Optimal Design of Noncoherent Cayley Unitary Space-Time Codes

Jibing Wang
Qualcomm Inc.
San Diego, CA 92121

Xiaodong Wang
EE Dept., Columbia Univ.
New York, NY 10027

Mohammad Madihian
NEC Labs
Princeton, NJ 08540

Abstract

The Cayley unitary (CU) codes constitute a systematic way of constructing unitary space-time modulations for noncoherent MIMO communications. For MIMO systems employing CU codes, there is no explicit expression for block (or bit) error probabilities. Hence, deterministic optimization tools cannot be employed to design the optimal CU codes. In this work, we propose to optimize the design of CU codes through simulation-based optimization techniques, in particular, stochastic approximation together with gradient estimation. The proposed methodology can be employed to design optimal CU codes under the maximum likelihood decoding or the suboptimal linearized sphere decoding. Simulation results show that new CU codes obtained by the proposed design significantly outperform those in the literature designed by minimizing the expected distance between codeword pairs. The new CU codes also enjoy comparable performance over training-based designs.

1 Introduction

For fast fading noncoherent MIMO channels, in particular the block fading channels [1][2], unitary space-time signals have been proposed and it has been shown that by combining them with channel coding a high fraction of the channel capacity can be achieved. It is shown in [3] that unitary signals also minimize the asymptotic union bound on the block error rate for equal energy signals. The design of unitary space-time codes have been discussed in various works.

In this paper, we focus on the optimal design of noncoherent Cayley unitary codes. For MIMO systems employing Cayley unitary space-time modulation, no explicit expressions exit for the block (or bit) error probabilities. Therefore, deterministic optimization tools are hard to use. In this paper, we propose to design optimal CU codes using simulation-based optimization together with gradient estimation. We employ the score function method [4] to obtain an unbiased estimate of the gradient of block error rate with respect to the dispersion matrices. We then optimize the CU codes through the well-known Robbins-Monro algorithm. Our simulation examples show that codes obtained by the proposed method significantly outperform the existing codes.

2 Cayley Unitary Code Design

In this section we present the signal model for MIMO systems employing unitary codes, and formulate the code design problem as a constrained stochastic optimization problem.

2.1 Signal Model

Consider a MIMO system with M_T transmit antennas and M_R receive antennas. Assume that the channel is frequency non-selective and remains constant for T symbol intervals, and changes independently from one realization to another. T is the coherent interval of this block fading model (see, e.g. [1]). The input-output relationship can be written in matrix form as

$$\boldsymbol{Y} = \sqrt{\frac{\rho T}{M_T}} \boldsymbol{X}\boldsymbol{H} + \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{Y}$ is the $T \times M_R$ matrix of the received signal, $\boldsymbol{X}$ is the $T \times M_T$ matrix of the transmitted signal, $\boldsymbol{W}$ is the $T \times M_R$ matrix of the additive white Gaussian noise, and $\boldsymbol{H}$ is the $M_T \times M_R$ MIMO channel matrix. Here, ρ is the expected SNR at each receive antenna regardless of the number of transmit antennas. We restrict ourselves to a Rayleigh fading scenario, therefore, the elements of $\boldsymbol{H}$ are composed of i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unit variance.

In the absence of the channel state information $\boldsymbol{H}$ at the receiver, conditioned on the transmitted signal matrix $\boldsymbol{X}$, the received signal $\boldsymbol{Y}$ has independent and identically distributed columns. At each antenna, the received T symbols are complex Gaussian random variables with zero mean and covariance matrix given by [1] $\boldsymbol{\Lambda} = \boldsymbol{I}_T + \frac{\rho T}{M_T}\boldsymbol{X}\boldsymbol{X}^H$. Therefore, conditioned on $\boldsymbol{X}$, the received signal $\boldsymbol{Y}$ has the following probability density function (pdf)

$$p(\boldsymbol{Y}|\boldsymbol{X}) = \frac{\exp\left(-\mathrm{Tr}\left\{\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}\boldsymbol{Y}^H\right\}\right)}{\pi^{TM_R}\det^{M_R}(\boldsymbol{\Lambda})}. \quad (2)$$

Unitary Space-Time Modulation: For noncoherent MIMO channels, it is shown in [1][2] that a capacity-achieving random signal matrix for the channel given by (1) may be constructed as a product $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{V}$, where $\boldsymbol{\Phi}$ is an isotropically distributed $T \times M_T$ matrix whose columns are orthonormal, and $\boldsymbol{V}$ is an independent $M_T \times M_T$ real, nonnegative, diagonal matrix. Furthermore, when either $T \gg M_T$, or for high SNR and $T > M_T$, the capacity achieving $\boldsymbol{V}$ is the identity matrix. Motivated by the above result, the so-called unitary space-time modulation (USTM) is introduced in [5] where the codewords satisfy the following property $\boldsymbol{X}^H\boldsymbol{X} = \boldsymbol{I}_{M_T}$. In USTM, the space-time code consists of a set of $T \times M_T$ unitary matrices $\{\boldsymbol{X}_\ell\}_{\ell=0}^{L-1}$. A transmission data rate of R bits per channel use needs a constellation set of $L = 2^{RT}$ signals. The maximum likelihood (ML) decoder is given by [5]

$$\widehat{\ell} = \arg\max_{\ell=1,\cdots,L} p(\boldsymbol{Y}|\boldsymbol{X}_\ell). \quad (3)$$

In particular, for unitary space-time codes, the ML receiver reduces to [5]

$$\widehat{\ell} = \arg\max_{\ell=1,\cdots,L} \left\|\boldsymbol{Y}^H\boldsymbol{X}_\ell\right\|_F, \quad (4)$$

where $\|\cdot\|_F$ denotes Frobenius norm, i.e., $\|\boldsymbol{R}\|_F = \sqrt{\mathrm{Tr}\left\{\boldsymbol{R}\boldsymbol{R}^H\right\}}$.

Cayley Unitary Codes: In [6], a systematic way is introduced to construct unitary space-time codes for arbitrary number of antennas and at any rate. The idea is to encode data onto a skew-Hermitian matrix and then apply the Cayley transform to get a unitary matrix. In essence, the Cayley unitary (CU) codes break the data stream into substreams, and these substreams are used to parameterize the unitary matrices $\boldsymbol{X}$ to be transmitted. Suppose we break the data stream into Q substreams (we will discuss the choice of Q in Section 3.2) and use these substreams to choose $\alpha_1, \cdots, \alpha_Q$ each from a set $\mathcal{A}$ with r real scalars.[1] A CU space-time code is given by

$$\boldsymbol{X} = (\boldsymbol{I}_T + i\boldsymbol{A})^{-1}(\boldsymbol{I}_T - i\boldsymbol{A})\begin{bmatrix}\boldsymbol{I}_{M_T}\\ 0\end{bmatrix}, \quad (5)$$

[1]For example, we can choose $\mathcal{A}$ to be the set of r-PAM constellations. We will discuss this in Section 3.2.

where the Hermitian matrix $\boldsymbol{A}$ is given by $\boldsymbol{A} = \sum_{q=1}^{Q}\alpha_q\boldsymbol{A}_q$, where $\boldsymbol{A}_q$ are fixed complex Hermitian matrices that specify the code. Note that when the dispersion matrices $\{\boldsymbol{A}_q\}_{q=1}^{Q}$ are Hermitian and $\{\alpha_q\}_{q=1}^{Q}$ are real, the Cayley transform in (5) guarantees that the codeword $\boldsymbol{X}$ is unitary [7][6]. Since we transmit Q substreams $\alpha_1, \cdots, \alpha_Q$ over T channel uses with each α_q taking on one of r possible values, the rate of the code is $R = (Q/T)\log_2 r$. It is shown in [6] that by further constraining that the matrix $\boldsymbol{A}$ to have some specific structures (see also Section 3.2), the Cayley unitary codes can be decoded efficiently using polynomial-time receivers such as the sphere-decoder.

3 Optimal Design of Noncoherent Cayley Unitary Codes

3.1 Problem Formulation

We consider the optimal design of noncoherent Cayley unitary (CU) codes. In [6], the CU codes are designed by maximizing the expected "distance" between codeword pairs. However, the methods proposed in [6] do not necessarily lead to unitary space-time codes with good block or bit error performance. In fact, the CU codes designed in [6] is outperformed by the optimized training-based scheme. In [5] the codes are designed or searched by optimizing certain bounds on the pairwise error probability (PEP). Still in general it is not true that the codes optimized with respect to the worst case PEP will end up with optimum bit or block error performance. Unfortunately, the average bit or block error rate is hard if not impossible to analytically characterize for arbitrary unitary space-time codes. Simulation-based optimization techniques turns out to be powerful for this scenario. In this work, we demonstrate how to optimize the average block error probability (BLEP) for CU codes through simulation-based optimization. The bit error performance can be optimized similarly.

We denote the set of dispersion matrices as $\boldsymbol{\theta} \triangleq \{\boldsymbol{A}_q, q = 1, \ldots, Q\}$. We define a vector $\boldsymbol{\alpha}$ that corresponds to the information streams $\boldsymbol{\alpha} = [\alpha_1 \ \ldots \ \alpha_Q]^T$. The set of all r^Q possible vectors $\boldsymbol{\alpha}$ is denoted as $\mathbb{C}$. We also denote $\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ as the empirical BLEP for a given set of dispersion matrices $\boldsymbol{\theta}$, a given information symbol vector $\boldsymbol{\alpha}$, and a given received signal matrix $\boldsymbol{Y}$. Therefore, $\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = 0$ if the decoded symbol vector (denoted as $\hat{\boldsymbol{\alpha}}$) is the same as the transmitted one (i.e., $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$), and $\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = 1$ otherwise (i.e., $\hat{\boldsymbol{\alpha}} \neq \boldsymbol{\alpha}$). Given $\boldsymbol{\theta}$, the average

BLEP is obtained by

$$\Upsilon(\boldsymbol{\theta}) = \mathbb{E}\{\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\} = \int\int \gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\, p(\boldsymbol{Y}, \boldsymbol{\alpha} \mid \boldsymbol{\theta})\, d\boldsymbol{Y} d\boldsymbol{\alpha} \quad (6)$$

where $p(\boldsymbol{Y}, \boldsymbol{\alpha} \mid \boldsymbol{\theta})$ is the joint probability density function (pdf) of $(\boldsymbol{Y}, \boldsymbol{\alpha})$ for a given $\boldsymbol{\theta}$. Note that the empirical BLEP $\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ usually does not have a closed-form expression. Actually $\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ also depends on the receiver structure (ML, or sub-optimal). The design goal is to solve the following optimization problemmin$_{\boldsymbol{\theta}\in\Theta}\, \Upsilon(\boldsymbol{\theta})$, with $\Theta = \left\{\boldsymbol{\theta} : \boldsymbol{A}_q = \boldsymbol{A}_q^H, q = 1, \cdots Q.\right\}$.

3.2 Simulation-based Code Design

We have $\Upsilon(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\alpha}}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\alpha},\boldsymbol{\theta}}\{\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\}$. Note $p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta})$ is a Gaussian distribution, and it is continuously differentiable in $\boldsymbol{\theta}$, hence $\Upsilon(\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$. We have

$$\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\alpha}} \int \nabla_{\boldsymbol{\theta}} [\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\, p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta})]\, d\boldsymbol{Y}$$
$$= \mathbb{E}_{\boldsymbol{\alpha}} \int \left[\nabla_{\boldsymbol{\theta}}\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\, p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta}) + \gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}} p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta})\right] d\boldsymbol{Y}, \quad (7)$$

where we have assumed that regularity conditions hold so that the derivative and expectation can be interchanged. We can show that for ML detection, with probability one (w.p.1) we have

$$\nabla_{\boldsymbol{\theta}}\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = 0. \quad (8)$$

Substituting (8) into (7) we have $\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta}) = E_{\boldsymbol{\alpha}}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\alpha},\boldsymbol{\theta}}\{\gamma(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta})\}$.

Assume at the k-th iteration the current set of dispersion matrices is $\boldsymbol{\theta}_k$, then we perform the following steps during the next iteration to generate $\boldsymbol{\theta}_{k+1}$:

1 - Generate samples :

   1). Draw M symbol vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_M$ uniformly from the set $\mathbb{C}$.

   2). Simulate M observations $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_M$ where each $\boldsymbol{Y}_i$ is generated according to (1) using symbol vector $\boldsymbol{\alpha}_i$.

   3). Using the given decoding algorithm, decode $\boldsymbol{\alpha}_i$ based on the observations $\boldsymbol{Y}_i$. Compute the empirical BLEP $\gamma(\boldsymbol{Y}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta}_k)$.

2 - Calculate the gradient estimate: $\widehat{g}(\boldsymbol{\theta}_k) = \frac{1}{M}\sum_{i=1}^{M} \gamma(\boldsymbol{Y}_i, \boldsymbol{\alpha}_i, \boldsymbol{\theta}_k)\left[\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i, \boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_k}\right]$, where an explicit formula for $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\theta})$ can be obtained but omitted here.

3 - Update dispersion matrices:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - a_k \widehat{g}(\boldsymbol{\theta}_k), \quad (9)$$

where $a_k = \frac{c}{k}$ for some positive constant c.

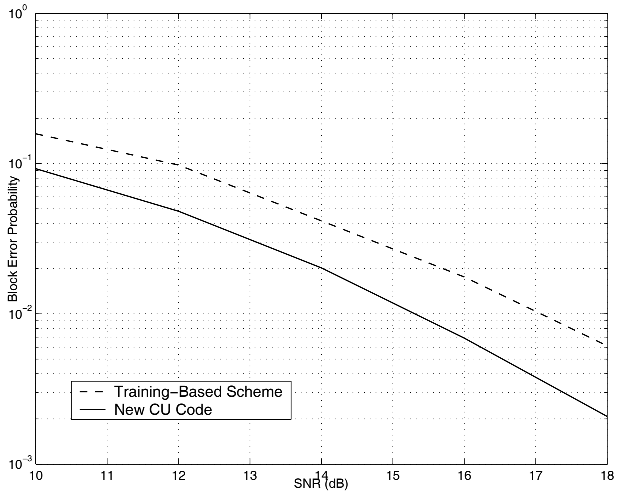

Figure 1: Example 1: $T = 2$, $M_T = 1$, $M_R = 3$, and $R = 2$.

4 Design Examples

Example 1: $T = 2$, $M_T = 1$, $M_R = 3$, and $R = 2$. In the first example, we consider the case of a single transmit antenna and three receive antennas with data rate 2 bits/sec/Hz and coherence interval $T = 2$. Fig. 1 reports the block error probability versus SNR for the new CU code and a training-based scheme. For the training-based scheme, to achieve data rate 2 bits/sec/Hz we transmit 16QAM constellations after the training phase. We also assume ML decoding for the training-based scheme. For the CU scheme, we use BPSK constellation and set $Q = 4$. We can see that the new CU code offers significant gain over the training-based scheme. The performance gain is about 1.7 dB when the BLEP is around 10^{-2}.

Example 2: $T = 4$, $M_T = 2$, $M_R = 2$, and $R = 2$. In this setting, for the training-based scheme, half of the coherence interval is used for training. For the data transmission phase, we consider two different space-time codes: the linear dispersion (LD) codes, and the threaded algebraic space-time (TAST) codes [8]. We consider both the suboptimal training-based decoder and the ML decoder. For the CU scheme, to achieve data rate 2 bits/sec/Hz, we can choose $Q = 4$ using 4-PAM constellation, or we can choose $Q = 8$ using BPSK constellation. We also provide simulation results for the CU code designed in [6] (denoted as

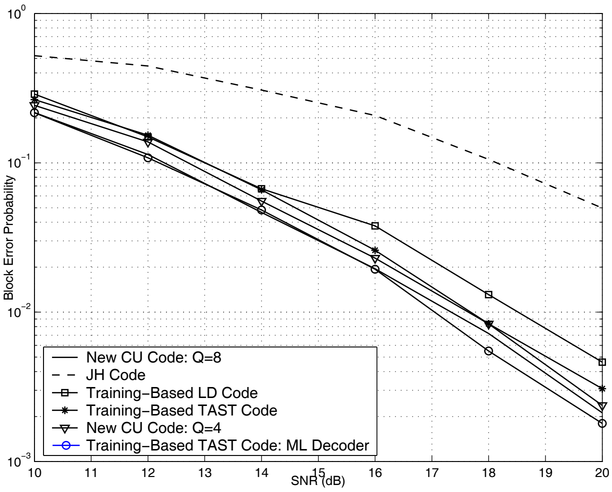

Figure 2: Example 2: $T = 4$, $M_T = 2$, $M_R = 2$, and $R = 2$.

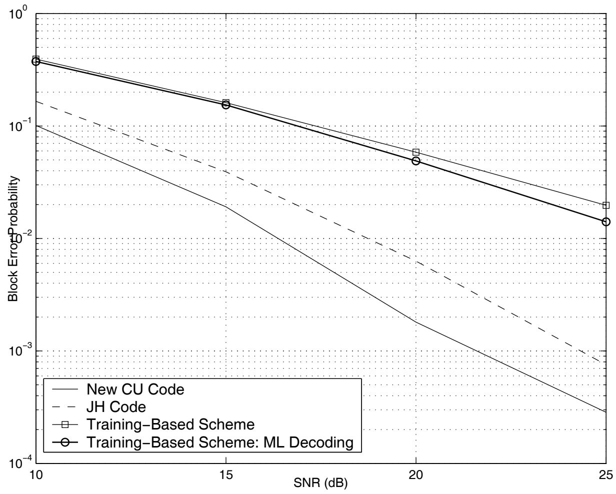

Figure 3: Example 3: $T = 5$, $M_T = 2$, $M_R = 1$, and $R = 1$.

JH code). All CU codes are decoded with ML decoding. From Fig. 2, we can see that the new CU codes perform dramatically better than the JH code. The performance gain can be as large as 5dB. The new CU code with $Q = 4$ performs similarly to the training-base TAST code (with suboptimal decoder) and better than the training-based LD code. The CU code with $Q = 8$ perform slightly better than the CU code with $Q = 4$. Note that the new CU code with $Q = 8$ also enjoys similar performance as the training-based TAST code with ML decoding.

Example 3: $T = 5$, $M_T = 2$, $M_R = 1$, and $R = 1$. In this case, as in [6], for the training-based scheme, two channel uses of each coherent interval are allocated to training. In the data transmission phase, an uncoded transmission scheme is employed, i.e., independent BPSK constellation is employed, resulting in rate 6/5. For the CU scheme, in Fig. 3 we include the new CU code generated by our algorithm and the JH code from [6]. The CU schemes are decoded using the ML decoder. We can see that the new CU code offers about 2dB gain over the JH code at BLEP 10^{-2}.

5 Conclusions

We have proposed to design Cayley unitary (CU) space-time modulations by employing simulation-based optimization with gradient estimation. We perform the gradient estimation through the score function method. We search the optimum dispersion matrices for specific scenarios taking into account the number of transmit and receive antennas, the constellation set used, and the operating SNR. Simulation results show that codes obtained by the simulation-based optimization algorithm generally outperform the CU codes designed by minimizing the expected distance between codeword pairs.

References

[1] T.L. Marzetta and B.M Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139- 157, Jan 1999.

[2] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: a geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359-383, Feb 2002.

[3] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2383-2399, Sept 2001.

[4] M.C. Fu, "Optimization via Simulation: A Review," Annals of Operations Research, Vol. 53, pp.199-248, 1994.

[5] B.M. Hochwald and T.L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543-564, Mar 2000.

[6] Y. Jing and B. Hassibi, "Unitary space-time modulation via Cayley transformation," *IEEE Trans. Signal Processing*, vol. 51, pp. 2891-2904, Nov 2003.

[7] B. Hassibi and B. M Hochwald, "Cayley differential space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1485-1503, Aug 2002.

[8] H. E. Gamal, and M. O. Damen, "Universal Space-Time Coding," *IEEE Trans. Inform. Theory,* vol. 48, pp. 1097-1119, May 2003.