_j\bigl(\boldsymbol{F}_i^H\boldsymbol{F}_j\bigr)^{-1}. \tag{3}$$

If we let $\boldsymbol{\Delta}(\boldsymbol{F}_i, \boldsymbol{F}_j) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\boldsymbol{H}}$ denote the compact singular value decomposition of $\boldsymbol{\Delta}(\boldsymbol{F}_i, \boldsymbol{F}_j)$, then the points on the geodesic from $\boldsymbol{F}_i$ to $\boldsymbol{F}_j$ can be written as [4]

$$\boldsymbol{F}(t) = \boldsymbol{F}_i\boldsymbol{V}\cos(\boldsymbol{\Sigma}t)\boldsymbol{V}^H + \boldsymbol{U}\sin(\boldsymbol{\Sigma}t)\boldsymbol{V}^H, \tag{4}$$

for $t \in [0, 1]$.

5. PROPOSED INCREMENTAL GRASSMANNIAN LIMITED FEEDBACK SCHEME

A straightforward way to take advantage of correlation between fading blocks is to consider an incremental quantization scheme. Rather than sending information regarding the actual precoder to be employed, as it does in memoryless schemes, in an incremental scheme the receiver sends information that enables the transmitter to update its current precoder. The potential of incremental quantization schemes in limited feedback applications has become apparent of late, and a number of different updating strategies have been considered [7–10]. The proposed incremental quantization scheme is based on geodesic updates, but rather than constructing a codebook in the tangent space, we compute the directions of the updates from the elements of a conventional Grassmannian codebook for memoryless quantization.

In the proposed incremental feedback scheme, the B bits that are available for feedback are partitioned into two sets of size B_{cb} and B_{step}, respectively. The set of B_{cb} bits is used to index the elements of a Grassmannian codebook $\mathcal{F}_{\text{cb}}$ of size $L_{\text{cb}} = 2^{B_{\text{cb}}}$ from

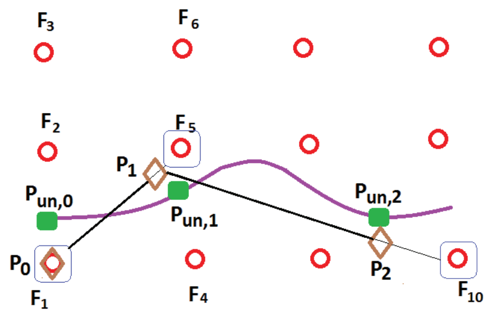

Fig. 1. A pictorial representation of the proposed technique. The points marked $\boldsymbol{P}_{\text{un},n}$ denote the sequence of precoders that would be chosen if the feedback were unlimited. The circles denote the points $\boldsymbol{F}_i$ in the Grassmannian codebook (which do not have to be uniformly spaced), and the points $\boldsymbol{P}_n$ denote the precoders generated by the technique. Note that the Grassmannian manifold is compact, and hence the edge effects that appear in this pictorial representation do not arise in practice.

a memoryless scheme, and and the set of B_{step} bits is used to index a quantization of the interval $[0, 1]$, $\mathcal{T} = \{t_1, \ldots, t_{L_{\text{step}}}\}$, where $L_{\text{step}} = 2^{B_{\text{step}}}$. To describe the principle behind the method, let $\boldsymbol{P}_n$ denote the precoder for block n. The precoder for the next block is obtained by informing the transmitter to take a step of size $t_n \in \mathcal{T}$ along the geodesic in the direction of an element $\boldsymbol{F}_n$ of $\mathcal{F}_{\text{cb}}$. The initial precoder, $\boldsymbol{P}_0$ is determined using conventional memoryless quantization using the codebook $\mathcal{F}_{\text{cb}}$. This principle is illustrated in the conceptual diagram in Fig. 1, where $\boldsymbol{P}_0$ is obtained using memoryless quantization, $\boldsymbol{P}_1$ is obtained by taking a step from $\boldsymbol{P}_0$ in the direction of $\boldsymbol{F}_5$, and $\boldsymbol{P}_2$ is obtained by taking a step from $\boldsymbol{P}_1$ in the direction of $\boldsymbol{F}_{10}$.

To formalize the procedure, let us consider the case in which the performance metric is the Gaussian mutual information,

$$\text{GMI}(\boldsymbol{P}) = \log\det\bigl(\boldsymbol{I}_K + \rho \boldsymbol{P}^H \boldsymbol{H}^H \boldsymbol{H} \boldsymbol{P}\bigr). \tag{5}$$

The algorithm proceeds as follows:

1. *Initialization:* The receiver selects the initial precoder $\boldsymbol{P}_0$ as $\boldsymbol{P}_0 = \arg\max_{\boldsymbol{F}_i \in \mathcal{F}_{\text{cb}}} \text{GMI}(\boldsymbol{F}_i)$, and feeds back the corresponding index.
2. *Incremental updates:* Given $\boldsymbol{P}_n$, the precoder for the nth fading block, the receiver uses (3) to construct the tangent $\boldsymbol{\Delta}(\boldsymbol{P}_n, \boldsymbol{F}_j)$ for each $\boldsymbol{F}_j \in \mathcal{F}_{\text{cb}}$. For each tangent, the receiver considers steps along the corresponding geodesic of size $t_i \in \mathcal{T}$, $i = 1, \ldots, L_{\text{step}}$. Using (4), this process generates $L_{\text{cb}} L_{\text{step}} = 2^B$ candidate precoders, which we denote by $\boldsymbol{G}_i \in \mathcal{G}_n$. The receiver then determines
$$\boldsymbol{P}_{n+1} = \arg\max_{\boldsymbol{G}_i \in \mathcal{G}_n} \text{GMI}(\boldsymbol{G}_i), \tag{6}$$
and feeds back indices that enable the transmitter to construct $\boldsymbol{P}_{n+1}$. These indices are the index of the direction $\boldsymbol{F}_i \in \mathcal{F}_{\text{cb}}$ and the step size $t_i \in \mathcal{T}$ that generated $\boldsymbol{P}_{n+1}$ in (6).

Although this incremental approach is based on a rather simple concept, it has a number of interesting properties.

a. The codebook that is used for the directions of the incremental update can also be used in the memoryless quantization process in the initialization step. As such, there is no additional storage requirement. (Actually, since $B_{\text{cb}} < B$, the codebook is actually smaller than the codebook for the same feedback rate in the memoryless case.) Furthermore, this enables a system designer to take advantage of existing codebooks for memoryless quantization [2,5,12]. This alleviates some of the difficulties associated with constructing codebooks for increments; cf. [8,9].
b. Since the proposed technique is based on updates along geodesics, the precoder in (6) lies on the manifold. As a result, the additional projection step employed in the method in [8] is not required.
c. In the proposed technique a fraction of the feedback budget is reserved for the size of the geodesic step, and this enables the step size to be adapted to the channel realization. This is in contrast to the methods in [8, 10], where the corresponding notions of step size are adapted to the channel statistics.
d. In the proposed scheme the quantization $\mathcal{T}$ of the step size can be adapted to the temporal correlation of the channel. Furthermore, for systems in which the feedback is subject to delay, the proposed technique can be modified to mitigate the effects of delay by performing prediction. A simple way to do so is to scale the value of the step size.
e. Given its ability to interpolate between points in the codebook for memoryless feedback (cf. Fig. 1), the performance of the proposed technique is less sensitive to the quality of the codebook than conventional memoryless feedback schemes. Indeed, if the channel is constant, this interpolation yields a feedback scheme with higher resolution than the underlying memoryless scheme.

The computational cost of the process in (6) in the proposed approach is the same as that of the corresponding step precoder selection in a memoryless quantization scheme with a codebook of 2^B elements. However, the proposed approach does incur additional computation cost in computing the $L_{\text{cb}} = 2^{B_{\text{cb}}}$ tangents, $\boldsymbol{\Delta}(\boldsymbol{P}_n, \boldsymbol{F}_j)$ and the corresponding geodesics. Like other incremental methods [8–10], the proposed technique is sensitive to errors in the feedback channel, but given the low rate of feedback these can be mitigated by error control coding, and the receiver can initiate a reset in the incremental scheme if its rate of repeat requests suggests that feedback errors have occurred.

6. SIMULATION RESULTS

In this section, we examine the performance of the proposed technique in a simple richly-scattered block-fading environment with $M = 4$, $N = 2$, and $K = 2$. The evolution of the channel matrix H is modeled by a first-order Gauss-Markov process, in which the channel at the nth block is

$$H_n = \alpha H_{n-1} + \sqrt{1-\alpha^2}\,\tilde{H}_n, \tag{7}$$

where $\tilde{H}_n$ has i.i.d. zero mean complex Gaussian entries of unit variance, and the temporal correlation is related to Jakes model via $\alpha = J_0(2\pi\beta)$, where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, and β is the normalized Doppler frequency. We evaluate the proposed technique in terms of the Gaussian mutual information, and we compare its performance to that of a memoryless scheme that employs the same number of feedback bits, the idealized case of unlimited feedback, and the differential approach proposed in [8]. For the memoryless scheme we employ a codebook of $2^4 = 16$ precoders designed using a tailored version of the smoothed optimization technique outlined in [14], and for the method in [8] we use a codebook of 16 precoders provided by the authors of that paper. For

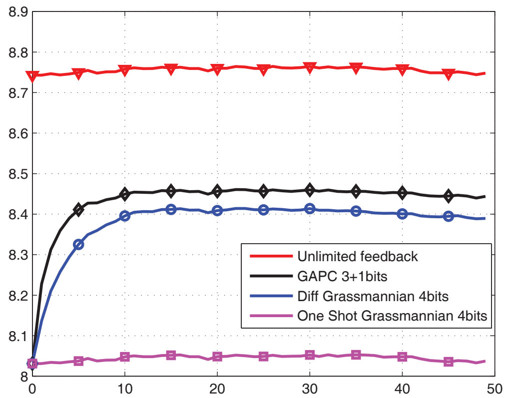

Fig. 2. Gaussian mutual information against channel use for the proposed approach (denoted GAPC), the memoryless approach (one-shot), the approach in [8] (Diff) and the case of unlimited feedback, in an environment with $\alpha = 0.997$.

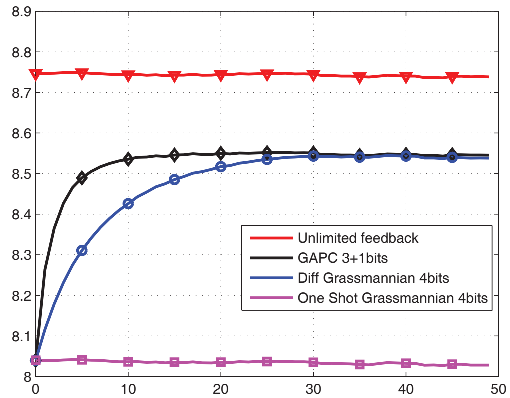

Fig. 3. Gaussian mutual information against channel use for the proposed approach (denoted GAPC), the memoryless approach (one-shot), the approach in [8] (Diff) and the case of unlimited feedback, in an environment with $\alpha = 0.988$.

the proposed technique, we reserve one bit for the quantization of the step size, and we consider a codebook of $2^3 = 8$ precoders designed using the tailored version of the method in [14].

In Figs 2 and 3 we provide the Gaussian mutual information achieved in scenarios with $\alpha = 0.997$ and 0.988, respectively. For a carrier frequency of 2.5 GHz and a feedback interval of 5 ms these values correspond to speeds of 1.5 and 3 km/h, respectively [15]. The quantization $\mathcal{T}$ of the step size was chosen to be $\{0.1, 0.5\}$. These figures show that in these slowly varying channels, the proposed incremental quantization technique is able to take a step across the gap between the performance of the memoryless quantization scheme and that of a system with unlimited feedback. In the case of the slower varying channel, the proposed technique converges faster than that in [8], and in the faster varying channel, the proposed technique provides a modest performance gain over the technique in [8].

7. ACKNOWLEDGEMENT

We would like to thank Taejoon Kim for providing us with the codebook of precoders used in [8].

8. REFERENCES

[1] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[2] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008

[3] W. Yu and J. M. Cioffi, "Constant power waterfilling: Performance bound and low-complexity implementation," *IEEE Trans. Commun.*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[4] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Oct. 1998.

[5] B. Clerckx, Y. Zhou and S. Kim, "Practical codebook design for limited feedback spatial multiplexing," in *Proc. IEEE Int. Conf. Commun.*, Beijing, 2008.

[6] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer, 1992.

[7] J. Yang and D. B. Williams, "Transmission subspace tracking for MIMO systems with low-rate feedback," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1629–1639, Aug. 2007.

[8] T. Kim, D. J. Love, and B. Clerckx, "MIMO systems with limited rate differential feedback in slowly varying channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1175–1189, Apr. 2011.

[9] T. Inoue and R. W. Heath Jr., "Grassmannian predictive coding for limited feedback multiuser MIMO systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Prague, May 2011. See also: `http://arxiv.org/abs/1105.5782`

[10] D. Zhu,Y. Zhang, G. Wang and M. Lei, "Grassmannian subspace prediction for precoded spatial multiplexing MIMO with delayed feedback," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 555–558, Oct. 2011.

[11] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

[12] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing system," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–1976, Aug. 2005

[13] I. S. Dhillon, T. Stromher, R. W. Heath Jr., and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Exper. Math.*, vol. 17, no. 1, pp. 9-35, 2008.

[14] R. H. Gohary and T. N. Davidson. "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, Mar. 2009.

[15] "IEEE P802.16m-2008 draft standard for local and metropolitan area network," IEEE Standard 802.16m, 2008.