A new approach to noncoherent Space-Time Block Codes

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik der
Technischen Universität Dortmund
vorgelegt von

Jens Diewald

im August 2017

Dissertation

A new approach to noncoherent Space-Time Block Codes

Fakultät für Mathematik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Detlev Hoffmann

Zweitgutachter: Prof. Thomas Unger, Ph.D.

Tag der mündlichen Prüfung: 29.09.2017

Contents

§1 Preface

Space-time block codes can be used for wireless communication in settings where multiple transmit and receive antennas are available. In these situations they allow for a significant improvement of communication. However, there are two main cases that have to be distinguished. Signals which are transmitted across the communication channel degrade before they are received at the decoder. If the decoder has precise information about this degradation it can predict how a transmitted signal is affected. In this case, the communication is referred to as *coherent*. This case has been extensively studied and many constructions are available in order to obtain corresponding codebooks. If no such information is available at the receiver, the communication is referred to as *noncoherent*. This case has received far less attention than the coherent one. In practice it corresponds to situations in which the communications channel changes too quickly for the receiver to obtain information on the degrading of transmitted signals. This mainly occurs when one of the communication parties is moving quickly, for example aboard a car, train or airplane. Another example where noncoherent communication is used is when large numbers of antennas are employed such that many signals are transmitted simultaneously. As a result, obtaining information about the degradation of each transmitted signal becomes increasingly complex and hence unfeasible from a certain point on.

There is no general theory available on how to obtain good, or even optimal noncoherent space-time block codes. These code were first systematically investigated in 1999 by Hochwald an Marzetta and they quickly restricted themselves to a special case which they referred to as unitary codes. Since then, unitary codes have been the focus of work on noncoherent space-time block codes. The problem of finding good unitary codebooks remains complicated and several suboptimal constructions have been proposed.

The first aim of this work is to establish a theory which allows the study and construction of noncoherent space-time block codes as generally as possible. In particular, the considered codebooks shall not be restricted to the case of unitary codes. Secondly, the developed theory is to be applied to introduce new classes of noncoherent space-time block codes which generalize unitary codes. For these new classes, concrete example codes will be constructed and compared to optimal unitary codebooks in simulations.

Part of the motivation for this work stems from Lie algebras and particularly from the exponential map, which provides the connection between Lie groups and Lie algebras. The first chapter starts with the introduction of some nonstandard facts from linear algebra and goes on to briefly explain this motivation as Lie algebras and Lie groups are introduced. Subsequently, the exponential map is introduced and important properties are discussed.

The second chapter provides an introduction to wireless communication in general and space-time block codes in particular. Also, the most important examples of known space-time block codes are briefly discussed.

Building on the preceding introduction, the third chapter summarizes the known results on noncoherent space-time block codes. Particularly, unitary codes are studied as they constitute the best known noncoherent codes so far. In this context, a connection between unitary codes and a packing problem on the Grassmann manifold is discussed and a numerical optimization algorithm based on this connection is presented. This algorithm can be used to obtain optimal unitary codebooks, which will be used in simulations in later chapters. Close attention is payed to the stochastic analysis of the noncoherent communications channel, which provides the basis of the study of unitary codes. For the development of a more general theory of noncoherent space-time block codes, these stochastic considerations also have to be the starting point.

In the fourth chapter, a new approach to noncoherent space-time block codes is developed. Starting with an analysis of the error probability, new criteria for designing good noncoherent space-time block codes are formulated as generally as possible. Based on these criteria, notions of reduced and equivalent codebooks are introduced and analyzed. It turns out that the problem of designing good noncoherent codebooks remains difficult. Therefore, in order to be able to systematically construct well performing codes, special cases are of interest for which the design problem can be simplified. To aid the systematic study of special cases, the notion of a distance function is introduced and it is clarified in which cases such functions are equivalent.

The fifth chapter introduces a specific class of codes which can be decoded by the same criterion that is commonly used to decode unitary space-time block codes, referred to as the *GLRT* criterion. The introduced class contains unitary codes as a special case. General codebooks arising from it are analyzed by means of the theory developed earlier and example constructions of corresponding codes are carried out. In simulations, these are compared to unitary codes that were obtained by the numerical algorithm presented at the end of the third chapter. To conclude, optimal codebooks within the considered class are characterized and their connection to unitary codes is discussed.

The appendix provides a description of the simulations which were run to compare example codes constructed in the course of this work.

I would like to express my gratitude towards my advisor Prof. Dr. Detlev Hoffmann for his guidance and support in the making of this thesis. In addition, I would like to thank Professor Frédérique Oggier for sharing her insights into the various branches of space-time block codes and her knowledge of simulations. Finally, I thank my family and my friends for supporting me throughout the writing process.

§ 2 Basics from linear algebra and Lie theory

A list of standard nomenclature being used can be found at the end of this work.

2.1 Some facts from linear algebra

This section introduces some non-standard but elementary facts from linear algebra which will be used frequently in the later chapters. To start with, some notation is fixed.

(2.1) Notation

(i) The $n \times n$ identity matrix is denoted by I_n.

(ii) Extending this notation, the $n \times m$ matrix $I_{n\times m}$ is defined by

$$(I_{n\times m})_{ij} = \begin{cases} 1 & \text{, if } i = j \\ 0 & \text{, otherwise} \end{cases} \quad \text{for } 1 \leq i \leq n \text{ and } 1 \leq j \leq m.$$

(iii) The $n \times m$ zero matrix is denoted by $0_{n\times m}$.

(iv) A diagonal $n \times n$ diagonal matrix with diagonal entries $d_1, \ldots, d_n$ is denoted by

$$\operatorname{diag}(d_1, \ldots, d_n) = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_2 \end{pmatrix}.$$

(v) The space of $n \times n$ Hermitian matrices is denoted by

$$\mathcal{H}(n) = \{X \in \mathbb{C}^{n\times n} \mid X = \overline{X}^\top\}.$$

(vi) The group of $n \times n$ unitary matrices is denoted by

$$U(n) = \{X \in \mathbb{C}^{n\times n} \mid X\overline{X}^\top = \overline{X}^\top X = I_n\}.$$

◇

Eigenvalues will play an important role throughout this work, hence a notion of the spectrum of a matrix is introduced.

(2.2) Definition
Consider a matrix $A \in \mathbb{C}^{n \times n}$, $n \in \mathbb{N}$ with eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ of respective algebraic multiplicities $\mu_1, \ldots, \mu_m \in \mathbb{N}$. The *spectrum* of A is defined as:

$$\operatorname{spec}(A) := \{(\lambda_1, \mu_1), \ldots, (\lambda_m, \mu_m)\}.$$

A spectrum $\mathcal{S} = \{(\lambda_1, \mu_1), \ldots, (\lambda_m, \mu_m)\}$ is called

(i) *real*, if all eigenvalues $\lambda_1, \ldots, \lambda_m$ are real.

(ii) *positive definite*, if all eigenvalues $\lambda_1, \ldots, \lambda_m$ are positive real numbers.

(iii) *positive semidefinite*, if all eigenvalues $\lambda_1, \ldots, \lambda_m$ are nonnegative real numbers.

◇

In the study of space-time block codes the matrix norm induced by the standard Hermitian inner product $(A, B) \mapsto \operatorname{Tr}(A\overline{B}^\top)$, commonly referred to as *Frobenius norm*, is of importance. It is also defined in [BO13]. A more detailed introduction in the context of matrix norms can be found in [SK09].

(2.3) Definition and Lemma (Frobenius norm)
The *Frobenius norm* of $A \in \mathbb{C}^{n \times m}$, $n, m \in \mathbb{N}$ is defined by

$$\|A\|_F := \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} |A_{ij}|^2}.$$

It may also be conveniently expressed by means of the matrix trace:

$$\|A\|_F^2 = \operatorname{Tr}(A\overline{A}^\top) = \operatorname{Tr}(\overline{A}^\top A).$$

The Frobenius norm is *submultiplicative*. That is, for $B \in \mathbb{C}^{m \times k}$, $k \in \mathbb{N}$ the following inequality holds:

$$\|AB\|_F \leq \|A\|_F \|B\|_F .$$

◇

Proof. The first equation is obtained by elementary computations:

$$\operatorname{Tr}(A\overline{A}^\top) = \sum_{i=1}^{n} (A\overline{A}^\top)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}\overline{A_{ij}} = \|A\|_F^2 .$$

Also $\mathrm{Tr}(\overline{A}^\top A) = \mathrm{Tr}(A\overline{A}^\top)$ holds since $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ is true for any matrices A, B for which the products AB and BA are defined.

Concerning submultiplicativity, denote the columns of $\overline{A}^\top$ by $a_1, \ldots, a_n \in \mathbb{C}^m$ and the columns of B by $b_1, \ldots, b_k \in \mathbb{C}^m$. The Cauchy–Schwarz inequality applied with respect to the canonical Hermitian inner product on $\mathbb{C}^m$ implies $\left|\overline{a_i}^\top b_j\right|^2 \leq \|a_i\|^2 \|b_j\|^2$ for $1 \leq i \leq n$, $1 \leq j \leq k$ and the norm $\|\cdot\|$ induced by the canonical Hermitian inner product. Writing out the Frobenius norm of AB then yields the assertion:

$$\|AB\|_F = \sum_{i=1}^{m} \sum_{j=1}^{k} \left|\overline{a_i}^\top b_j\right|^2 \leq \sum_{i=1}^{m} \sum_{j=1}^{k} \|a_i\|^2 \|b_j\|^2 = \left\|\overline{A}^\top\right\|_F^2 \|B\|_F^2 = \|A\|_F^2 \|B\|_F^2 .$$

□

In the following the notion of a diagonal matrix will be needed for matrices which are not necessarily square.

(2.4) Notation
A matrix $D \in \mathbb{C}^{n \times m}$, $n, m \in \mathbb{N}$ is called a *diagonal matrix* if and only if it has nonzero entries only on its main diagonal. That is, for $1 \leq i \leq n$ and $1 \leq j \leq m$ the implication $i \neq j \implies D_{ij} = 0$ holds. ◇

Using this notation, an important matrix decomposition which will be used throughout this work can be introduced.

(2.5) Definition and Lemma (Singular-Value-Decomposition (SVD))
Let $A \in \mathbb{C}^{n \times m}$, $n, m \in \mathbb{N}$ be any matrix over the complex numbers. There exist unitary matrices $U \in U(n), V \in U(m)$ and a diagonal matrix $D \in \mathbb{C}^{n \times m}$ such that $A = UDV$ holds. The diagonal elements D_{ii} for $1 \leq i \leq \min(n, m)$ are nonnegative real numbers and called the *singular values* of A.

Additionally, for any singular value σ of A, there are vectors $u \in \mathbb{C}^n$ and $v \in \mathbb{C}^m$ with $\|u\|_F = \|v\|_F = 1$ which satisfy $Av = \sigma u$ and $\overline{A}^\top u = \sigma v$. A vector u satisfying this condition is called a *left singular vector* of A, a respective vector v is called a *right singular vector* of A. ◇

The definition of the singular value decomposition along with some basic properties can be found in standard texts on numerical mathematics or matrix analysis, see for example [Bha96]. There is a connection between singular values and eigenvalues that will also be of importance.

(2.6) Proposition
Consider a matrix $A \in \mathbb{C}^{n \times m}$, $n, m \in \mathbb{N}$ with the nonzero singular values given by $\sigma_1, \dots, \sigma_r \in \mathbb{R}_{\geq 0}$ for $r \in \mathbb{N}$.

The nonzero eigenvalues of $A\overline{A}^\top$ and of $\overline{A}^\top A$, respectively, are given by $\sigma_1^2, \dots, \sigma_r^2$. Furthermore, the eigenvectors of $A\overline{A}^\top$ are exactly the left singular vectors of A and the eigenvectors of $\overline{A}^\top A$ are exactly the right singular vectors of A. ◇

Some particular consequences of that connection which will be used later on are listed in the following lemma.

(2.7) Lemma
Consider a matrix $A \in \mathbb{C}^{n \times m}$ for $n \leq m$.

Suppose the singular values of A are given by $\sigma_1, \dots, \sigma_n$ and a singular value decomposition is given by $A = UDV$ for $U \in U(n)$, $V \in U(m)$ and a real diagonal matrix $D \in \mathbb{R}^{n \times m}$. The following is true:

1) The eigenvalues of $A\overline{A}^\top$ are given by $\lambda_i = \sigma_i^2$ for $i = 1, \dots, n$.
2) The matrix $A\overline{A}^\top$ is diagonalized by U, that is $A\overline{A}^\top = U(DD^\top)\overline{U}^\top$.

Conversely, suppose the eigenvalues of $A\overline{A}^\top$ are given by $\lambda_1, \dots, \lambda_n \in \mathbb{R}_{\geq 0}$ and $U \in U(n)$ is such that $A\overline{A}^\top = UD\overline{U}^\top$ for $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n \times n}$. The following is true:

1) The singular values of A are given by $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \dots, n$.
2) A singular value decomposition of A is given by $A = UD'V$ for some $V \in U(m)$ and a real diagonal matrix $D' \in \mathbb{R}^{n \times m}$ satisfying $D'D'^\top = D$. ◇

By the definition of the Frobenius norm of a matrix it is clear that it can be easily computed in terms of its singular values.

(2.8) Lemma
The Frobenius norm of $A \in \mathbb{C}^{n \times m}$, $n, m \in \mathbb{N}$ with singular values $\sigma_1, \dots, \sigma_l$ for $l = \min(n, m)$ is given by

$$\|A\|_F = \sqrt{\sum_{i=1}^{l} \sigma_i^2}.$$

Now, an important class of matrices which have a particularly simple singular value decomposition is introduced.

(2.9) Definition (normal matrices)
A matrix A is said to be *normal* if it commutes with its conjugate transpose, that is $A\overline{A}^\top = \overline{A}^\top A$. ◇

Normal matrices may be classified easily by means of their eigenvalue decomposition.

(2.10) Lemma
Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix. Then A may be diagonalized by a unitary matrix, that is: there exists $U \in U(n)$ such that $UA\overline{U}^\top = D$, where $D \in \mathbb{C}^{n\times n}$ is a diagonal matrix.

The singular values of A are given by the absolute values of its eigenvalues. ◇

Important classes of normal matrices are listed in the following examples.

(2.11) Example (normal matrices)

1) Let $A \in \mathcal{H}(n)$ be a Hermitian matrix.
   Then A is normal: $A\overline{A}^\top = A^2 = \overline{A}^\top A$.

   Hermitian matrices are exactly the normal matrices with only real eigenvalues.

2) Let $A \in U(n)$ be a unitary matrix.
   Then A is normal: $A\overline{A}^\top = I_n = \overline{A}^\top A$.

   Unitary matrices are exactly the normal matrices with only roots of unity as eigenvalues.

3) Let $A \in \mathbb{C}^{n\times n}$ be a skew-Hermitian matrix, that is $A = -\overline{A}^\top$.
   Then A is normal: $A\overline{A}^\top = -A^2 = \overline{A}^\top A$.

   Skew Hermitian matrices are exactly the normal matrices with only purely imaginary eigenvalues. ◇

By lemma 2.8 the Frobenius norm of a matrix can be expressed by means of its singular values. Since for normal matrices these are given by the absolute values of the eigenvalues, the following result is obtained.

(2.12) Corollary
The Frobenius norm of a normal matrix $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1, \dots, \lambda_n$ is given by

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} |\lambda_i|^2}.$$

◇

The trace of the product of two positive semidefinite Hermitian matrices will be an important value later on. It can be seen to be a nonnegative real number.

(2.13) Lemma
Let $A, B \in \mathbb{C}^{n\times n}$, $n \in \mathbb{N}$ be two positive semidefinite Hermitian matrices.

Then $\operatorname{Tr}(AB) \in \mathbb{R}_{\geq 0}$ holds. ◇

Proof. Any positive semidefinite Hermitian matrix $X \in \mathbb{C}^{n\times n}$ can be decomposed as $X = UD\overline{U}^\top$ for $U \in U(n)$ and $D = \operatorname{diag}(d_1, \ldots, d_n)$ for $d_1, \ldots, d_n \in \mathbb{R}_{\geq 0}$. Therefore, X may be written as $X = \hat{X}^2$ where $\hat{X} = U\hat{D}\overline{U}^\top$ for $\hat{D} = \operatorname{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n})$. In particular, $\hat{X}$ is also positive semidefinite Hermitian.

Consequently, there are positive semidefinite Hermitian matrices $\hat{A}, \hat{B} \in \mathbb{C}^{n\times n}$ such that $A = \hat{A}^2$ and $B = \hat{B}^2$ hold. Hence, one obtains

$$\operatorname{Tr}(AB) = \operatorname{Tr}(\hat{A}^2\hat{B}^2) = \operatorname{Tr}(\hat{A}\hat{B}\hat{B}\hat{A}) = \operatorname{Tr}(\overline{\hat{B}\hat{A}}^\top \hat{B}\hat{A}) = \left\|\hat{B}\hat{A}\right\|_F^2 \in \mathbb{R}_{\geq 0}.$$

□

2.2 Lie groups and Lie algebras

Some problems that are considered later require the minimization of $\|XY\|_F$ or the maximization of $\left\|X^{-1}Y\right\|_F$ for matrices X, Y. The Lie algebra $\mathfrak{g}$ of a matrix Lie group G also consists of matrices and the two structures are connected by a map $\exp : \mathfrak{g} \to G$. By means of this map, the multiplicative structure of G is connected to the additive structure of $\mathfrak{g}$. Part of the motivation for this work stems from the idea to make use of the geometric structure of a Lie algebra and, by means of the map exp, translate it to a multiplicative structure which helps to solve the problems described above. In the following, Lie groups, Lie algebras and the map exp are introduced. This section mainly follows [Kir08].

(2.14) Definition (complex Lie group)
A *complex Lie group* G is a group which is also a complex analytic manifold over $\mathbb{C}$. The group structure agrees with the manifold structure in the sense that the multiplication map $G \times G \to G$, $(g, h) \mapsto gh$ and the inversion map $G \to G$, $g \mapsto g^{-1}$ are analytic maps.

A *morphism* of complex Lie groups G_1, G_2 is an analytic map $f : G_1 \to G_2$ which is also a group homomorphism. ◇

(2.15) Example
A simple example of a complex Lie group is the additive group of $\mathbb{C}$ itself.

Moreover, many subgroups of $\mathrm{GL}_n(\mathbb{C})$, such as $\mathrm{SL}_n(\mathbb{C})$ or $\mathrm{GL}_n(\mathbb{C})$ itself can be regarded as complex Lie groups. ◇

Lie algebras were originally obtained from Lie groups and due to that, their structure was completely determined by the corresponding Lie group. It is, however, possible to define Lie algebras independently.

(2.16) Definition (Lie algebra)
A *Lie algebra* L over a field K is a vector space together with a K-bilinear operation

$$L \times L \to L,\ (x,y) \mapsto [x,y]$$

satisfying the following identities:

(1) $[x,y] = -[y,x]$ for all $x, y \in L$,

(2) $[x,[y,z]] + [z,[x,y]] + [y,[z,x]] = 0$ for all $x,y,z \in L$.

The expression $[x,y]$ for $x, y \in L$ is commonly referred to as the *Lie bracket* of x, y. ◇

(2.17) Example
For any field K and any $n \in \mathbb{N}$, the ring of matrices $K^{n\times n}$ can be given the structure of a Lie algebra by defining the Lie bracket to be the ring theoretic commutator $[X,Y] := XY - YX$ for $X, Y \in K^{n\times n}$. ◇

In the following, the connection between Lie groups and Lie algebras will be developed. To this end, a certain tangent space of a Lie group is essential.

(2.18) Lemma
Let G be a complex Lie group and $\mathfrak{g} := T_1 G$ the corresponding tangent space at the identity $1 \in G$.

For any $x \in \mathfrak{g}$, there exists a unique morphism of Lie groups $\gamma_x : \mathbb{C} \to G$ such that $\gamma_x'(0) = x$ holds, where γ_x' denotes the derivative of γ_x. The map γ_x is called the *one-parameter subgroup* with respect to x. ◇

This lemma allows to give the following definition.

(2.19) Definition (exponential map)
Let G be a complex Lie group and $\mathfrak{g} = T_1 G$ the corresponding tangent space at the identity $1 \in G$.

The *exponential map* $\exp : \mathfrak{g} \to G$ is defined by

$$\exp(x) = \gamma_x(1).$$

The exponential map is analytic and locally invertible. ◇

The local invertibility of the exponential map allows to describe the multiplication of two elements of the Lie group in terms of their two corresponding elements of the tangent space.

(2.20) Corollary
There is an analytic map $\mu : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, defined in a neighborhood of $(0, 0)$, which satisfies $\exp(\mu(x, y)) = \exp(x) \exp(y)$.

Analytic maps may be presented by their Taylor expansion. For μ, one obtains

$$\mu(x, y) = x + y + \lambda(x, y) + \cdots ,$$

where the dots represent terms of order larger than two and $\lambda : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is a bilinear map satisfying $\lambda(x, y) = -\lambda(y, x)$. ◇

By means of the map λ, the space $T_1 G$ may now be endowed with the structure of a Lie algebra.

(2.21) Lemma
Consider a complex Lie group G. The tangent space $\mathfrak{g} := T_1 G$ at the identity $1 \in G$ is a Lie Algebra with Lie bracket defined as $[x, y] := \lambda(x, y)$ for $x, y \in \mathfrak{g}$. ◇

For the applications that will be studied later matrices over the complex numbers will be of central interest. Therefore, in particular matrix groups will be considered.

(2.22) Definition
Closed subgroups of $\mathrm{GL}_n(\mathbb{C})$ are complex Lie groups in their own right and are called *linear Lie groups*. ◇

(2.23) Example
Examples of linear Lie groups include $\mathrm{SL}_n(\mathbb{C})$ or again, $\mathrm{GL}_n(\mathbb{C})$ itself. In particular, the Lie algebra of $\mathrm{GL}_n(\mathbb{C})$ is given by $\mathbb{C}^{n \times n}$, as introduced in example 2.17. ◇

2.3 The exponential map for linear Lie algebras

As discussed above, the exponential map will be employed to parametrize matrices in later chapters. Therefore, some important properties of the exponential map for linear Lie algebras are introduced. This section also follows [Kir08].

(2.24) Proposition
For linear Lie-groups G, the exponential map is given by the well known series

$$\exp : T_1 G \to G,\ X \mapsto \sum_{i=0}^{\infty} \frac{1}{i!} X^i.$$

◇

As a consequence of the above proposition, the exponential map for linear Lie groups may be regarded as a generalization of the well known real or complex exponential map. In particular, it acts on the diagonal entries of diagonal matrices by $\exp : \mathbb{C} \to \mathbb{C}$.

(2.25) Corollary
Consider a complex diagonal matrix $D \in \mathbb{C}^{n \times n}$ with diagonal entries $d_1, \ldots, d_n \in \mathbb{C}$. The value $\exp(D)$ is given by

$$\exp \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = \begin{pmatrix} \exp(d_1) & & \\ & \ddots & \\ & & \exp(d_n) \end{pmatrix}.$$

◇

Some further important properties are listed in the following lemma.

(2.26) Lemma (Properties of the exponential map)
Consider a linear Lie group G with corresponding Lie algebra $\mathfrak{g} = T_1 G$. The following are true:

(i) $xy = yx \implies \exp(x)\exp(y) = \exp(x+y)$ for $x, y \in \mathfrak{g}$.

(ii) $\exp((t+s)x) = \exp(tx)\exp(sx)$ for $s, t \in K$ and $x \in \mathfrak{g}$.

(iii) $X \exp(y) X^{-1} = \exp(XyX^{-1})$ for $X \in G$ and $y \in \mathfrak{g}$. ◇

Part (iii) of this lemma, combined with the observation how the exponential map acts on diagonal matrices in the corollary before, yields the following result concerning eigenvalues.

(2.27) Corollary
Consider a matrix $Y \in \mathbb{C}^{n \times n}$, $n \in \mathbb{N}$.

1) Suppose Y is diagonalizable with eigenvalues $\lambda_1, \dots, \lambda_n$. Then $\exp(Y)$ is diagonalizable with eigenvalues $\exp(\lambda_1), \dots, \exp(\lambda_n)$.

2) The matrix Y is Hermitian $\iff$ $\exp(Y)$ is positive definite Hermitian. $\diamond$

For the special case of normal matrices further consequences are immediate.

(2.28) Corollary
Let $A \in \mathbb{C}^{n \times n}$, $n \in \mathbb{N}$ be a normal matrix with $A = \overline{U}^\top DU$ for $U \in U(n)$ and a diagonal matrix $D \in \mathbb{C}^{n \times n}$.

Then $\exp(A)$ is normal and satisfies $\exp(A) = \overline{U}^\top \exp(D)U$. $\diamond$

2.4 Matrix series

In a later chapter, some explicit calculations involving the exponential map for matrices have to be performed. In order to do so, some basic properties of series of matrices have to be established. The space $\mathbb{C}^{n \times n}$ is a Banach space with respect to the Frobenius norm $\|\cdot\|_F$. Therefore, series of matrices are a special case of series in Banach spaces. An introduction to Banach spaces and the following basic properties of corresponding series can be found in [Meg98].

(2.29) Proposition
Consider two Banach spaces V_1, V_2 over a field K and two families of elements $(x_i)_{i \in \mathbb{N}}, (y_i)_{i \in \mathbb{N}}$ for $x_i, y_i \in V_1$ for all $i \in \mathbb{N}$.

(i) If $\sum_{i=1}^{\infty} x_i$ and $\sum_{i=1}^{\infty} y_i$ are convergent, then so is $\sum_{i=1}^{\infty}(x_i + y_i)$ and $\sum_{i=1}^{\infty} x_i + \sum_{i=1}^{\infty} y_i = \sum_{i=1}^{\infty}(x_i + y_i)$ holds.

(ii) If $\sum_{i=1}^{\infty} x_i$ is convergent and $\alpha \in K$ is a scalar, then $\sum_{i=1}^{\infty} \alpha x_i$ is convergent and satisfies $\sum_{i=1}^{\infty} \alpha x_i = \alpha \sum_{i=1}^{\infty} x_i$.

(iii) If $\sum_{i=1}^{\infty} x_i$ is convergent and $T : V_1 \to V_2$ as a continuous linear operator, then $\sum_{i=1}^{\infty} T(x_i)$ is convergent in V_2 and $\sum_{i=1}^{\infty} T(x_i) = T\left(\sum_{i=1}^{\infty} x_i\right)$ holds.

(iv) Denote the norm of V_1 by $\|\cdot\|$. Then $\|\sum_{i=1}^{\infty} x_i\| \leq \sum_{i=1}^{\infty} \|x_i\|$ holds.

(v) If $\sum_{i=1}^{\infty} \|x_i\|$ is convergent, the series $\sum_{i=1}^{\infty} x_i$ is said to be *absolutely convergent*. In particular, this implies that $\sum_{i=1}^{\infty} x_i$ is convergent and so is $\sum_{i=1}^{\infty} x_{\pi(i)}$ for any permutation π of $\mathbb{N}$. $\diamond$

Multiplication by a matrix $C \in \mathbb{C}^{n\times n}$ yields a linear continuous operator. Thus, the following corollary is obtained.

(2.30) Corollary
Consider families of matrices $(A^{(i)})_{i\in\mathbb{N}}, (B^{(i)})_{i\in\mathbb{N}}$ for $A^{(i)}, B^{(i)} \in \mathbb{C}^{n\times n}$ for all $i \in \mathbb{N}$ and some $C \in \mathbb{C}^{n\times n}$.

If $\sum_{i=1}^{\infty} A^{(i)}$ is convergent and $C \in \mathbb{C}^{n\times n}$, then $\sum_{i=1}^{\infty} CA^{(i)}$ is convergent and satisfies

$$\sum_{i=1}^{\infty} CA^{(i)} = C\sum_{i=1}^{\infty} A^{(i)}.$$

Consequently, if $\sum_{i=1}^{\infty} A^{(i)}$ and $\sum_{j=1}^{\infty} B^{(j)}$ are convergent, one obtains

$$\sum_{i=1}^{\infty} A^{(i)} \sum_{j=1}^{\infty} B^{(j)} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A^{(i)}B^{(j)}.$$

◇

When introducing the series expansion of the exponential map for matrices, it was implicitly claimed that it is convergent and hence well defined. This can now be shown formally. It is easiest to show that it even converges absolutely.

(2.31) Lemma
The series $\sum_{i=0}^{\infty} \frac{A^i}{i!}$ converges absolutely for any $A \in \mathbb{C}^{n\times n}$. ◇

Proof. The submultiplicativity of the Frobenius norm yields $\left\|A^i\right\|_F \leq \|A\|_F^i$ for all $i \in \mathbb{N}$. Hence, the assertion of the lemma can be obtained as follows:

$$\sum_{i=0}^{\infty} \left\| \frac{A^i}{i!} \right\|_F \leq \sum_{i=0}^{\infty} \frac{\|A\|_F^i}{i!} = \exp(\|A\|_F).$$

□

The trace of a matrix series will be of importance later. It can be shown to be the series of the traces of the summands.

(2.32) Corollary
Suppose $\sum_{i=0}^{\infty} A^{(i)}$ converges, then $\mathrm{Tr}\left(\sum_{i=0}^{\infty} A^{(i)}\right) = \sum_{i=0}^{\infty} \mathrm{Tr}(A^{(i)})$ holds and the series on the right hand side is also convergent. ◇

Proof. Clearly, if $\sum_{i=0}^{\infty} A^{(i)}$ converges, for any $k, l \in \{1, \ldots, n\}$ the series $\sum_{i=0}^{\infty} A_{kl}^{(i)}$ also converges. Additionally making use of proposition 2.29 part (i) yields

$$\operatorname{Tr}\left(\sum_{i=0}^{\infty} A^{(i)}\right) = \sum_{j=1}^{n} \left(\sum_{i=0}^{\infty} A^{(i)}\right)_{jj} = \sum_{j=1}^{n} \sum_{i=0}^{\infty} A_{jj}^{(i)} = \sum_{i=0}^{\infty} \sum_{j=1}^{n} A_{jj}^{(i)} = \sum_{i=0}^{\infty} \operatorname{Tr}(A^{(i)}).$$

This is the assertion of the corollary. □

§ 3 Introduction to space-time block coding

This chapter begins by explaining how data are transferred over a wireless channel in general. After the basic setting has been described, space-time block codes are introduced and the challenges in designing such codes are discussed.

3.1 Modeling wireless communication

In general, information can be transmitted over a wireless channel by modulating an electromagnetic wave transmitted by one antenna and then demodulating it accordingly at a receive antenna. An electromagnetic wave has three properties which can be modulated and hence used to encode information. These properties are its frequency, its amplitude and its phase. In order for the receiver to be able to identify a signal sent by the corresponding transmitter, at least one of these properties has to be fixed. A commonly used setup is fixing a frequency and then encoding information using a finite amount of different states of the phase and the amplitude. Denoting the phase of the signal by φ and the amplitude by R, a piece of information transmitted over the channel can then be regarded as a complex number $Re^{i\varphi}$.

An important difference between wireless communication and the wire based case is that the signal will spread out from the transmitting antenna in any direction. It may then be blocked by objects in the way or be reflected towards different directions multiple times. This causes multiple possibly degraded copies of the original signal to reach the receive antenna via several different paths. The resulting effect is called *fading* of the signal and is commonly modeled by multiplying the sent signal by a complex number called the *fading-coefficient*. This coefficient depends on the actual physical environment between the transmitter and the receiver at the time the signal is sent.

To get a complete model of the wireless communication channel, also the presence of background-noise at the receiving antenna needs to be included. This part is simply modeled by adding another complex number to the faded signal.

Altogether the following setting is obtained:

$$x\bullet \overset{h}{\rightsquigarrow} \bullet y = hx + v$$

A signal $x \in \mathbb{C}$ is sent from the antenna on the left, it undergoes the fading process modeled by the fading-coefficient $h \in \mathbb{C}$ along the way and at the receive antenna on the right side a message $y = hx + v$ is received, thereby $v \in \mathbb{C}$ represents the additive background-noise.

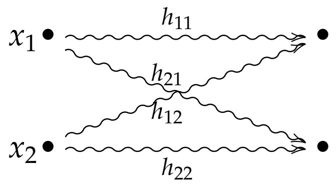
It is worth noting at this point that the transmission could always be made more reliable by transmitting a scaled signal ρx for a large real factor $\rho > 1$ as this would reduce the relative influence of the additive noise on the received signal. However, scaling up the signal which is sent across the channel corresponds to using more energy for that transmission in practice. Therefore the absolute values of signals that are sent across the channel are bound by the amount of energy available at the transmitter. Consequently, the ratio $\frac{|hx|^2}{|v|^2}$ is an important factor in wireless communication. It is referred to as the *signal-to-noise ratio*, commonly abbreviated as SNR.

In many applications the signal strength will be significantly higher than the strength of the background noise and therefore the SNR is commonly measured by the logarithmic decibel (dB) scale. The SNR ρ in dB is given by

$$\rho = 10 \cdot \log_{10}\left(\frac{|hx|^2}{|v|^2}\right) dB.$$

3.2 Space-time block codes

In order to improve the reliability of the communication described above systems using multiple transmit and receive antennas have been introduced. For instance when using two antennas at both ends the setting is as follows:



$$y_1 = h_{11}x_1 + h_{12}x_2 + v_1$$

$$y_2 = h_{21}x_1 + h_{22}x_2 + v_2$$

There are two signals $x_1, x_2 \in \mathbb{C}$ now which are sent from the two respective transmit antennas at the same time. These two signals are reaching the two receive antennas via different paths. Consequently there are now four different fading coefficients h_{ij} for $i, j \in \{1, 2\}$ in total. The coefficient h_{ij} models the fading process which the signal x_i, that was sent at the i-th transmit antenna, undergoes while traveling to the j-th receive antenna. Also there are two values $v_1, v_2 \in \mathbb{C}$ representing the respective background noise at each receive antenna. The signal y_j which is received at the j-th receive antenna is then a superimposition of the two faded signals that have been transmitted plus the additive noise present at this antenna: $y_j = h_{1j}x_1 + h_{2j}x_2 + v_j$.

This setting can be further generalized to employ M transmit and N receive antennas for any $M, N \in \mathbb{N}$. Signals $x_1, \ldots, x_M \in \mathbb{C}$ are transmitted simultaneously at the transmit antennas and travel along $N \cdot M$ different paths toward the receive antennas. Let $h_{ij} \in \mathbb{C}$ again denote the coefficient which models the fading that the signal x_i, which was sent at the i-th transmit antenna, undergoes while traveling to the j-th receive antenna. The signal y_j received at the j-th receive antenna is then given by $y_j = \sum_{i=1}^{N} h_{ij}x_i + v_j$, where v_j again denotes the additive noise present at this antenna.

The vector of the received received signals can hence be expressed by the vector equation

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} h_{11} & \cdots & h_{1M} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{NM} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} + \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}.$$

The matrix $H := (h_{ij})_{1 \le i \le N, 1 \le j \le M} \in \mathbb{C}^{N \times M}$ is called the *fading-matrix* or the *channel-matrix*.

So far, the spatial dimension has been exploited by spreading out multiple antennas to different spatial locations and by combining the information symbols sent and received in vectors.

Moreover, it is reasonable to assume that the physical environment of the communication channel does not change for short periods of time. Therefore the fading-matrix may be assumed to be constant for several consecutive transmissions. The corresponding time period is called a *coherence interval*. In the following it will be assumed that T information symbols can be transmitted per antenna during this interval. Let the symbol that is transmitted from the i-th transmit antenna during the j-th time step be denoted by $x_{ij} \in \mathbb{C}$. This yields T vectors

$$\begin{pmatrix} x_{11} \\ \vdots \\ x_{M1} \end{pmatrix}, \ldots, \begin{pmatrix} x_{1T} \\ \vdots \\ x_{MT} \end{pmatrix} \in \mathbb{C}^M$$

which can be transmitted while the fading matrix H remains constant. These vectors can be combined into a matrix $X \in \mathbb{C}^{M \times T}$ such that the ij-th entry of X contains the information symbol sent from the i-th transmit antenna during the j-th time step. Furthermore, set v_{ij} to be the background-noise at the i-th receive antenna during the j-th time step for $1 \leq i \leq N, 1 \leq j \leq T$ and y_{ij} to be the received signal at the i-th receive antenna during the j-th time step for $1 \leq i \leq N, 1 \leq j \leq T$. The channel equation becomes

$$\begin{pmatrix} y_{11} & \cdots & y_{1T} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{NT} \end{pmatrix} = \begin{pmatrix} h_{11} & \cdots & h_{1M} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{NM} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1T} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MT} \end{pmatrix} + \begin{pmatrix} v_{11} & \cdots & v_{1T} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NT} \end{pmatrix}.$$

Information that shall be sent across the wireless channel can now be encoded into the matrix X. A finite collection $\mathcal{C}$ of such codewords is referred to as a *space-time block code* (STBC). Such a set $\mathcal{C}$ is commonly also referred to as a *code* or a *codebook*.

An important parameter of an STBC is the amount of codewords it contains, as that controls the amount of information that is essentially conveyed by one codeword. In information theory, the quantity of information is commonly measured in *bit*. If a codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ contains $|\mathcal{C}| = n$ codewords, it holds $\log_2(n)$ bit of information. Therefore, $\log_2(|\mathcal{C}|)$ bit of information can be transmitted across the channel over one coherence interval of T time steps. This motivates the definition of the *rate* of an STBC. The rate of a STBC $\mathcal{C}$ is defined as

$$\operatorname{rate}(\mathcal{C}) := \frac{\log_2(|\mathcal{C}|)}{T}.$$

It corresponds to the amount of information, measured in bits, which can be transmitted across the channel during one time step using the codebook $\mathcal{C}$.

In the situation introduced in this section, a message is received at several antennas over several time steps. Therefore it is necessary to extend the notion of the signal-to-noise ratio. In case a codeword $X \in \mathbb{C}^{M \times T}$ has been sent over a channel with the corresponding channel matrix $H \in \mathbb{C}^{N \times M}$ and the background noise at the receive antennas represented by the matrix $V \in \mathbb{C}^{N \times T}$, the SNR ρ at the receiver is defined as

$$\rho = \rho(X, H, V) = \frac{\|HX\|_F^2}{\|V\|_F^2}.$$

3.3 The stochastic model

Now that the notion of a space-time-block code is established, the focus is placed on the problem of finding codes which will perform well in practice.

There are two main challenges which have to be met in order to create well-performing STBC. At first there needs to be a preferably efficient algorithm that decodes a received signal correctly to the codeword that was transmitted with high probability. In addition it has to be investigated how a codebook should be designed in order for the decoding algorithm to make as few mistakes as possible.

These problems depend on the channel matrices H and the noise matrices V that occur in the channel equation. As it cannot be assumed that these matrices are known when designing a codebook, they need to be modeled as random variables.

The coefficients v_{ij} representing the background noise at the i-th receive antenna during the j-th time step can be assumed to be statistically independent and are commonly assumed to be complex Gaussian distributed. This will also be assumed throughout this work.

The probability distributions of the fading coefficients does depend more strongly on the specific setting in which the codes are to be used. Therefore different models do exist, providing different random distributions for the entries of the channel matrix. However, in most cases a model known as *Rayleigh*-fading can be applied and it has therefore received by far the most attention. In this case, the fading-coefficients are all supposed to be independent and complex Gaussian distributed with variance 1. In the following only this case will be considered.

When the channel and noise coefficients are considered as random variables, it is possible to introduce the notion of signal-to-noise ratio for the channel as opposed to defining it for one particular transmission. The signal-to-noise ratio ρ of a channel at the receiver is defined as

$$\rho = \frac{E[\|HX\|_F^2]}{E[\|V\|_F^2]},$$

where E denotes the expected value of the respective expressions.

3.4 Types of STBC

When it comes to the decoding of STBC, a distinction has to be made whether or not the channel matrix H is known to the receiver. In practice, this knowledge can be obtained by sending some predefined training signals over the channel, allowing the receiver to estimate the channel matrix. If the channel matrix is known at the receiving end, the transmission is referred to as a *coherent* transmission. Otherwise it is called a *noncoherent* transmission. Extending this notion, codes specifically designed for one of these cases are also referred to as respectively coherent or noncoherent codes.

Employing the process of learning the channel coefficients does only make sense if the channel matrix stays invariant for a sufficiently long period of time. If the physical channel between the sender and the receiver changes too quickly, it may not be worthwhile to learn the channel coefficients at the receiver. The most obvious case when this does happen is when at least one of the communication partners is inside of a vehicle moving above a certain speed. The particular speed at which it is no longer beneficial to estimate the channel coefficients depends on several properties of the channel. Besides the frequency used for transmitting the signal and the statistical distribution of the occurring coefficients, the number of antennas plays a limiting role. The more antennas used to transmit and receive the signals, the more channel coefficients have to be estimated and therefore a larger amount of time steps has to be allocated as a training period.

In [HM00] the authors briefly discuss an example which gives an idea in which cases it is not feasible to use coherent coding any more.

(3.1) Example (Wireless communication involving a moving party)
For a transmitter aboard a vehicle moving at 60 miles per hour (about 100 kilometers per hour) transmitting at a frequency of 1.9 Gigahertz the channel stays invariant for about 3 milliseconds. In this setting a reasonable assumption is that about 30000 symbols can be transmitted per second. This allows for about $50 - 100$ symbols to be transmitted during the period over which the channel remains invariant.

This severely limits the amount of antennas for which it makes sense to use training based coherent coding.

The authors further argue that at a speed of 500 kilometers per hour it would not even be feasible to learn the single coefficient of a channel using only one transmit and one receive antenna. $\diamond$

The focus of this work is on studying the noncoherent case. However, an overview about the most important topics concerning space-time block codes in general will be presented in the following. In particular, besides the noncoherent case, a short overview of the coherent case as well as a special case of noncoherent coding, known as differential coding, will be given. The challenges which arise with each of these cases will be discussed briefly.

3.4.1 The coherent case

For many practical applications it is reasonable to assume that the channel matrix is known at the receiving end. The problem of designing codes for this case has therefore received the greatest attention and very many constructions of corresponding codes have been proposed.

The first question that has to be answered when considering a class of codebooks is how decoding can be done. To that end, assume the codewords of a space-time block code $C \subseteq \mathbb{C}^{M \times T}$ are transmitted over the channel using coherent transmission. For a received matrix Y, the decoding problem is to find the codeword $\tilde{X}$ which satisfies the equation

$$\left\|H\tilde{X} - Y\right\|_F = \min_{X \in C} \|HX - Y\|_F .$$

Since in the coherent case the fading-matrix H is known to the decoder, the set of all faded codewords $\{HX \mid X \in C\}$ can be computed and the minimal solution to $\|HX - Y\|_F$ can be found. In general, this leads to an exhaustive search in a set with $|C|$ elements. However, if the codewords are additionally assumed to be linear in the information symbols, they may be embedded into a lattice Λ. The decoding problem then translates into a closest lattice point problem in the distorted lattice $H\Lambda$ which can be solved efficiently, for example using the *sphere-decoder*. (See [VB99, HV05].)

The problem of designing good codes, however, turns out to be harder. By examining the pairwise error probability it was found that in order to design good codes, the absolute value of the determinant of the difference of two codewords needs to be maximized. This, in particular, forces the difference of any two codewords to be invertible, which may be achieved by constructing codebooks based on division algebras. Many good codes have been found using this approach. Several exemplary constructions can be found in [BO13] along with an extensive introduction to central simple algebras and an explanation how they can be used to find suitable division algebras.

3.4.2 Differential coding

If the channel matrix is not known to the receiver, the decoding and the design problem differ a lot from the ones described in the previous section. However, there is a technique called differential coding that can be used for noncoherent communication and for which the design criteria are similar to the ones obtained for the coherent case. As this also allows to use interesting algebraic constructions to obtain good codebooks, this special case shall also be introduced here.

In order to use differential coding, the codebook is required to consist only of square unitary matrices. In addition, it is necessary that the channel stays invariant for $(T+1) \cdot M$ time steps. This allows $T+1$ matrices of dimension $M \times M$ to be transmitted during that time frame. Thus, codebooks are finite subsets $\mathcal{C} \subseteq U(M)$. The transmitted $T+1$ matrices allow to communicate T codewords $S_1, \dots, S_T \in \mathcal{C}$ to the receiver. In particular, the matrices

$$X_0 := S_0 := I_M, \quad X_1 := S_1, \quad X_2 := S_1 S_2, \quad \dots, \quad X_T := S_1 \cdots S_T$$

are consecutively transmitted over the channel. Therefore, for $j > 0$ after the j-th transmission, the matrix

$$Y_j = HX_j + V_j = HS_0 \cdots S_j + V_j = Y_{j-1}S_j - V_{j-1}S_j + V_j$$

is received. In this formula, the right hand side for none of the $j = 1, \dots, T$ depends on the channel matrix H. Since the signal Y_{j-1} has been received in the time step before, it is known to the decoder, and therefore the message received in the j-t time step may be decoded to the codeword $\tilde{S}_j$ satisfying

$$\left\| Y_j - Y_{j-1}\tilde{S}_j \right\|_F = \min_{S \in \mathcal{C}} \left\| Y_j - Y_{j-1}S \right\|_F .$$

It has been found that, as in the coherent case, the most important design criterion is to maximize $|\det(S_1 - S_2)|$ for $S_1 \neq S_2$, $S_1, S_2 \in \mathcal{C}$. Therefore, the design problem becomes the same as in the coherent case with the additional restriction that all codewords have to be unitary matrices. A way to construct suitable codebooks from division algebras is also described in [BO13].

To conclude, it shall be noted that while technically differential coding does not require the knowledge of the channel matrix at the receiver, it may also be regarded as a special case of coherent coding. To see that, recall that coherent coding requires a training period at the beginning of every coherence interval before actual information can be transmitted. During this period an estimation of the channel coefficients is obtained. In the case of differential coding, transmitting the initial signal $X_0 = I_M$ may be regarded as such a training period. For all later transmissions, the knowledge of the channel coefficients is implicitly handed over from the preceding transmission. Also, differential coding requires the channel to remain invariant for at least $2M$ time steps in order to transmit one codeword. In that case, it would also be possible to reserve the first few time steps to estimate the channel coefficients and to use coherent coding to transmit information afterwards.

3.4.3 The noncoherent case

The most general case is the one where the channel information is not known at the receiver and no additional restrictions are imposed upon the codewords. As less information is available to the receiver, naturally the rate which can be achieved by noncoherent coding is less than what may be achieved if the channel coefficients are known. Noncoherent coding is therefore only used in cases in which the channel does in fact change too quickly to make an estimation of the channel coefficients feasible. For these reasons noncoherent codes have received less attention than their coherent counterparts so far.

So far the best known STBCs for the noncoherent channel only make use of $\lfloor \frac{T}{2} \rfloor$ transmit antennas. These codes are constructed by means of a coding scheme known as *unitary coding*.

For such codes the maximum likelihood decoding problem of finding the codeword X^* for a received message Y becomes

$$X^* = \text{argmax}_{X \in \mathcal{C}} \left(\left\| X^\top Y \right\|_F \right).$$

The corresponding decoder is referred to as the GLRT-decoder, and for the decoding problem efficient means of decoding are available, see for instance [RCC07].

Concerning the problem of designing good unitary codebooks, several criteria have been proposed. These criteria are usually based on the matrix products $X_1 \overline{X_2}^\top$ for codewords $X_1, X_2 \in \mathcal{C} \subseteq \mathbb{C}^{M \times T}$. For instance, the *GLRT-criterion* requires to minimize $\left\| X_1 \overline{X_2}^\top \right\|_F$ for distinct codewords X_1, X_2. Good unitary codes have been constructed with respect to this criterion. It will also be discussed later on and it will be investigated to what extent it can be generalized to noncoherent STBC if one does not require the codewords to be unitary.

§ 4 Known results and constructions of noncoherent STBC

In this chapter the known results on noncoherent STBC will be summarized and the best known constructions will be presented.

The systematic work on noncoherent STBC started in 1999 with an analysis of the capacity of the corresponding channel by B. Hochwald and T. Marzetta ([MH99]). The capacity of a transmission channel is an information theoretic measure of the maximal amount of information which can be transmitted over this channel in a given amount of time without any data loss. It therefore provides an upper bound on the performance which may be achieved by any code used to transmit information over the channel.

The main results concern the structure which codewords need to have in order for a corresponding code to achieve the channel capacity. Furthermore the dependence of the capacity on channel parameters such as the number of transmit antennas M, the length of the time period over which the channel stays invariant T and the signal-to-noise ratio are investigated.

The paper was followed in 2000 by a second one ([HM00]) which introduced a coding scheme referred to as *unitary space-time modulation*. This scheme restricts codewords to matrices $X \in \mathbb{C}^{M \times T}$ for $T > M$ which satisfy $X\overline{X}^{\top} = \rho I_M$ for a positive real factor ρ depending on the signal-to-noise ratio. In doing so, the design- and decoding problem could be significantly simplified and some bounds on the capacity could be given for certain specific choices of M and T.

Unitary space-time modulation has also been referred to as *unitary coding*. Following its introduction, a connection to a packing problem on the Grassmann manifold has been investigated and the best known noncoherent STBC that have been constructed make use of this connection. The problem of finding optimal noncoherent STBC, in the sense that they yield a transmission error rate which is as low as possible, is, however, still very challenging. Since optimal codes are only known in some special cases, numerical algorithms to find near optimal codebooks are of interest. At the end of this chapter one algorithm is presented, outlining the complexity of the corresponding optimization problem.

4.1 General results on noncoherent STBC

This section introduces and elaborates on the known results on general noncoherent STBC with no further restrictions. It mainly follows the initial paper by Hochwald and Merzetta [MH99]. However, many details have been added to the stochastic computations in order to provide a good foundation for the generalizations which will be made in later chapters.

As introduced in section 3.3, the channels which are going to be considered throughout this work will be assumed to be subject to Rayleigh fading. To start with, the necessary notation is fixed. During a period over which the channel stays invariant a codeword $X \in \mathbb{C}^{M\times T}$ is transmitted. It undergoes some fading by the channel-matrix $H \in \mathbb{C}^{N\times M}$ and is further degraded by some additive noise $V \in \mathbb{C}^{N\times T}$ at the receiving antennas. The received codeword is then given by $Y = HX + V \in \mathbb{C}^{N\times T}$.

In the case of Rayleigh fading, the entries of the channel matrix are complex Gaussian distributed with expected value 0 and variance 1. The channel equation will be assumed to be normalized in the sense that the Gaussian distributed additive noise also has variance 1 at any antenna and in any time step. If the variance of the entries of V is given by σ^2, this can be achieved by accordingly scaling the channel equality to

$$\frac{1}{\sigma}Y = H\left(\frac{1}{\sigma}X\right) + \frac{1}{\sigma}V.$$

While investigating the performance of a code with respect to a certain channel a few basic stochastic principles will have to be used. These are briefly introduced in the following remark.

(4.1) Remark

(i) The expected value is a linear operator. That is, for two random variables X, Y and a deterministic value α the equalities $E[X+Y] = E[X] + E[Y]$ and $E[\alpha X] = \alpha E[X]$ are satisfied.

(ii) The *variance* of a complex random variable X is defined as $\mathrm{Var}(X) = E[\overline{X}X]$.

(iii) The expected value is not multiplicative. The *covariance* of two complex random variables X and Y, denoted by $\mathrm{Cov}(X,Y)$, may be defined as

$$\mathrm{Cov}(X,Y) = E[\overline{X}Y] - E[\overline{X}]E[Y].$$

In the special case where X and Y are independently distributed the covariance is zero and hence the expected value is multiplicative in X and Y.

(iv) If $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ are two multivariate random variables, their covariance matrix Λ is defined by $\Lambda_{ij} = \mathrm{Cov}(X_i, Y_j)$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. ◇

Using the respective assumptions that have been made on the random distributions of the channel matrix and the noise matrix, the signal-to noise ratio of the considered channel can be simplified. This is very helpful for investigating the performance of codebooks. In particular, two distinct codebooks may only be compared in a fair manner if they yield the same signal-to-noise ratio.

(4.2) Lemma
Assume that the entries of H and V are complex Gaussian distributed with variance 1 and expected value 0. Then, the signal-to-noise ration of the channel is given by

$$\frac{E[\|HX\|_F^2]}{E[\|V\|_F^2]} = \frac{E[\|X\|_F^2]}{T}.$$

Proof. The expected values in the numerator and the denominator will be evaluated separately.

To start with, the numerator can be written out as

$$E[\|HX\|_F^2] = \sum_{i=1}^{N} \sum_{j=1}^{T} E[(HX)_{ij}\overline{(HX)_{ij}}] = \sum_{i=1}^{N} \sum_{j=1}^{T} \sum_{k=1}^{M} \sum_{l=1}^{M} E[H_{ik}X_{kj}\overline{H_{il}X_{lj}}].$$

The entries of X and H are stochastically independent and hence the equality

$$E[H_{ik}X_{kj}\overline{H_{ik}X_{kj}}] = E[H_{ik}\overline{H_{il}}] \cdot E[X_{kj}\overline{X_{lj}}]$$

holds for all suitable i, j, k, l. Furthermore the entries of H have variance 1 and are pairwise stochastically independent. Therefore, they satisfy

$$E[H_{ik}\overline{H_{il}}] = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise.} \end{cases}$$

Altogether, one obtains

$$E[\|HX\|_F^2] = \sum_{i=1}^{N} \sum_{j=1}^{T} \sum_{k=1}^{M} E[X_{kj}\overline{X_{kj}}] = N \cdot E[\|X\|_F^2].$$

Using analogous arguments, the value of the denominator can be computed as

$$E[\|V\|_F^2] = \sum_{i=1}^{N} \sum_{j=1}^{T} E[V_{ij}\overline{V_{ij}}] = N \cdot T.$$

□

The lemma implies that the SNR of the channel depends only on the signal strength $\|X\|_F$. Therefore, two codebooks for which the codewords have the same Frobenius norms yield the same SNR. This is important, as it only makes sense to compare two codebooks which yield the same SNR.

As a next step the probability of transmission errors shall be determined. To that end, suppose $X \in \mathbb{C}^{M\times T}$ has been sent over the channel and consider the channel equation $Y = HX + V$. The entries of H and V are all independently distributed and the rows of Y only depend on disjoint subsets of these variables, hence they are independent as well. Denote the covariance matrix of Y_i by $\Lambda_X^{(i)}$. It will later be shown to be invertible for any $X \in \mathbb{C}^{M\times T}$ and $i \in \{1,\ldots,N\}$. The probability of a row $Y_i \in \mathbb{C}^{1\times T}$ being received at the i-th receive antenna if a codeword X has been sent can be computed by means of the probability density function of the multivariate complex normal distribution. It is given by

$$P(Y_i|X) = \frac{1}{\pi^T \det\left(\Lambda_X^{(i)}\right)} \exp\left(-Y_i(\Lambda_X^{(i)})^{-1}\overline{Y_i}^\top\right).$$

The authors of [MH99] quickly skip over the derivation of the error probability and do not explicitly use this density function. A detailed introduction of the multivariate normal distribution including its density function can be found in [Gut95].

The matrices $\Lambda_X^{(i)}$ can be calculated directly.

(4.3) Lemma
Suppose a codeword X has been sent across the channel. For any $i \in \{1,\ldots,N\}$ the covariance matrix of the i-th row of Y is given by:

$$\Lambda_X^{(i)} = I_T + \overline{X}^\top X.$$

This matrix does not depend on the choice of the row i. Therefore is is not necessary to index the matrix by i and in the following the notation

$$\Lambda_X := \Lambda_X^{(i)} = I_T + \overline{X}^\top X \quad \text{for} \quad i \in \{1,\ldots,N\}$$

will be used. ◇

Proof. The formula can be deduced by directly computing the covariance matrix of the i-th row of Y. By definition, its (j,k)-th entry is given by

$$\left(\Lambda_X^{(i)}\right)_{jk} = \operatorname{Cov}(Y_{ij}, Y_{ik}) = E[\overline{Y_{ij}}Y_{ik}] - E[\overline{Y_{ij}}]E[Y_{ik}] \quad \text{for } 1 \le j,k \le T.$$

Considering the second summand first, on the one hand

$$E[Y_{ij}] = E[\sum_{m=1}^{M} H_{im}X_{mj} + V_{ij}] = \sum_{m=1}^{M} E[H_{im}]X_{mj} + E[V_{ij}] = 0$$

holds and hence also $E[\overline{Y_{ij}}]E[Y_{ik}] = 0$.

On the other hand one obtains:

$$\begin{aligned}
E[\overline{Y_{ij}}Y_{ik}] &= E\left[\left(\overline{V_{ij}} + \sum_{s=1}^{M} \overline{H_{is}X_{sj}}\right)\left(V_{ik} + \sum_{t=1}^{M} H_{it}X_{tk}\right)\right] \\
&= E[\overline{V_{ij}}V_{ik}] + \sum_{s=1}^{M} E[\overline{H_{is}}]E[V_{ik}]\overline{X_{sj}} + \sum_{t=1}^{M} E[\overline{V_{ij}}]E[H_{it}]X_{tk} + \sum_{s=1}^{M}\sum_{t=1}^{M} E[\overline{H_{is}}H_{it}]\overline{X_{sj}}X_{tk} \\
&= E[\overline{V_{ij}}V_{ik}] + \sum_{s=1}^{M}\sum_{t=1}^{M} E[\overline{H_{is}}H_{it}]\overline{X_{sj}}X_{tk} \\
&= \begin{cases} \sum_{s=1}^{M} \overline{X_{sj}}X_{sk} & \text{, if } j \neq k \\ 1 + \sum_{s=1}^{M} \overline{X_{sj}}X_{sk} & \text{, if } j = k. \end{cases}
\end{aligned}$$

For the last equation, the equalities

$$E[\overline{V_{ij}}V_{ik}] = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad E[\overline{H_{is}}H_{it}] = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$$

were used, as in the proof of lemma 4.2. This yields the assertion of the lemma. □

By this explicit presentation, the matrices Λ_X can also be seen to be invertible.

(4.4) Lemma
The matrix $\Lambda_X = I_T + \overline{X}^\top X$ is positive definite Hermitian and hence invertible for any $X \in \mathbb{C}^{M\times T}$. ◇

Proof. Suppose a singular value decomposition of $X \in \mathbb{C}^{M\times T}$ is given by $X = UDV$ for $U \in U(M)$, $V \in U(T)$ and a real diagonal matrix $D \in \mathbb{R}^{M\times T}$ with nonnegative diagonal entries.

Then $\overline{X}^\top X = \overline{V}^\top(D^\top D)V$ is positive semidefinite Hermitian. Consequently, the matrix $\Lambda_X = I_T + \overline{X}^\top X$ is positive definite Hermitian and hence invertible. □

Since the rows of the matrix Y are stochastically independent, the conditional probability that Y is received equals the product of the probabilities for each of its rows to be received. Thus, one obtains the following result. Note that $\text{Tr}(AB) = \text{Tr}(BA)$ holds for any matrices A, B for which the products AB and BA are defined.

(4.5) Proposition
The probability that a message Y is received if a codeword X has been sent is given by

$$P(Y|X) = \frac{\exp(-\operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^\top Y))}{\pi^{TN}\det^N(\Lambda_X)}.$$

◇

The error-probability depends mainly on the covariance matrix Λ_X. For this reason, it plays a central role in the performance of noncoherent STBC.

Two more important results which were obtained by Hochwald and Marzetta in their initial paper concern the capacity of the noncoherent channel. In particular, they derive necessary criteria which codewords have to satisfy in order to achieve the channel capacity. This does essentially mean that it is not worthwhile to study codewords which do not satisfy these criteria.

Later in this work analogous results will be obtained using a different approach Therefore, the two respective theorems are cited here without going into too much detail regarding the used terminology. Nonetheless, their ramifications are elaborated upon afterwards.

(4.6) Theorem ([MH99] Theorem 1)
The capacity that can be obtained with $M > T$ transmitter antennas is the same as the capacity obtained with $M = T$ antennas. ◇

This result provides the fact that it is not possible to enhance the communication by making the number of transmitter antennas larger than the number T of time steps over which the channel remains invariant. Therefore, only codewords $X \in \mathbb{C}^{M\times T}$ for $M \leq T$ need to be considered.

Recall that an *isotropically distributed unitary matrix* has a probability distribution that remains unchanged when that matrix is multiplied by any deterministic unitary matrix.

(4.7) Theorem ([MH99] Theorem 2)
The signal matrix that achieves capacity can be written as $X = DU$, where U is a $T \times T$ isotropically distributed unitary matrix, and D is an independent $T \times M$ real, nonnegative, diagonal matrix. Furthermore the joint density of the diagonal elements of V can be chosen such that it is unchanged by rearrangements of its arguments. ◇

The main implication of this theorem is that the capacity does not change when a codeword is multiplied by a unitary matrix from the left.

Using these results the authors aimed at finding bounds on the capacity of the communications channel. These are hard to obtain in general and therefore, they restricted themselves to the case of codewords of the form $X = DU$ with a real diagonal matrix $D = \rho I_{M \times T} \in \mathbb{C}^{M \times T}$ for $M < T$, some $\rho \in \mathbb{R}_{>0}$ and a unitary matrix $U \in U(T)$. These codewords satisfy $X\overline{X}^\top = \rho \cdot I_M$.

This restriction lead to the suggestion of unitary space-time modulation in their subsequent work ([HM00]).

4.2 Unitary codes

In the following, unitary codes will be formally defined, it will be studied how they may be decoded and how good unitary codes can be constructed. The reasoning mainly follows the consideration of the noncoherent channel in [HM00].

(4.8) Definition
Suppose $M < T$ holds and consider $\rho \in \mathbb{R}_{>0}$. A unitary noncoherent STBC is a finite subset

$$\mathcal{C} \subseteq \{X \in \mathbb{C}^{M \times T} \mid X\overline{X}^\top = \rho I_M\}.$$

◇

This definition implies $\|X\|_F = \sqrt{\mathrm{Tr}(X\overline{X}^\top)} = \sqrt{\rho M}$ for any codeword X of a unitary code. Thus, by means of lemma 4.2, the signal-to-noise ratio can be explicitly computed.

(4.9) Lemma
Consider a unitary codebook $\mathcal{C} \subseteq \{X \in \mathbb{C}^{M \times T} \mid X\overline{X}^\top = \rho I_M\}$ for $\rho \in \mathbb{R}_{>0}$. If this codebook is used to transmit information, the signal-to noise ratio of the channel is given by

$$\frac{E[\|HX\|_F^2]}{E[\|V\|_F^2]} = \frac{E[\|X\|_F^2]}{T} = \rho \frac{M}{T}.$$

◇

Furthermore, the restriction to unitary codes yields a significant simplification to the error probability. In order to obtain this simplification, the following lemma is needed.

(4.10) Lemma
Suppose $X\overline{X}^\top = \rho I_M$ for $\rho \in \mathbb{R}_{>0}$, $X \in \mathbb{C}^{M\times T}$ and consider $\Lambda_X = I_T + \overline{X}^\top X$.

Then, the inverse of Λ_X is given by

$$\Lambda_X^{-1} = I_T - \frac{1}{1+\rho}\overline{X}^\top X.$$

Furthermore, its determinant is given by $\det(\Lambda_X) = (1+\rho)^M$. ◇

Proof. The statement on the inverse of Λ_X is easily checked by direct calculations:

$$\begin{aligned}(I_T + \overline{X}^\top X)(I_T - \frac{1}{1+\rho}\overline{X}^\top X) &= I_T + \overline{X}^\top X - \frac{1}{1+\rho}\overline{X}^\top X - \frac{\rho}{1+\rho}\overline{X}^\top X \\ &= I_T.\end{aligned}$$

Concerning the determinant, note that Λ_X is Hermitian and hence diagonalizable. Therefore, its determinant is given by the product of its eigenvalues. By lemma 2.7 the eigenvalues of $\overline{X}^\top X$ are the same as the ones of $X\overline{X}^\top = \rho \cdot I_M$ plus additional $T-M$ times the eigenvalue 0. Therefore, $\Lambda_X = I_T + \overline{X}^\top X$ has the eigenvalues $1+\rho$ of multiplicity M and 1 of multiplicity $T-M$, which yields the assertion. □

Using the formulas for Λ_X^{-1} and $\det(\Lambda_X)$ from the lemma, the error-probability from proposition 4.5 can be reformulated.

(4.11) Proposition
Consider a unitary codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ and suppose $X \in \mathcal{C}$ is sent across the noncoherent channel. The probability that a message $Y \in \mathbb{C}^{N\times T}$ is received at the decoder is given by

$$P(Y|X) = \frac{\exp(-\operatorname{Tr}(\overline{Y}^\top Y))}{\pi^{TN}(1+\rho)^{MN}} \exp\left(\frac{1}{1+\rho}\operatorname{Tr}(\overline{X}^\top X\overline{Y}^\top Y)\right).$$

◇

With this simplified probability, the problem of decoding a received codeword will now be addressed. To that end, suppose a unitary codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ is used to transmit a codeword across the channel and a message $Y \in \mathbb{C}^{N\times T}$ is received. With the information available at the decoder, the codeword that was most likely sent is the one $X \in \mathcal{C}$ for which $P(Y|X)$ is maximal. By the preceding proposition, $P(Y|X)$ does depend on the specific codeword X only through the expression $\operatorname{Tr}(\overline{X}^\top X\overline{Y}^\top Y) = \left\|X\overline{Y}^\top\right\|_F^2$. This yields the following simple corollary.

(4.12) Corollary (GLRT criterion)
Suppose that a unitary code $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ is used to transmit over the channel and a message $Y \in \mathbb{C}^{N \times T}$ is received. The maximum likelihood decoding problem is to find the codeword

$$\operatorname{argmax}_{X \in \mathcal{C}} \left(\left\| X \overline{Y}^{\top} \right\|_F \right).$$

This decoding criterion is commonly referred to as the *GLRT criterion* and a decoder which applies this criterion is referred to as a *GLRT decoder*. ◇

The abbreviation GLRT stands for *generalized likelihood ratio test*. This is a common statistical test, the details of which are not important in this context.

Now that the maximum likelihood decoding rule has been established, it will be investigated how a codebook has to be constructed in order to minimize the probability of decoding errors. To this end, consider the case that a codeword X_2 has been sent and a message Y is received at the decoder. In order for a codebook to perform well, the probability of Y being decoded erroneously to $X_1 \neq X_2$ should be as small as possible. Hence, the probability $P(X_1 \mid X_2)$, of a codeword X_1 being decoded if X_2 was sent, has to be investigated.

Using elaborate computations, in [HM00] Hochwald and Marzetta gave an exact formula for the probability $P(X_1 \mid X_2)$ for codewords X_1, X_2 of a unitary code. This formula involves residues of functions depending on the singular values of the matrix $X_1 \overline{X_2}^{\top}$. It is not easy to evaluate, but they were able to give its Chernoff upper bound, yielding the assertion of the following proposition.

(4.13) Proposition
Suppose that a unitary code $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ is used to transmit over the channel and that at the receiving end, a decoder applying maximum likelihood decoding is used. The probability that the received message is decoded to the codeword $X_1 \in \mathcal{C}$ if the codeword $X_2 \in \mathcal{C}$ was sent is bounded by

$$P(X_1 \mid X_2) \leq \frac{1}{2} \prod_{i=1}^{M} (1 + (\vartheta(1 - \sigma_i^2))^{-N}.$$

Here, $\sigma_1, \ldots, \sigma_M$ denote the singular values of $X_1 \overline{X_2}^{\top}$ and ϑ is some positive real number depending on the signal-to-noise ratio of the channel. ◇

In order to find unitary codebooks which guarantee a small probability of decoding errors, the given bound should be minimized for distinct codewords X_1, X_2. Constructions which build on this approach usually use even simpler estimates for that bound. These are obtained by multiplying out $\prod_{i=1}^{M}(1+(\vartheta(1-\sigma_i^2))$ and considering it as polynomial in ϑ. The square roots of the coefficients of ϑ and ϑ^M are respectively given by

1) $\sqrt{\sum_{i=1}^{M}(1-\sigma_i^2)}$, this term is commonly referred to as the *diversity sum* and

2) $\sqrt{\prod_{i=1}^{M}(1-\sigma_i^2)}$, which is commonly referred to as the *diversity product*.

Minimizing the bound for the error probability can now be approximated by minimizing the diversity sum and/or the diversity product. If the signal-to-noise ratio is very high, so is ϑ^M and therefore the diversity product has the larger influence on the bound. For the most common signal-to-noise ratio ranges, the diversity sum is used as the main criterion to optimize codebooks.

(4.14) Remark
By lemma 2.12 the equation $\left\|X_1\overline{X_2}^\top\right\|_F^2 = \sum_{i=1}^{M}\sigma_i^2$ holds for the singular values $\sigma_1,\ldots,\sigma_M$ of $X_1\overline{X_2}^\top$. Therefore, the diversity sum may be written as

$$\sqrt{\sum_{i=1}^{M}(1-\sigma_i^2)} = \sqrt{M - \left\|X_1\overline{X_2}^\top\right\|_F^2}.$$

This only depends on $\left\|X_1\overline{X_2}^\top\right\|_F$ and hence maximizing the diversity sum is equivalent to minimizing $\left\|X_1\overline{X_2}^\top\right\|_F$ for distinct codewords X_1, X_2. Due to the connection with the GLRT-criterion, the value $\left\|X_1\overline{X_2}^\top\right\|_F$ is often referred to as the *GLRT distance* of X_1 and X_2. ◇

As the diversity sum and the diversity product depend on codewords X_1 and X_2 only through the singular values of $X_1\overline{X_2}^\top$, they agree with the assertion from theorem 4.7: If a codeword X_1 or X_2 is multiplied by a square unitary matrix from the left, the singular values of $X_1\overline{X_2}^\top$ do not change and hence a codebook cannot be enhanced by doing that. On the other hand, suppose two codewords X_1, X_2 only differ by left multiplication from a unitary matrix, say $X_1 = UX_2$ for $U \in U(M)$. The GLRT-decoder cannot distinguish these codewords:

$$\left\|X_1\overline{Y}^\top\right\|_F = \left\|UX_2\overline{Y}^\top\right\|_F = \left\|X_2\overline{Y}^\top\right\|_F.$$

Therefore, it makes sense to further restrict the set from which the codewords are taken from rectangular unitary matrices to equivalence classes of such with respect to left multiplication by a square unitary matrix. In the following section, it will be shown that any of these equivalence classes may be identified with the vector space which is spanned by the rows of any representative of such a class. In doing so, codewords are identified with M-dimensional subspaces of $\mathbb{C}^T$. The set of M-dimensional subspaces of $\mathbb{C}^T$ can be given the structure of a manifold and as such it is known as the *Grassmann manifold*.

4.3 A connection to packing problems on the Grassmann manifold

In the following, the problem of finding good unitary codebooks will be formulated as a packing problem on the Grassmann manifold. The connection between unitary codes and the Grassmann manifold has been described by Zeng and Tse in [ZT02]. Many constructions of unitary codes have been proposed using this connection.

(4.15) Definition (Grassmann manifold)
For $X \in \mathbb{C}^{M \times T}$ denote by $\langle X \rangle$ the span of the rows of X.

The complex *Grassmann manifold* $G^{\mathbb{C}}_{M,T}$ is defined as the set of M-dimensional subspaces of $\mathbb{C}^T$. It may be written as

$$G^{\mathbb{C}}_{M,T} = \{\langle X \rangle \mid X \in \mathbb{C}^{M \times T}, X\overline{X}^{\top} = \rho I_M\}$$

for any $\rho \in \mathbb{R}_{>0}$. ◇

Take $k \in \mathbb{N}$ and consider a set of k distinct subspaces $\mathcal{C}' = \{\langle X_1 \rangle, \ldots, \langle X_k \rangle\} \subseteq G^{\mathbb{C}}_{M,T}$. Additionally suppose that the representatives $X_1, \ldots, X_k$ are chosen such that they satisfy $X_i\overline{X_i}^{\top} = \rho I_M$ for $i = 1, \ldots, k$ and some $\rho \in \mathbb{R}_{>0}$. A corresponding unitary code may then be defined as

$$\mathcal{C} := \{X_1, \ldots, X_k\}.$$

The definition guarantees that for $X \in \mathcal{C}$ the matrix UX cannot be in $\mathcal{C}$ for any unitary matrix $U \in U(M) \setminus \{I_M\}$. This becomes obvious by the following remark which is easily checked by standard linear algebra methods.

(4.16) Remark
Two matrices $X_1, X_2 \in \mathbb{C}^{M \times T}$ satisfying $X_1\overline{X_1}^{\top} = X_2\overline{X_2}^{\top} = I_M$ have the same row span if and only if there is a unitary matrix $U \in U(M)$ such that $X_1 = UX_2$ holds. ◇

Therefore, the problem that the GLRT-decoder may not be able to distinguish between two codewords, as discussed at the end of the previous section, is avoided.

For general unitary codes, it was found that the GLRT-distance $\left\| X_1 \overline{X_2}^\top \right\|_F$ of two distinct codewords X_1, X_2 has to be maximized in order for the codebook to perform well. It will now be investigated what this condition translates into for codebooks built from the Grassmann manifold. To that end the following definition is given.

(4.17) Definition and Lemma
Consider $X_1, X_2 \in \mathbb{C}^{M \times T}$ satisfying $X_1 \overline{X_1}^\top = X_2 \overline{X_2}^\top = I_M$ and denote by $\sigma_1, \dots, \sigma_M$ the singular values of $X_1 \overline{X_2}^\top$.

The values $\theta_i := \arccos(\sigma_i)$ for $i = 1, \dots, M$ are called the *principle angles* between $\langle X_1 \rangle$ and $\langle X_2 \rangle$.

By the preceding remark, this definition is independent of the choice of representatives of the equivalence classes. It may also be checked that for $i = 1, \dots, M$ all σ_i satisfy $0 \leq \sigma_i \leq 1$. Therefore, the expression $\arccos(\sigma_i)$ is well defined. ◇

This enables the definition of two notions of distance on the Grassmann manifold.

(4.18) Definition
Consider $\langle X_1 \rangle, \langle X_2 \rangle \in G^{\mathbb{C}}_{M,T}$ and denote by $\theta_1, \dots, \theta_M$ the *principle angles* between X_1 and X_2.

(i) The *chordal distance* between $\langle X_1 \rangle$ and $\langle X_2 \rangle$ is defined as:

$$d_c(\langle X_1 \rangle, \langle X_2 \rangle) = \sqrt{\sum_{i=1}^{M} \sin(\theta_i)^2}.$$

(ii) The *geodesic distance* between $\langle X_1 \rangle$ and $\langle X_2 \rangle$ is defined as:

$$d_g(\langle X_1 \rangle, \langle X_2 \rangle) = \sqrt{\sum_{i=1}^{M} \theta_i^2}$$

◇

The chordal distance between $\langle X_1 \rangle$ and $\langle X_2 \rangle$ can now be put in relation with the GLRT-distance of X_1 and X_2.

(4.19) Proposition
Consider a unitary codebook $\mathcal{C} \subseteq \{X \in \mathbb{C}^{M \times T} \mid X\overline{X}^\top = \rho I_M\}$ for some positive $\rho \in \mathbb{R}$.

The GLRT-distance $\left\|X_1\overline{X_2}^\top\right\|_F$ for distinct $X_1, X_2 \in \mathcal{C}$ is minimized if and only if the chordal distance $d_c(\langle X_1\rangle, \langle X_2\rangle)$ is maximized. ◇

Proof. Consider two distinct codewords $X_1, X_2 \in \mathcal{C}$. In terms of the singular values $\sigma_1, \dots, \sigma_M$ of $X_1\overline{X_2}^\top$, their GLRT-distance is given by

$$\left\|X_1\overline{X_2}^\top\right\|_F = \sqrt{\sum_{i=1}^{M} \sigma_i^2}.$$

Note that $\langle X_i\rangle = \langle \frac{1}{\sqrt{\rho}} X_i\rangle$ and $(\frac{1}{\sqrt{\rho}}X_i)\overline{(\frac{1}{\sqrt{\rho}}X_i)}^\top = I_M$ hold for $i = 1, 2$. Therefore, the singular values of $(\frac{1}{\sqrt{\rho}}X_1)\overline{(\frac{1}{\sqrt{\rho}}X_2)}^\top$ are given by $\frac{\sigma_1}{\rho}, \dots, \frac{\sigma_M}{\rho}$ and hence, the principal angles between $\langle X_1\rangle$ and $\langle X_2\rangle$ are given by $\arccos(\frac{\sigma_1}{\rho}), \dots, \arccos(\frac{\sigma_M}{\rho})$. The chordal distance between $\langle X_1\rangle$ and $\langle X_2\rangle$ may now be computed explicitly:

$$\begin{aligned} d_c(\langle X_1\rangle, \langle X_2\rangle) &= \sqrt{\sum_{i=1}^{M} \sin(\theta_i)^2} = \sqrt{\sum_{i=1}^{M} 1 - \cos(\theta_i)^2} \\ &= \sqrt{\sum_{i=1}^{M} (1 - \frac{\sigma_i^2}{\rho^2})} = \sqrt{M - \frac{1}{\rho^2}\left\|X_1\overline{X_2}^\top\right\|_F^2}. \end{aligned}$$

The right hand side of the equation is maximized if and only if $\left\|X_1\overline{X_2}^\top\right\|_F$ is minimized. This yields the assertion of the proposition. □

By the preceding proposition, good unitary codes can be obtained by finding packings on the Grassmann manifold that possess a large minimal distance with respect to the chordal distance. This section will now be concluded by a brief summary of selected results on finding such packings.

In [Cre07], Creignou was able to find a class of optimal packings for this case, using representations of finite groups. However, the applied method only yields packings for certain tuples $(|\mathcal{C}|, M, T)$ which limits its practical relevance.

In order to find good but suboptimal packings, Henkel showed in [Hen05] that the chordal distance is locally equivalent to the geodesic distance on the Grassmann manifold. Building on this, in [UCL08] Utkovski, Chen and Lindner present a method to obtain good packings on the Grassmann manifold with respect to the geodesic distance.

Altogether it remains a hard problem to find good packings on the Grassmann manifold. Therefore, numerical methods are an important tool and worth discussing.

4.4 A numerical algorithm for Grassmannian packings

Since the problem of finding optimal packings on the Grassmann manifold with respect to the chordal distance turns out to be hard, it is of interest to compute good packings by numerical means. This section elaborates on one approach to this end. It is included mainly because the resulting algorithm was implemented to obtain approximately optimal unitary codes which could be compared to newly constructed codes in simulations. However, the apparent complexity of the optimization algorithm also underlines the difficulty of the problem of finding optimal unitary codes.

Analogously to the complex Grassmann manifold defined in the previous section, one may define a real Grassmann manifold $G^{\mathbb{R}}_{M,T}$ as the set of M-dimensional subspaces of $\mathbb{R}^T$. The problem of finding optimal packings on the real Grassmann manifold has originally been investigated by Conway, Hardin and Sloane in [CHS96]. Their methods were extended to the complex Grassmann manifold by Agrawal, Richardson and Urbanke in [ARU01]. They propose using a relaxation technique combined with a gradient search algorithm which will be described in this subsection.

Formally, in order to obtain an optimal unitary codebook of cardinality n, the map

$$\left(G^{\mathbb{C}}_{M,T}\right)^n \to \mathbb{R},\ (\langle X_1\rangle, \ldots, \langle X_n\rangle) \mapsto \min_{1\leq i<j\leq n} d_c(\langle X_i\rangle, \langle X_j\rangle)$$

has to be maximized. In practice a unitary codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ is considered and its codewords are identified with the corresponding elements in $G^{\mathbb{C}}_{M,T}$. According to proposition 4.19 minimizing the above map is then equivalent to maximizing the value

$$\sigma^*(\mathcal{C}) := \max_{X_1\neq X_2\in\mathcal{C}} \left(\left\|X_1\overline{X_2}^\top\right\|_F\right).$$

This defines a continuous function on a compact differentiable manifold and hence possesses a global minimum. Unfortunately the function σ^* is not differentiable everywhere and therefore standard gradient descent algorithms are not directly applicable to obtain that minimum. In order to overcome this obstacle it is possible to approximate σ^* closely by a smooth map. It then turns out, however, that in general σ^* has a large number of local minima which may be far from the global minimum. Therefore standard gradient descent algorithms applied to a close approximation of σ^* will get caught in these local minima and yield highly suboptimal results.

The authors propose to overcome this obstacle by relaxing the original problem with the help of a family of surrogate functionals $(f_\alpha)_{\alpha\in\mathbb{R}}$ which satisfy the following properties:

1) For all $\alpha \in \mathbb{R}$ the map f_α is smooth.

2) For small values of α the map f_α has few local minima.

3) The maps f_α mimic σ^*, that is for large values of α, the f_α closely track σ^* and in particular satisfy $\lim_{\alpha\to\infty} f_\alpha = \sigma^*$.

As an example of such a family they propose

$$f_\alpha(\mathcal{C}) = \frac{1}{\alpha}\log\left(\sum_{X_1\neq X_2\in\mathcal{C}} \exp\left(\alpha \left\| X_1\overline{X_2}^\top \right\|_F^2\right)\right).$$

With these prerequisites a starting set $\mathcal{C}$ may be chosen and optimized by a gradient search algorithm with respect to f_α for gradually growing values of α.

Using this method it is guaranteed to obtain a local minimum of σ^* and the chances are increased that it may be in fact the global minimum. However, it may still be far from it. In practice it is possible to find good packings by randomly choosing many different input sets, running them through the optimization process and choosing the best result in the end.

The described method is not very satisfactory from a theoretical point of view as there is no guarantee that the optimal result will be obtained and no estimation on how close the result is to the optimum can be given. Furthermore the relaxation process requires the repeated application of a gradient descend algorithm and that process itself has also to be repeated multiple times with various random input sets until a satisfying result is found. In conclusion it is computationally very expensive to use this technique.

However, the authors achieved close approximations of the optimal value for $\sigma^*(\mathcal{C})$ in instances where that value is known. For cases where the optimum is not known they could enhance on the best previously known results. For this reason unitary codes obtained by this method will be used for comparison with other codes later on.

§ 5 A new approach to noncoherent STBC

All known constructions of noncoherent STBC focus on the case where all codewords are (rectangular) unitary matrices. This assumption will not be made in the following. Instead, the objective will be to develop a theory that allows studying noncoherent STBC as generally as possible.

The starting point for developing such a theory has to be the communication model of the noncoherent channel. For this model some basic assertions on the probability distribution of received codewords have been introduced in section 4. These probabilities will be used to formulate the maximum likelihood decoding rule as generally as possible. This rule in turn will be closely examined in order to find design criteria which allow the construction of codebooks that guarantee a small error rate.

Particularly finding useful design criteria turns out to be a challenging problem. Therefore, at some points, additional assumptions will be made. These will, however, still generalize the case of unitary coding.

5.1 Decoding and the design criterion

It was seen in Proposition (4.5) that the probability of a message $Y \in \mathbb{C}^{N \times T}$ being received if the codeword $X \in \mathbb{C}^{M \times T}$ was sent is given by

$$P(Y|X) = \frac{\exp(-\operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^\top Y))}{\pi^{TN}\det(\Lambda_X)^N}.$$

In this formula Λ_X denotes the covariance matrix of any row of the received message Y. It was shown in lemma 4.3 that it is given by $\Lambda_X = I_T + \overline{X}^\top X$. For a given codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ the values M and T are fixed. Therefore the value π^{TN} is a scalar independent of the codeword X and the received message Y. Hence it can be dropped in order to find the largest probability throughout the codebook and the following result for the the maximum likelihood decoding problem is obtained immediately.

(5.1) Proposition
For a codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ and a received message $Y \in \mathbb{C}^{N \times T}$ the maximum likelihood decoding problem is to find the codeword

$$\operatorname{argmax}_{X \in \mathcal{C}} P(Y|X) = \operatorname{argmax}_{X \in \mathcal{C}} \frac{\exp(-\operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^\top Y))}{\det(\Lambda_X)^N}.$$

Since the values $\det(\Lambda_X)^N$ and the elements Λ_X^{-1} can be precomputed for all codewords C, maximum likelihood decoding by exhaustive search may be implemented efficiently using this criterion. However, decoding by exhaustive search is not satisfactory for most practical applications and therefore a simpler decoding criterion for which lower complexity algorithms are available is desirable.

An even larger issue is that the decoding rule does not provide us with a useful criterion for the design of noncoherent codebooks which ensures that the probability of decoding errors is small. To overcome this issue restrictions will be imposed on the codewords that allow the decoding rule to be simplified. This simplified decoding rule will then be used to deduce a useful design criterion.

In particular, suppose $\det(\Lambda_X)$ takes a fixed value for all codewords X in a codebook $\mathcal{C}$. Under that assumption the decoding rule from the above proposition can be equivalently reformulated as follows.

(5.2) Corollary
Let $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ be a codebook such that $\det(\Lambda_X)$ takes a constant value for all $X \in \mathcal{C}$. Then the maximum likelihood decoding problem is to find

$$\operatorname{argmax}_{X \in \mathcal{C}} P(Y|X) = \operatorname{argmin}_{X \in \mathcal{C}} \operatorname{Tr}(\Lambda_X^{-1} \overline{Y}^\top Y).$$

Proof. At first note that according to lemma 4.3 the matrix Λ_X is a positive definite Hermitian matrix and therefore $\det(\Lambda_X)$ is a positive real number for any $X \in \mathcal{C}$. Hence if $\det(\Lambda_X)$ is constant for all $X \in \mathcal{C}$, the equality

$$\operatorname{argmax}_{X \in \mathcal{C}} \frac{\exp(-\operatorname{Tr}(\Lambda_X^{-1} \overline{Y}^\top Y))}{\det(\Lambda_X)^N} = \operatorname{argmax}_{X \in \mathcal{C}} \exp(-\operatorname{Tr}(\Lambda_X^{-1} \overline{Y}^\top Y))$$

is obvious.

Furthermore note that with Λ_X also Λ_X^{-1} is positive definite Hermitian and $\overline{Y}^\top Y$ is positive semidefinite Hermitian. By lemma 2.13 this implies that the trace of $\Lambda_X^{-1} \overline{Y}^\top Y$ is a real number. Since the exponential map $\exp : \mathbb{R} \to \mathbb{R}$ is strictly monotonously growing, the value $\exp(-\operatorname{Tr}(\Lambda_X^{-1} \overline{Y}^\top Y))$ is maximized if and only if $\operatorname{Tr}(\Lambda_X^{-1} \overline{Y}^\top Y)$ is minimized. This yields the assertion of the corollary. □

After having established this simplified decoding rule the next aim is to find a criterion which allows the systematical design of good codebooks. In the process, a codebook is considered to be good if it yields a low pairwise error probability. That means, if a codeword X_1 from a codebook $\mathcal{C}$ was sent across the channel and a message Y is received, the values $P(X_2|Y)$ are desired to be as small as possible for $X_2 \in \mathcal{C} \setminus \{X_1\}$. By the previous corollary this is equivalent to the values $\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y)$ being as large as possible for $X_2 \neq X_1$. These values depend on the matrix Y which may be regarded as a random variable depending on the codeword X_1 that has been sent across the channel. Therefore, the values in question need to be investigated depending on the probability distribution of Y.

One possible approach to do that would be to find a lower bound on $\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y)$ depending on X_1 and to look for codebooks which are optimal with respect to that bound. This has essentially been done to obtain a design criterion for unitary codes as described in section 4.2.

Another approach, and the one that will be pursued here, is to evaluate the expected value of $\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y)$. It will be used to deduce the main criterion for designing good codebooks in the following. Contrary to a lower bound that expected value is easy to calculate.

(5.3) Lemma
Suppose that the codeword $X_1 \in \mathbb{C}^{M\times T}$ is sent across the channel and that a message $Y \in \mathbb{C}^{N\times T}$ is received. Furthermore consider another codeword $X_2 \in \mathbb{C}^{M\times T}$. The expected value of $\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y)$ is given by

$$E(\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y))) = N \cdot \mathrm{Tr}(\Lambda_{X_2}^{-1}\Lambda_{X_1}).$$

Proof. In the proof of lemma 4.3 it was seen that for any $i \in \{1,\dots,N\}$ and $j,k \in \{1,\dots,T\}$ the equality $(\Lambda_X)_{jk} = E[\overline{Y_{ij}}Y_{ik}]$ holds. This implies

$$E[(\overline{Y}^\top Y)_{kj}] = \sum_{i=1}^{N} E[\overline{Y_{ij}}Y_{ik}] = N \cdot (\Lambda_X)_{kj}.$$

That is $E[\overline{Y}^\top Y] = N \cdot \Lambda_X$. Considering this as well as the linearity of the expected value and the fact that the entries of Λ_{X_2} are deterministic, the following equations are seen to be true:

$$\begin{aligned} E\left[\mathrm{Tr}(\Lambda_{X_2}^{-1}\overline{Y}^\top Y)\right) &= E\left(\sum_{i=1}^{T}\sum_{j=1}^{T}(\Lambda_{X_2}^{-1})_{ij}(\overline{Y}^\top Y)_{ji}\right] \\ &= \sum_{i=1}^{T}\sum_{j=1}^{T}(\Lambda_{X_2}^{-1})_{ij}E\left[(\overline{Y}^\top Y)_{ji}\right] \\ &= \mathrm{Tr}(\Lambda_{X_2}^{-1}E[\overline{Y}^\top Y]) \\ &= \mathrm{Tr}(\Lambda_{X_2}^{-1}(N\cdot\Lambda_{X_1})) \\ &= N\cdot\mathrm{Tr}(\Lambda_{X_2}^{-1}\Lambda_{X_1}). \end{aligned}$$ □

With the help of this lemma the problem of designing good codebooks can be regarded as a packing problem with respect to the expression that was computed for the expected value. In other words, the respective expression yields a measure of distance between codewords and will therefore play an important role. This notion is formalized in the following definition.

(5.4) Definition
The following map serves as a distance function between codewords:

$$\Delta: \mathbb{C}^{M\times T}\times\mathbb{C}^{M\times T}\to\mathbb{R},\ (X_1,X_2)\mapsto\mathrm{Tr}((\Lambda_{X_2})^{-1}\Lambda_{X_1}).$$

Utilizing this notation the problem of designing good noncoherent codebooks that has been derived up to this point can be formally summarized.

(5.5) Design Criteria for noncoherent STBC (1)
Let $\mathcal{C}\subseteq\mathbb{C}^{M\times T}$ be a finite set of complex matrices. For $\mathcal{C}$ to be used as a codebook for communicating over a channel with a SNR of ρ the following conditions have to be satisfied:

1) $\frac{1}{|\mathcal{C}|}\sum_{X\in\mathcal{C}}\|X\|_F=\rho$,

2) $\det(\Lambda_X)$ is constant for all $X\in\mathcal{C}$,

3) $\min_{X_1,X_2\in\mathcal{C},\,X_1\neq X_2}\Delta(X_1,X_2)$ is maximized. ◇

Part 1) can be satisfied for any finite set of codewords by rescaling them appropriately. It mainly influences the design problem as a constraint to the optimization problem given by part 3). Part 2) may be achieved in several ways by restricting codewords to suitable subsets of $\mathbb{C}^{M\times T}$. Some approaches will be discussed later on. This does however leave part 3) as a difficult optimization problem.

5.2 Reduced codebooks

In a first effort to simplify the design problem, note that the map

$$\Delta : \mathbb{C}^{M\times T} \times \mathbb{C}^{M\times T} \to \mathbb{R},\ (X_1, X_2) \mapsto \operatorname{Tr}((\Lambda_{X_2})^{-1}\Lambda_{X_1})$$

depends on its arguments X_1, X_2 only through the matrices Λ_{X_1} and Λ_{X_2}. In particular, for any codewords X_1, X_2 with $\Lambda_{X_1} = \Lambda_{X_2}$ their "distance" is given by $\Delta(X_1, X_2) = T$ which is equal to the "distance" of any codeword X to itself. Furthermore when decoding a message $Y \in \mathbb{C}^{N\times T}$ with the decoding rule

$$\operatorname{argmin}_{X\in\mathcal{C}} \operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^{\top}Y),$$

the two codewords X_1 and X_2 yield the same value and can therefore not be distinguished by the receiver.

For these reasons it makes sense to prevent codebooks from containing two codewords X_1, X_2 that yield the same matrix $\Lambda_{X_1} = \Lambda_{X_2}$. A property which makes sure this cannot be the case is introduced in the following definition.

(5.6) Definition
A codebook $\mathcal{C} \subseteq C$ is called Λ-*reduced* if $|\mathcal{C}| = |\{\Lambda_X \mid X \in \mathcal{C}\}|$ holds. $\diamond$

This property is characterized in more detail in the next proposition which also makes clear that the problem described above cannot arise for Λ-reduced codes.

(5.7) Proposition
Consider a codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$. The following are equivalent:

(i) $\mathcal{C}$ is Λ-reduced.

(ii) For distinct $X_1, X_2 \in \mathcal{C}$ there is no $U \in U(M)$ such that $X_1 = UX_2$ holds.

(iii) The map $\varphi : \mathcal{C} \to \{\Lambda_X \mid X \in \mathcal{C}\}$, $X \mapsto \Lambda_X$ is a bijection. $\diamond$

Proof. The map φ in (iii) is surjective by definition. Since $\mathcal{C}$ and hence $\{\Lambda_X \mid X \in \mathcal{C}\}$ are finite the equivalence (i) $\iff$ (iii) follows.

Now suppose there are $X_1, X_2 \in \mathcal{C}$ and $U \in U(M)$ such that $X_1 = UX_2$ holds. This implies

$$\Lambda_{X_1} = I_T + \overline{X_1}^{\top}X_1 = I_T + \overline{X_1}^{\top}\overline{U}^{\top}UX_1 = I_T + \overline{X_2}^{\top}X_2 = \Lambda_{X_2},$$

therefore the map φ is not injective. That shows (iii) $\Longrightarrow$ (ii).

Finally suppose φ is not bijective. As it is always surjective by definition, this implies that φ is not injective. In this case there are distinct codewords $X_1, X_2 \in \mathcal{C}$ such that $\Lambda_{X_1} = I_T + \overline{X_1}^\top X_1 = I_T + \overline{X_2}^\top X_2 = \Lambda_{X_2}$ holds. This implies the equality $\overline{X_1}^\top X_1 = \overline{X_2}^\top X_2$.

Particularly $\overline{X_1}^\top X_1$ and $\overline{X_2}^\top X_2$ have the same eigenvalues and corresponding eigenvectors. By lemma 2.7 the matrices X_1 and X_2 therefore have the same singular values and they have singular value decompositions $X_1 = U_1 DV$ and $X_2 = U_2 DV$ for $U_1, U_2 \in U(M)$ and a real diagonal matrix D and $V \in U(T)$.

In conclusion, $X_1 = (U_1 U_2^{-1}) X_2$ holds with $U_1 U_2^{-1} \in U(M)$ and therefore the implication (ii) $\Longrightarrow$ (iii) holds. □

In order to formalize the process of finding Λ-reduced codes, property (ii) of the proposition can be reformulated in a group theoretic manner. To that end, a brief description of the left group action of $U(M)$ on $\mathbb{C}^{M \times T}$ is given.

(5.8) Remark
The group $U(M)$ acts from the left on $\mathbb{C}^{M \times T}$ by the usual matrix multiplication. The orbits under this action are given by $U(M) \cdot X = \{UX \mid U \in U(M)\}$ for $X \in \mathbb{C}^{M \times T}$. $\mathbb{C}^{M \times T}$ can be decomposed into a disjoint union of these orbits. ◇

These group theoretic terms allow for the following characterization of Λ-reduced codebooks.

(5.9) Corollary
Consider $\mathbb{C}^{M \times T}$ endowed with the canonical left group action of $U(M)$.

$\mathcal{C}$ is Λ-reduced if and only if each group orbit contains at most one element of $\mathcal{C}$. ◇

This characterization can be used to connect Λ-reduced codebooks to certain unitary codes which have been introduced earlier. In particular, in section 4.3 it was concluded that it is sufficient to take unitary codewords from the Grassmann manifold. This can now be justified as a strategy which guarantees for the resulting codebooks to be Λ-reduced.

(5.10) Example (Λ-reduced unitary codes)
By remark 4.16, for $M \leq T$ two matrices $X_1, X_2 \in \mathbb{C}^{M \times T}$ satisfying an equation $X_1 \overline{X_1}^\top = X_2 \overline{X_2}^\top = \rho I_M$ for some $\rho \in \mathbb{R}_{>0}$ have the same row span if and only if there is a unitary matrix $U \in U(M)$ satisfying $X_1 = UX_2$. Therefore, in light of the

previous corollary, a unitary code is Λ-reduced if and only if the row spans of all its elements are distinct. That is, all elements of the code correspond to different elements of the Grassmann manifold and consequently, unitary codes constructed from the Grassmann manifold as described in section 4.3 are Λ-reduced. ◇

5.3 Equivalent codebooks

In order to further simplify the design problem, the aim of this section is to characterize codes which perform equally well.

(5.11) Definition
Two codebooks $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ and $\mathcal{C}' \subseteq \mathbb{C}^{M' \times T}$ are called Λ-*equivalent* if

$$\{\Lambda_X \mid X \in \mathcal{C}\} = \{\Lambda_{X'} \mid X' \in \mathcal{C}'\}$$

holds. ◇

This notion does obviously define an equivalence relation. The following proposition elaborates to what extent it guarantees for two Λ-reduced codebooks, which are Λ-equivalent, to perform equally well.

(5.12) Proposition
Consider two Λ-reduced codebooks $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ and $\mathcal{C}' \subseteq \mathbb{C}^{M' \times T}$.

If $\mathcal{C}$ and $\mathcal{C}'$ are Λ-equivalent, there is a bijective map $\varphi : \mathcal{C} \to \mathcal{C}'$ which satisfies

$$\|X\|_F = \|\varphi(X)\|_F \quad \text{and} \quad \Delta(X_1, X_2) = \Delta(\varphi(X_1), \varphi(X_2))$$

for $X, X_1, X_2 \in \mathcal{C}$. ◇

Proof. If $\mathcal{C}$ and $\mathcal{C}'$ are Λ-reduced, by proposition 5.7 the maps

$$\varphi_1 : \mathcal{C} \to \{\Lambda_X \mid X \in \mathcal{C}\},\ X \mapsto \Lambda_X \quad \text{and} \quad \varphi_2 : \mathcal{C}' \to \{\Lambda_{X'} \mid X' \in \mathcal{C}'\},\ X' \mapsto \Lambda_{X'}$$

are bijections. If $\mathcal{C}$ and $\mathcal{C}'$ are furthermore Λ-equivalent, the map

$$\varphi := \varphi_2^{-1} \circ \varphi_1 : \mathcal{C} \to \mathcal{C}'$$

is well defined and hence a bijection that satisfies $\Lambda_X = \Lambda_{\varphi(X)}$ for $X \in \mathcal{C}$.

Using the latter equation to evaluate the map Δ for two codewords $X_1, X_2 \in \mathcal{C}$ yields

$$\Delta(X_1, X_2) = \operatorname{Tr}(\Lambda_{X_2}^{-1} \Lambda_{X_1}) = \operatorname{Tr}(\Lambda_{\varphi(X_2)}^{-1} \Lambda_{\varphi(X_1)}) = \Delta(\varphi(X_1), \varphi(X_2)).$$

For any codeword $X \in \mathcal{C}$ the value

$$\|X\|_F^2 = \operatorname{Tr}(\overline{X}^\top X) = \operatorname{Tr}(I_T + \overline{X}^\top X - I_T) = \operatorname{Tr}(\Lambda_X) - T$$

is uniquely determined by Λ_X. Therefore $\|X\|_F = \|\varphi(X)\|_F$. □

The previous proposition implies that Λ-reduced codebooks which are Λ-equivalent yield the same rate, as the codebooks clearly need to be of the same size. Furthermore, if one codebook is replaced by the other via the map φ, the communication will work at the same signal-to-noise-ratio since codewords identified via φ have the same Frobenius norm. Finally, since that identification also preserves the value of the map Δ, the expected probability of error is the same as well. Altogether, Λ-equivalent codebooks exchanged via the map φ can be used in the same situation and will perform equally well.

It is sufficient to make this observation for Λ-reduced codebooks as it was established earlier that codebooks which are not Λ-reduced will perform strictly worse. Furthermore it is easily seen that any codebook is Λ-equivalent to a Λ-reduced codebook.

(5.13) Lemma
Consider a codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$. There is a Λ-reduced codebook $\mathcal{C}'$ which is Λ-equivalent to $\mathcal{C}$. ◇

Proof. Define $\mathcal{C}' \subseteq \mathcal{C}$ to be a codebook such that for every $\Lambda \in \{\Lambda_X \mid X \in \mathcal{C}\}$ exactly one $X \in \mathcal{C}$ is contained in $\mathcal{C}'$ satisfying $\Lambda_X = \Lambda$. Then the codebook $\mathcal{C}'$ is Λ-reduced and Λ-equivalent to $\mathcal{C}$. □

The previous section on Λ-reduced codebooks lead to the conclusion that at most one codeword from an orbit of the left action of $U(T)$ on $\mathbb{C}^{M\times T}$ should be used in a codebook. Using the notion of Λ-equivalence, it can be seen that it does not matter which element from an orbit is chosen.

(5.14) Lemma
Consider a codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$. For some $X \in \mathcal{C}$ choose any $X' \in U(M)\cdot X$.

The codebook $\mathcal{C}' := (\mathcal{C} \setminus \{X\}) \cup \{X'\}$ is Λ-equivalent to $\mathcal{C}$. ◇

Proof. For $X \in \mathcal{C}$ consider $X' = UX \in U(M)\cdot X$ for $U \in U(M)$. The matrix Λ_X is the same as $\Lambda_{X'}$:

$$\Lambda_{X'} = I_T + \overline{UX}^\top UX = I_T + \overline{X}^\top \overline{U}^\top UX = I_T + \overline{X}^\top X = \Lambda_X.$$

Therefore the assertion of the lemma follows. □

Combining this with the earlier results on Λ-reduced codebooks yields the following important corollary.

(5.15) Corollary
Consider a codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ and a system of representatives $R \subseteq \mathbb{C}^{M \times T}$ for the orbits of $\mathbb{C}^{M \times T}$ under the left group action of $U(M)$.

There is a code $\mathcal{C}' \subset R$ which is Λ-equivalent to $\mathcal{C}$. ◇

Proof. By lemma 5.13 the codebook $\mathcal{C}$ is Λ-equivalent to a Λ-reduced codebook $\mathcal{C}''$. Iteratively applying lemma 5.14 to replace any element of $\mathcal{C}''$ by one from R yields a codebook $\mathcal{C}' \subseteq R$ which is Λ-equivalent to $\mathcal{C}''$ and hence also to $\mathcal{C}$. □

This result allows the space from which the codewords are drawn to be restricted significantly and so simplifies the design problem.

More simplifications can be made regarding the amount of transmit antennas used. By using information-theoretic arguments, it was already argued in [MH99] that it does not make sense to make the number of transmit antennas M larger than the number of timesteps T over which the channel remains invariant. This result can also be explained in terms of Λ-equivalence. In particular it will be shown that any codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ for $M > T$ is Λ-equivalent to a code consisting of square codewords. In fact it will even be shown that a codebook employing any arbitrary amount of transmit antennas is Λ-equivalent to such a code.

(5.16) Proposition
Let $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ be a codebook. There is a codebook $\mathcal{C}' \subseteq \mathbb{C}^{T \times T}$ which is Λ-equivalent to $\mathcal{C}$. ◇

Proof. The two cases $M > T$ and $M < T$ are distinguished. If $M = T$ holds, the assertion is obviously satisfied by choosing $\mathcal{C}' = \mathcal{C}$.

1) $M > T$:

Consider a codeword $X \in \mathcal{C}$ with singular value decomposition $X = UDV$ for $U \in U(M)$, $V \in U(T)$ and $D = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_M \\ & 0_{M-T \times T} & \end{pmatrix}$ where $\sigma_1, \ldots, \sigma_M$ denote the singular values of X.

Now define $D' := \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_M \end{pmatrix}$ and note that $\overline{D'}^\top D' = \overline{D}^\top D$ holds. Finally, setting $X' := D'V$ one obtains

$$\Lambda_{X'} = I_T + \overline{V}^\top \overline{D'}^\top D'V = I_T + \overline{V}^\top \overline{D}^\top DV = I_T + \overline{V}^\top \overline{D}^\top \overline{U}^\top UDV = \Lambda_X.$$

With that notation $\mathcal{C}' := \{X' \mid X \in \mathcal{C}\} \subseteq \mathbb{C}^{T\times T}$ is Λ-equivalent to $\mathcal{C}$.

2) $M < T$:

As in the first case consider a codeword $X = UDV$ for $U \in U(M)$, $V \in U(T)$ and $D = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & 0_{M\times T-M} \\ & & \sigma_M & \end{pmatrix}$ where $\sigma_1, \dots, \sigma_M$ denote the singular values of X.

The diagonal matrix $D' := \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_M & \\ & & & 0_{T-M\times T-M} \end{pmatrix}$ satisfies $\overline{D'}^\top D' = \overline{D}^\top D$.

Analogously to the first case set $X' := D'V$ and $\mathcal{C}' = \{X' \mid X \in \mathcal{C}\}$ to obtain the assertion. □

Due to this proposition it is sufficient to study square codes, that is codebooks containing square codewords. It may, of course, be the case that an optimally performing square code is Λ-equivalent to a codebook $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ for $M < T$. In that case the latter code would be preferable since it offers the same performance using fewer transmit antennas. This code may, however, be constructed from the square code that it is Λ-equivalent to. Case 2) of the proof of the above proposition yields a concrete way of doing that construction.

Summarizing the two simplifications found in this section, it is sufficient to study square codebooks $\mathcal{C} \subseteq R$, where R denotes a system of representatives of the orbits of $\mathbb{C}^{T\times T}$ under the left action of $U(M)$. It is therefore of interest to identify such a system of representatives.

(5.17) Lemma
A system of representatives for the orbits of the left group operation of $U(T)$ on $\mathbb{C}^{T\times T}$ is given by the set of positive semidefinite Hermitian $T \times T$ matrices

$$\mathcal{H}_0^+(T) = \{X \in \mathbb{C}^{T\times T} \mid X = \overline{X}^\top, \overline{v}^\top X v \geq 0 \text{ for all } v \in \mathbb{C}^T\}.$$

Proof. The positive semidefinite Hermitian $T \times T$ matrices are exactly the matrices of the form $UD\overline{U}^\top$ for $U \in U(T)$ and a real diagonal matrix D with nonnegative diagonal entries.

Consider any $X \in \mathbb{C}^{T\times T}$. It has a singular value decomposition $X = UDV$ for $U, V \in U(T)$ and a real diagonal matrix $D \in \mathbb{R}^{T\times T}$ with nonnegative diagonal entries, hence the positive semidefinite Hermitian matrix $\overline{V}^\top DV = \overline{(UV)}^\top X$ is in the orbit of X.

It remains to be shown that there cannot be two distinct positive semidefinite Hermitian matrices in one orbit.

To that end, consider $X \in \mathcal{H}_0^+(T)$ and $U \in U(T)$. It was seen in example 2.11 that all eigenvalues of unitary matrices are roots of unity. Two cases will be distinguished depending on the eigenvectors of $\overline{U}^\top$.

1) There is an eigenvector $v \in \mathbb{C}^T$ of $\overline{U}^\top$ with respect to an eigenvalue $\xi \neq 1$ and $v \notin \ker(X)$.

   Since X is positive semidefinite Hermitian and $v \notin \ker(X)$, the value $\overline{v}^\top Xv$ is real and strictly larger than zero. In addition ξ is a root of unity not equal to 1. In particular ξ is not a nonnegative real number. Therefore, the value

$$\overline{v}^\top UXv = \overline{\overline{U}^\top v}^\top Xv = \overline{\xi} \cdot \overline{v}^\top Xv$$

   is not a nonnegative real number and hence UX is not positive semidefinite Hermitian.

2) All eigenvectors $v \in \mathbb{C}^T$ of $\overline{U}^\top$ with $v \notin \ker(X)$ satisfy $\overline{U}^\top v = 1 \cdot v$.

   Suppose $v \in \mathbb{C}^T$ is an eigenvector of $\overline{U}^\top$ with respect to the eigenvalue ξ. If $\xi = 1$, then obviously $\overline{v}^\top UX = \overline{v}^\top X$ follows. If on the other hand $\xi \neq 1$, then $v \in \ker(X)$ and hence $\overline{v}^\top UX = \overline{\xi} \cdot \overline{v}^\top X = 0 = \overline{v}^\top X$ is obtained.

   In conclusion, $\overline{v}^\top UX = \overline{v}^\top X$ holds for all eigenvectors v of $\overline{U}^\top$. Since there is a basis of $\mathbb{C}^T$ consisting of eigenvectors of $\overline{U}^\top$ the equality $UX = X$ follows.

Altogether either $UX = X$ holds or UX is not positive semidefinite Hermitian. Therefore there cannot be two distinct positive semidefinite hermititan matrices in one orbit. □

The results of this section can now be combined and allow to conclude the following theorem.

(5.18) Theorem
Let $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ be a noncoherent STBC. There is a Λ-reduced code $C^* \subseteq \mathcal{H}_0^+(T)$ which is Λ-equivalent to $\mathcal{C}$. $\diamond$

Proof. By proposition 5.16 there is a square codebook $\mathcal{C}' \in \mathbb{C}^{T\times T}$ such that $\mathcal{C}$ is Λ-equivalent to $\mathcal{C}'$.

Furthermore, the preceding lemma 5.17 states that $\mathcal{H}_0^+(T)$ is a system of representatives of the orbits of $\mathbb{C}^{T\times T}$ under the left action of $U(T)$. Due to proposition 5.15 there is another codebook $\mathcal{C}'' \subseteq \mathcal{H}_0^+(T)$ which is Λ-equivalent to $\mathcal{C}'$ and hence also to $\mathcal{C}$.

It remains to be seen that $\mathcal{C}'$ is Λ-reduced. This follows directly from corollary 5.9.□

In conclusion, the problem of designing arbitrary noncoherent codebooks can be reduced to designing square codebooks containing only positive semidefinite matrices.

5.4 Distance functions for codewords

So far, efforts have been made to simplify the design problem by restricting the space from which codewords are drawn. The problem of finding good codebooks in such restricted spaces with respect to the distance function Δ, however, remains difficult. Therefore, the map Δ will be studied in the following and it will be investigated how it may be simplified.

To start with, a closer look is taken at two problematic properties of Δ.

1) The "distance" of a codeword X to itself is given by

$$\Delta(X,X) = \mathrm{Tr}(\Lambda_X^{-1}\Lambda_X) = \mathrm{Tr}(I_T) = T.$$

But for example

$$\Delta(I_T, 0) = \mathrm{Tr}\left(\frac{1}{2}I_T \cdot I_T\right) = \frac{1}{2}T$$

is strictly smaller than that.

More generally speaking this means that in a badly designed codebook there may be a codeword X_1 and a different codeword X_2 satisfying $\Delta(X_1, X_2) < \Delta(X_1, X_1)$.

As the distance function Δ was derived from the expected value of the decoding rule, that means that on average the codeword X_1 is more likely to be decoded erroneously as the codeword X_2 than to be decoded correctly. This is obviously not desirable.

2) The value of the map Δ does not scale with its arguments. Therefore, if a good codebook is designed with respect to a certain SNR, it will in general not perform well at a different SNR. This is, however, a property that is desirable in practice, since the SNR is not known beforehand in many cases or is even changing over the course of several transmissions.

   For example, unitary codes do have this property. If two unitary codewords X_1, X_2 are scaled to guarantee a certain SNR, the GLRT-distance does scale with them:
$$\left\| \rho X_1 \overline{\rho X_2}^\top \right\|_F = \rho^2 \left\| X_1 \overline{X_2}^\top \right\|_F .$$
   However, for arbitrary codewords X_1, X_2 and the distance function Δ this does not work in general:
$$\Delta(\rho X_1, \rho X_2) = \mathrm{Tr}((\rho^2 \overline{X_2}^\top X_2 + I_T)^{-1} (\rho^2 \overline{X_1}^\top X_1 + I_T)).$$

These drawbacks of the distance function Δ make it reasonable to look for simplifications. In order to present a systematic way of doing so, a formal notion of a distance function is introduced. Building on this, the aim will be to find other distance functions which are, in a reasonable sense, equivalent to Δ. However, it will not be possible to simplify Δ on the whole set $\mathbb{C}^{M \times T}$. Therefore, subsets $C \subseteq \mathbb{C}^{M \times T}$ will be considered. Such a subset $C \subseteq \mathbb{C}^{M \times T}$ from which the codewords are drawn is commonly referred to as the *coding space*.

(5.19) Definition
Let $C \subseteq \mathbb{C}^{M \times T}$ be a set of complex matrices. A *distance function* on C is an arbitrary map $d : C \times C \to \mathbb{R}$. ◇

Note that this definition of a distance function is very general. Axioms like nonnegativity, symmetry, the identity of indiscernibles and the triangle inequality as they are required for distances on metric spaces do not have to be satisfied here. There have been examples of distance functions which have been studied in earlier sections that will now illustrate the definition.

(5.20) Example (Distance functions)

1) One example of a distance function is given by the map Δ. It may be restricted from $\mathbb{C}^{M\times T}$ to any subset C:

$$\Delta : C \times C \to \mathbb{R},\ (X_1, X_2) \mapsto \mathrm{Tr}((\Lambda_{X_2})^{-1}\Lambda_{X_1}).$$

2) Another example was encountered when discussing unitary codes in section 4.2.

   Consider the set of scaled unitary matrices $C = \{X \in \mathbb{C}^{M\times T} \mid X\overline{X}^{\top} = \rho \cdot I_M\}$ for some $\rho \in \mathbb{R}_{>0}$ and a unitary codebook $\mathcal{C} \subseteq C$. It was argued that for distinct codewords $X_1, X_2 \in \mathcal{C}$ the GLRT-distance $\left\|X_1\overline{X_2}^{\top}\right\|_F$ needs to be minimized in order to minimize the probability of decoding errors.

   Putting this in terms of the notation which was introduced here, a corresponding distance function on C can be defined as

$$d_{\mathrm{GLRT}} : C \to C,\ (X_1, X_2) \mapsto -\left\|X_1\overline{X_2}^{\top}\right\|_F.$$

   The negative sign in the definition of the map d_{GLRT} was chosen such that the value of the distance function for two distinct codewords has to be as large as possible in order for the corresponding codebook to perform well. This makes the design of codebooks with respect to this distance function a more natural packing problem and is analogous to the distance function Δ. ◇

The aim now is to find suitable subsets $C \subseteq \mathbb{C}^{M\times T}$ on which a simpler distance function can be used which is equivalent to the original one. Hereby "equivalent" is supposed to mean that optimizing a codebook on that subset with respect to the one distance function yields the same result as optimizing it with respect to the other one. In particular, for two distance functions $d_1, d_2 : C \times C \to \mathbb{R}$ the following is required to hold for any $X_1, X_2, X_3, X_4 \in C$:

$$d_1(X_1, X_2) > d_1(X_3, X_4) \iff d_2(X_1, X_2) > d_2(X_3, X_4).$$

That is, d_1 and d_2 differ by a strictly monotonous transformation. This will be used to formalize the notion of equivalent distance functions. To that end, the notion of a strictly monotonously growing map on an arbitrary subset of $\mathbb{R}$ is clarified and necessary basic properties are stated.

(5.21) Definition and Lemma
Consider two nonempty subsets $D_1, D_2 \subseteq \mathbb{R}$ and a map $\varphi : D_1 \to D_2$.

The map φ is called *strictly monotonously growing* if for $x, y \in D_1$ the implication $x > y \implies \varphi(x) > \varphi(y)$ holds.

If φ is strictly monotonously growing the following assertions are true:

(i) φ is injective.

(ii) If φ is bijective, for $x, y \in D_1$ the equivalence $x > y \iff \varphi(x) > \varphi(y)$ holds.

(iii) If φ is bijective, φ^{-1} is also strictly monotonously growing.

(iv) If $D_3 \subset \mathbb{R}$ is another subset and $\psi : D_2 \to D_3$ is strictly monotonously growing, then $\psi \circ \varphi$ is strictly monotonously growing.

(v) For any nonempty subset $D \subseteq \mathbb{R}$, $\mathrm{Id}_D : D \to D, x \mapsto x$ is strictly monotonously growing. ◇

Proof. Statement (i) is obvious. Regarding (ii), consider $x, y \in D_1$ with $\varphi(x) > \varphi(y)$. Then clearly $x \neq y$. If $x < y$, then $\varphi(x) < \varphi(y)$ contradicting $\varphi(x) > \varphi(y)$. Therefore $x > y$ holds and hence the equivalence. Statement (iii) may be shown by using the implication " $\Longleftarrow$ " in case (ii) by setting $x = \varphi^{-1}(x')$ and $y = \varphi^{-1}(y')$ for $x', y' \in D_2$ with $x' > y'$. Statement (iv) is easily checked by successively applying the definition of a strictly monotonously growing map for φ and then for ψ. Lastly, (v) is easily checked by means of the definition. □

Having established this notion, the formal definition of equivalent distance functions can be given.

(5.22) Definition (Equivalent distance functions)
Let $C \subseteq \mathbb{C}^{M \times T}$ be a set of complex matrices.

Denote by $d_1 : C \times C \to \mathbb{R}$ and $d_2 : C \times C \to \mathbb{R}$ two distance functions.

These functions d_1 and d_2 are called equivalent if and only if a strictly monotonously growing map $\varphi : d_2(C, C) \to d_1(C, C)$ exists such that $d_1 = \varphi \circ d_2$ holds. ◇

Using the elementary statements that have been made in lemma 5.21 about strictly monotonously growing maps, it can be verified that it is in fact appropriate to use the term equivalence.

(5.23) Lemma
Equivalence of distance functions defines an equivalence relation on the set of all distance functions. ◇

Proof. By means of lemma 5.21, the axioms of an equivalence relation can be verified quickly.

To start with, reflexivity becomes obvious by setting $\varphi = \mathrm{Id}_{d(C,C)}$.

Concerning symmetry, note that any φ meeting the conditions from the definition has to be bijective: Injectivity is clear by lemma 5.21 part (i), whereas surjectivity follows from the fact that any element $d_1(X_1, X_2) \in d_1(C, C)$ does have the preimage $d_2(X_1, X_2)$ under φ. Therefore, by lemma 5.21 part (iii), the map φ^{-1} is strictly monotonously growing and it also satisfies $\varphi^{-1} \circ d_1 = d_2$

In order to prove transitivity one needs to consider the composition of two strictly monotonously growing maps. This is strictly monotonously growing as well by lemma 5.21 part (iv). □

It has been described in section 4.3 that the construction of good unitary codes can be regarded as packing elements on the Grassmann manifold with respect to the chordal distance. We can now motivate this differently by interpreting the chordal distance as a distance function on the set of rectangular unitary matrices and by then putting it in relation with the design criterion 5.5.

(5.24) Example (Equivalent distance functions for unitary codes)
Consider the set of scaled $M \times T$ unitary matrices

$$C_u := \{X \in \mathbb{C}^{M\times T} \mid X\overline{X}^\top = \rho I_M\}$$

for $\rho \in \mathbb{R}_{>0}$ representing the SNR of a unitary codebook drawn from this set. Furthermore, consider the distance function

$$\Delta : C_u \times C_u \to \mathbb{R}, (X_1, X_2) \mapsto \mathrm{Tr}(\Lambda_{X_2}^{-1}\Lambda_{X_1})$$

as well as the the distance function induced by the chordal distance on the Grassmann manifold

$$d_c(X_1, X_2) = \sqrt{M - \left\| X_1\overline{X_2}^\top \right\|_F^2}.$$

In section 4.2 the chordal distance was derived as an important measure for designing unitary codes. It will now be put in relation to the distance function Δ which was directly derived from the general maximum likelihood decoding rule. Particularly it will be shown that these two distance functions are equivalent on C_u by explicitly computing the strictly monotonic transformation required in the definition of equivalency.

By lemma 4.10 the equation $\Lambda_X^{-1} = I_T - \frac{1}{1+\rho}\overline{X}^\top X$ holds for $X \in C_u$ and the definition of the chordal distance between two codewords $X_1, X_2 \in C_u$ may be rearranged as $\left\| X_1 \overline{X_2}^\top \right\|_F^2 = M - d_c(X_1, X_2)^2$. Utilizing these equations for $X_1, X_2 \in C_u$ yields

$$\begin{aligned}
\Delta(X_1, X_2) &= \operatorname{Tr}\left((I_T - \frac{1}{1+\rho}\overline{X_2}^\top X_2)(I_T + \overline{X_1}^\top X_1) \right) \\
&= \operatorname{Tr}\left(I_T - \frac{1}{1+\rho}\overline{X_2}^\top X_2 + \overline{X_1}^\top X_1 - \frac{1}{1+\rho}\overline{X_2}^\top X_2 \overline{X_1}^\top X_1 \right) \\
&= T - \frac{1}{1+\rho}\|X_2\|_F^2 + \|X_1\|_F^2 - \frac{1}{1+\rho}\left\| X_1 \overline{X_2}^\top \right\|_F^2 \\
&= T - \frac{1}{1+\rho}\rho M + \rho M - \frac{1}{1+\rho}\left\| X_1 \overline{X_2}^\top \right\|_F^2 \\
&= T - \frac{1}{1+\rho}\rho M + \rho M - \frac{1}{1+\rho}(M - d_c(X_1, X_2)^2) \\
&= T + (\rho - 1)M + \frac{1}{1+\rho} d_c(X_1, X_2)^2.
\end{aligned}$$

Note that $d_c(C_u, C_u) \subseteq \mathbb{R}_{\geq 0}$ to conclude that the map

$$\varphi : d_c(C_u, C_u) \to \Delta(C_u, C_u),\ x \mapsto T + (\rho - 1)M + \frac{1}{1+\rho}x^2$$

is a monotonously growing map satisfying $\Delta = \varphi \circ d_c$.

It also becomes clear that the distance function

$$d_{\mathrm{GLRT}} : C_u \times C_u \to \mathbb{R}, (X_1, X_2) \mapsto -\left\| \overline{X_1}^\top X_2 \right\|_F$$

is equivalent to both the distance functions above. In particular, the map

$$\psi : d_c(C_u, C_u) \to d_{\mathrm{GLRT}}(C_u, C_u), x \mapsto -\sqrt{M - x^2}$$

is strictly monotonously growing and satisfies $d_{\mathrm{GLRT}} = \psi \circ d_c$. ◇

This example shows how the distance function Δ may be simplified on the coding space $C_u = \{X \in \mathbb{C}^{M \times T} \mid X\overline{X}^\top = \rho I_M\} \subseteq \mathbb{C}^{M \times T}$. Similar courses of action will be taken later on to find suitable distance functions on different coding spaces.

5.5 Conclusions about the general design problem

This section concludes the simplifications of the original design criteria. These were stated in proposition 5.5 and deduced directly from the expected value of the probability of error. In the following, the notions of Λ-reduced and Λ-equivalent codes were introduced. By their means it was possible to significantly restrict the set of potential codewords which should be used to construct noncoherent STBC. Particularly, in theorem 5.18 it was shown that it is sufficient to consider (square) positive semidefinite hermitian matrices as codewords. Thereafter, the notion of a distance function was formally introduced and it was studied how such distance functions can be simplified.

These results will be used to construct noncoherent STBC in the following sections. In order to guide this process, the design criteria are reformulated accordingly.

(5.25) Design Criteria for noncoherent STBC (2)
Consider a coding space $C \subseteq \mathcal{H}_0^+(T)$, the restricted distance function $\Delta : C \times C \to \mathbb{R}$ and a distance function $d : C \times C \to \mathbb{R}$ which is equivalent to Δ.

Let $\mathcal{C} \subseteq C$ be a finite set of complex matrices. For $\mathcal{C}$ to be used as a codebook for communicating over a channel with an SNR of $\rho \in \mathbb{R}_{>0}$ we stipulate that the following conditions shall be satisfied:

1) $\frac{1}{|\mathcal{C}|} \sum_{X \in \mathcal{C}} \|X\|_F = \rho$,

2) $\det(\Lambda_X)$ is constant for all $X \in \mathcal{C}$,

3) $\min_{X_1, X_2 \in \mathcal{C}, X_1 \neq X_2} d(X_1, X_2)$ is maximized. ◇

In general, the optimization problem resulting from these criteria is still hard to solve and there is no apparent algebraic structure to exploit in order to obtain good codes. The solution used in the upcoming section to systematically construct good codebooks will be to restrict the coding space C to a specific subset $C \subseteq \mathcal{H}_0^+(T)$. This subset will specifically be chosen such that there is a sufficiently simple distance function d on C which satisfies the above criteria. Namely, a suitable set C will be defined such that d can be chosen to be the GLRT-distance which was first introduced in section 4.1 and formally defined in example 5.20.

§ 6 Noncoherent STBC for the GLRT-receiver

In this chapter a new class of noncoherent STBC will be introduced and studied. This will be done by using the theory developed in the preceding part of this work. Particularly, a class of codes will be derived for which it is optimal to use the GLRT criterion (4.12) as the design criterion and for which maximum likelihood decoding is performed by the GLRT decoder. Within the class of these codes, optimal codebooks will be characterized and some example constructions will be presented.

To begin with, the aim is to find a set of complex matrices C for which it can be shown that the GLRT distance is Λ-equivalent to the distance function Δ. This is, for example, true for unitary codes, shown in example 5.24. The key to the proof of the equivalence of d_{GLRT} and Δ for unitary codes was the linear dependence of Λ_X^{-1} on Λ_X. In the following section, a class of codewords which also yield such a dependence is presented.

It was shown in the previous chapter and included in the design criteria 5.25 that it suffices to consider positive semidefinite Hermitian matrices as codewords. In the first section of this chapter, the codewords will, however, not be required to be positive semidefinite Hermitian. In fact, even nonsquare matrices will be considered. In doing so, the results of this section will hold for a greater class of codewords without complicating the required arguments.

6.1 A class of codebooks for the GLRT-receiver

This section introduces a class of codebooks $\mathcal{C} \subseteq \mathbb{C}^{M\times T}$ for $M \leq T$ for which the matrices Λ_X for all codewords $X \in \mathcal{C}$ have exactly two fixed eigenvalues with fixed multiplicities. Formally, throughout this section the coding space

$$C_{\mathcal{S}} := \{X \in \mathbb{C}^{M\times T} \mid \operatorname{spec}(\Lambda_X) = \mathcal{S}\}$$

will be considered, where $\mathcal{S} := \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ denotes a fixed spectrum which contains two distinct positive real eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ of respective multiplicities $\mu_1, \mu_2 \in \mathbb{N}$. Note that all matrices $\Lambda_X = I_T + \overline{X}^{\top} X$ are positive definite Hermitian and hence only possess positive eigenvalues.

Restricting to codebooks $\mathcal{C} \subseteq C_{\mathcal{S}}$ yields the following simple assertion which makes sure that part 2) of the design criteria 5.25 will always be satisfied for these codebooks.

(6.1) Remark
For $X \in C_{\mathcal{S}}$ the equation $\det(\Lambda_X) = \lambda_1^{\mu_1} \cdot \lambda_2^{\mu_2}$ holds.

In particular the determinant of Λ_X is fixed for all $X \in C_{\mathcal{S}}$. ◇

As a first step in the analysis of codes that are built from elements of $C_{\mathcal{S}}$, it shall be clarified what kind of codewords can actually occur. The following simple connection between the singular values of the matrices $X \in \mathbb{C}^{M \times T}$ and the eigenvalues of the respective matrices Λ_X provides a first step toward this end.

(6.2) Lemma
For $M \leq T$ consider $X \in \mathbb{C}^{M \times T}$ with singular values $\sigma_1, \ldots, \sigma_M$. The eigenvalues of Λ_X are given by $1 + \sigma_1^2, \ldots, 1 + \sigma_M^2$ and by additional $(T - M)$-times the value 1. ◇

Proof. Suppose the singular value decomposition of $X \in \mathbb{C}^{M \times T}$ is given by $X = UDV$ for $U \in U(M)$, $V \in U(T)$ and a diagonal matrix $D \in \mathbb{R}^{M \times T}$. In these terms, the matrix Λ_X may be written as

$$\Lambda_X = I_T + \overline{X}^\top X = \overline{V}^\top (I_T + \overline{D}^\top D) V.$$

In order to further evaluate the right hand side of that equation, denote the singular values of X by $\sigma_1, \ldots, \sigma_M$ to obtain

$$I_T + \overline{D}^\top D = \begin{pmatrix} 1 + \sigma_1^2 & & & \\ & \ddots & & \\ & & 1 + \sigma_M^2 & \\ & & & I_{T-M} \end{pmatrix}.$$

This yields the assertion of the lemma. □

By means of this lemma, the singular values of any $X \in C_{\mathcal{S}}$ can be explicitly described dependent on the eigenvalues of the respective matrices Λ_X.

(6.3) Corollary
Consider $X \in C_{\mathcal{S}}$. In order to describe the singular values of X three cases are distinguished:

(i) If $M < T$ and $\mu_1 = M$, then one of the eigenvalues of Λ_X must be 1. Suppose without loss of generality $\lambda_2 = 1$. The single singular value of X is of multiplicity M and given by $\sigma_1 = \sqrt{\lambda_1 - 1}$.

(ii) If $M < T$ and $\mu_1 < M$, then again $\lambda_2 = 1$ can be assumed. The two singular values of X are then given by $\sigma_1 = \sqrt{\lambda_1 - 1}$ of multiplicity μ_1 and $\sigma_2 = \sqrt{\lambda_2 - 1} = 0$ of multiplicity $M - \mu_1$.

(iii) If $M = T$, then the two singular values of X are given by $\sigma_1 = \sqrt{\lambda_1 - 1}$ and $\sigma_2 = \sqrt{\lambda_2 - 1}$ of respective multiplicities μ_1 and μ_2. ◇

As a direct result of the preceding corollary it can be concluded that the Frobenius norm of all potential codewords drawn from the set $C_{\mathcal{S}}$ will be equal.

(6.4) Corollary
All $X \in C_{\mathcal{S}}$ satisfy

$$\|X\|_F = \sqrt{\mu_1(\lambda_1 - 1) + (M - \mu_1)(\lambda_2 - 1)}.$$

In particular the Frobenius norm is fixed for all $X \in C_{\mathcal{S}}$. ◇

Proof. Consider $X \in C_{\mathcal{S}}$. By lemma 2.12 the Frobenius norm of the matrix X is given by $\|X\|_F = \sqrt{\sum_{i=1}^{M} \sigma_i^2}$ for the singular values $\sigma_1, \ldots, \sigma_M$ of X.

Therefore, the assertion may be checked by going through the three cases for the singular values of X from the preceding corollary.

Note that in the case (iii) the equality $\mu_2 = M - \mu_1$ holds. Hence, for this case the result is obtained by simply plugging in the determined singular values into the equation for the Frobenius norm given above.

For cases (i) and (ii) note that $\lambda_2 = 1$ holds and hence the assertion of this corollary reduces to $\|X\|_F = \sqrt{\mu_1(\lambda_1 - 1)}$. This is again apparent by plugging in the singular values into the above equation. □

Corollary 6.3 completely characterizes the singular values of the elements of $C_{\mathcal{S}}$ for all possible cases. Conversely, the singular values of a codeword X uniquely determine the eigenvalues of $\Lambda_X = I_T + \overline{X}^\top X$ and so all matrices $X \in \mathbb{C}^{M \times T}$ with these singular values lie in $C_{\mathcal{S}}$. Making use of this observation, the coding space $C_{\mathcal{S}}$ can now be described in more detail.

(6.5) Corollary
The coding space $C_{\mathcal{S}}$ is described by one of the following three cases:

1) $M < T$ and $\mu_1 = M$.

The codewords have exactly one singular value $\sigma := \sqrt{\lambda_1 - 1}$ of multiplicity M. Therefore, any element $X \in C_{\mathcal{S}}$ has a singular value decomposition

$$X = U(\sigma \cdot I_{M,T})V \quad \text{for } U \in U(M),\ V \in U(T).$$

This is equivalent to $X\overline{X}^{\top} = \sigma^2 I_M$ and hence the set $C_{\mathcal{S}}$ may be described as

$$C_{\mathcal{S}} = \{X \in \mathbb{C}^{M \times T} \mid X\overline{X}^{\top} = \sigma^2 I_M\}.$$

This does exactly correspond to the case of unitary codes. It has been studied in section 4.2 and again in example 5.24.

2) $M < T$ and $\mu_1 < M$.

The codewords have exactly two distinct singular values $\sigma_1 = \sqrt{\lambda_1 - 1}$ of multiplicity μ_1 and $\sigma_2 = 0$ of multiplicity $M - \mu_1$.

By means of lemma 6.2 it is easy to check that such a code is always Λ-equivalent to a code $\mathcal{C}' \subseteq \mathbb{C}^{\mu_1 \times T}$ which corresponds to the first case. The code $\mathcal{C}'$ uses strictly fewer transmit antennas than $\mathcal{C}$ and this case will therefore not be of importance.

3) $M = T$.

The codewords have two distinct singular values $\sigma_1 = \sqrt{\lambda_1 - 1}$ and $\sigma_2 = \sqrt{\lambda_2 - 1}$ of respective multiplicities μ_1 and μ_2. The set $C_{\mathcal{S}}$ may hence be given as

$$C_{\mathcal{S}} = \left\{ X \in \mathbb{C}^{T \times T} \,\middle|\, X = U \begin{pmatrix} \sigma_1 I_{\mu_1} & \\ & \sigma_2 I_{\mu_2} \end{pmatrix} V \text{ for } U, V \in U(T) \right\}.$$

◇

It was shown earlier that it suffices to study codebooks which contain only square matrices. Therefore, the upcoming sections will focus on the third case. However, for the course of this section no further restrictions will be imposed on the elements of C and hence all three cases are treated simultaneously. Keeping that in mind, the next aim is to find a distance function which is equivalent to Δ on the coding space $C_{\mathcal{S}}$. To that end, some technical lemmas are required.

It is easily checked by means of linear algebra that diagonalizable matrices with exactly two distinct eigenvalues have a minimal polynomial of degree two:

(6.6) Lemma
Let $A \in \mathbb{C}^{n \times n}$ for $n \in \mathbb{N}$ be diagonalizable with exactly two eigenvalues $\lambda_1, \lambda_2 \in \mathbb{C}$. The following equality holds:

$$(A - \lambda_1 I_n)(A - \lambda_2 I_n) = A^2 - (\lambda_1 + \lambda_2)A + \lambda_1 \lambda_2 I_n = 0.$$

◇

This fact will be very helpful to deduce distance functions which are equivalent to Δ on $C_{\mathcal{S}}$. To begin with, it is helpful for proving the following lemma.

(6.7) Lemma
Let $A \in \mathbb{C}^{n \times n}$ for $n \in \mathbb{N}$ be a diagonalizable matrix with exactly two eigenvalues $\lambda_1, \lambda_2 \in \mathbb{C} \setminus \{-1\}$.

Then $A + I_n$ is invertible with $(A + I_n)^{-1} = \frac{-1}{(\lambda_1+1)(\lambda_2+1)}(A - (\lambda_1 + \lambda_2 + 1)I_n)$. ◇

Proof. With the notation as in the lemma the following equations hold:

$$\begin{aligned}(A + I_n)(A - (\lambda_1 + \lambda_2 + 1)I_n) &= A^2 - (\lambda_1 + \lambda_2 + 1)A + A - (\lambda_1 + \lambda_2 + 1)I_n \\ &= A^2 - (\lambda_1 + \lambda_2)A - (\lambda_1 + \lambda_2 + 1)I_n \\ &\overset{(6.6)}{=} -\lambda_1 \lambda_2 I_n - (\lambda_1 + \lambda_2 + 1)I_n \\ &= -(\lambda_1 \lambda_2 + \lambda_1 + \lambda_2 + 1)I_n \\ &= -(\lambda_1 + 1)(\lambda_2 + 1)I_n.\end{aligned}$$

□

In order to apply these facts to matrices of the form $\Lambda_X = (I_T + \overline{X}^\top X)$, it is sufficient to restrict the matrix A from the above lemma to be positive semidefinite Hermitian. This implies $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ and hence the following simple corollary is obtained for this case:

(6.8) Corollary
Let A be a positive definite Hermitian matrix with exactly two different eigenvalues.

Then $(A + I_n)^{-1} = -aA + bI_n$ holds for some real $a, b > 0$ that depend only on the eigenvalues of A. ◇

This result can now be used to find simpler distance functions on C which are equivalent to Δ.

(6.9) Proposition
The following distance functions are equivalent on $C_{\mathcal{S}}$:

(1) $\Delta : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto \Delta(X_1, X_2)$,

(2) $\Delta' : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto \Delta(X_2, X_1)$,

(3) $d_{\text{GLRT}} : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto - \left\| X_1 \overline{X_2}^\top \right\|_F$,

(4) $d'_{\text{GLRT}} : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto - \left\| X_2 \overline{X_1}^\top \right\|_F$.

If the elements of $C_{\mathcal{S}}$ are additionally assumed to be square and invertible, that is $M = T$ and $\lambda_1, \lambda_2 > 0$, the following distance functions are also equivalent to the above:

(5) $d_5 : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto \left\| X_2 X_1^{-1} \right\|_F$,

(6) $d_6 : C_{\mathcal{S}} \times C_{\mathcal{S}} \to \mathbb{R}$, $(X_1, X_2) \mapsto \left\| X_1 X_2^{-1} \right\|_F$. ◇

Proof. For any $X_1, X_2 \in \mathbb{C}^{M \times T}$ it is easily seen that

$$\left\| X_1 \overline{X_2}^\top \right\|_F = \left\| \overline{X_1 \overline{X_2}^\top}^\top \right\|_F = \left\| X_2 \overline{X_1}^\top \right\|_F,$$

hence $d'_{\text{GLRT}} = d_{\text{GLRT}}$ holds and in particular d_{GLRT} and d'_{GLRT} are equivalent. Due to the similarities of the given maps it is now sufficient to prove $\Delta \sim d_{\text{GLRT}} \sim d_5$. The remaining equivalences $\Delta' \sim d'_{\text{GLRT}} \sim d_6$ may be shown analogously by renaming the variables.

To start with, the equivalence $\Delta \sim d_{\text{GLRT}}$ is proven. First, note that for any $X \in C_{\mathcal{S}}$ the matrix $\Lambda_X = I_T + \overline{X}^\top X$ is Hermitian and positive definite and by definition of C possesses exactly two eigenvalues. Therefore, for $X_2 \in C_{\mathcal{S}}$ corollary (6.8) can be applied to Λ_{X_2} which yields $\Lambda_{X_2}^{-1} = (I_T + \overline{X_2}^\top X_2)^{-1} = -a\overline{X_2}^\top X_2 + bI_T$ for real numbers $a, b > 0$. This may be applied to the definition of Δ for $X_1, X_2 \in C_{\mathcal{S}}$:

$$\begin{aligned} \Delta(X_1, X_2) &= \operatorname{Tr}((\Lambda_{X_2})^{-1} \Lambda_{X_1}) \\ &= \operatorname{Tr}((bI_T - a\overline{X_2}^\top X_2)(I_T + \overline{X_1}^\top X_1)) \\ &= \operatorname{Tr}(bI_T + b\overline{X_1}^\top X_1 - a\overline{X_2}^\top X_2 - a\overline{X_2}^\top X_2 \overline{X_1}^\top X_1) \\ &= b \cdot T + b \left\| X_1 \right\|_F^2 - a \left\| X_2 \right\|_F^2 - a \left\| X_1 \overline{X_2}^\top \right\|_F^2. \end{aligned}$$

By the corollaries 6.4 and 6.8 the values a, b and $\|X_1\|_F$, $\|X_2\|_F$ are fixed via the eigenvalues of Λ_{X_1} and Λ_{X_2} and hence are identical for all possible choices of $X_1, X_2 \in C_{\mathcal{S}}$.

Therefore, the last expression is only dependent on $\left\| X_1 \overline{X_2}^\top \right\|_F$. Denote $c := \|X\|_F$ for $X \in C_{\mathcal{S}}$. Because of the above computation, of $d_{\text{GLRT}}(C_{\mathcal{S}}, C_{\mathcal{S}}) \subseteq \mathbb{R}_{<0}$ and of $a > 0$, the map

$$\varphi : d_{\text{GLRT}}(C_{\mathcal{S}}, C_{\mathcal{S}}) \to \Delta(C_{\mathcal{S}}, C_{\mathcal{S}}),\ x \mapsto bT + bc^2 - ac^2 - ax^2$$

is well defined, strictly monotonously growing and it also satisfies $\Delta = \varphi \circ d_{\text{GLRT}}$. So, the equivalence of Δ and d_{GLRT} follows.

Now suppose all elements of $C_{\mathcal{S}}$ are square and invertible. Under this assumption any $X \in C_{\mathcal{S}}$ has exactly two distinct strictly positive singular values and hence the matrix $\overline{X}^\top X$ is positive definite Hermitian. Its eigenvalues are uniquely determined by the singular values of X and hence by the eigenvalues of Λ_X. In particular, the eigenvalues of $\overline{X}^\top X$ are the same for all $X \in C_{\mathcal{S}}$.

To show the equivalence $\Delta \sim d_5$, lemma (6.6) is applied to $\overline{X_1}^\top X_1$ for $X_1 \in C_{\mathcal{S}}$ which yields the existence of real numbers $d, e > 0$ such that $(\overline{X_1}^\top X_1)^2 = d\overline{X_1}^\top X_1 - eI_T$ holds. These numbers depend only on the eigenvalues of $\overline{X_1}^\top X_1$ and, by the comment above, they are therefore independent of the element $X_1 \in \mathcal{C}_{\mathcal{S}}$. The equation is multiplied by $X_1^{-1}\overline{X_1^{-1}}^\top$ to obtain $\overline{X_1}^\top X_1 = dI_T - eX_1^{-1}\overline{X_1^{-1}}^\top$. Analogously to the above computation for $X_1, X_2 \in C_{\mathcal{S}}$, this yields

$$\begin{aligned}\Delta(X_1, X_2) &= \operatorname{Tr}(bI_T + b\overline{X_1}^\top X_1 - a\overline{X_2}^\top X_2 - a\overline{X_2}^\top X_2 \overline{X_1}^\top X_1) \\ &= \operatorname{Tr}(bI_T + b\overline{X_1}^\top X_1 - a\overline{X_2}^\top X_2 - a\overline{X_2}^\top X_2(dI_T - eX_1^{-1}\overline{X_1^{-1}}^\top)) \\ &= b \cdot T + b\,\|X_1\|_F^2 - a(1+d)\,\|X_2\|_F^2 + ae\left\| X_2 X_1^{-1} \right\|_F^2.\end{aligned}$$

Again, in the last formula all the values but $\left\| X_2 X_1^{-1} \right\|_F$ are independent of the choices of X_1 and X_2. Since the inequality $ae > 0$ holds, the well defined map

$$\psi : d_5(C_{\mathcal{S}}, C_{\mathcal{S}}) \to \Delta(C_{\mathcal{S}}, C_{\mathcal{S}}),\ x \mapsto b \cdot T + bc - a(1+d)c + aex^2$$

is strictly monotonously growing and it satisfies $\Delta = \psi \circ d_5$. □

An analogous argument can also be applied to simplify the maximum likelihood decoding problem.

(6.10) Proposition
The maximum likelihood decoding problem for codebooks $\mathcal{C} \subseteq C_{\mathcal{S}}$ and a received signal $Y \in \mathbb{C}^{N \times T}$ is to find the element

$$\operatorname{argmax}_{X \in \mathcal{C}} \left\| X\overline{Y}^\top \right\|_F.$$

Proof. As in the proof of the previous proposition, for $X \in C_{\mathcal{S}}$, real numbers $a, b > 0$ independent of X are obtained such that $\Lambda_X^{-1} = bI_T - a\overline{X}^\top X$ holds. So, for a received message Y the following equation holds:

$$\operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^\top Y) = \operatorname{Tr}((bI_T - a\overline{X}^\top X)\overline{Y}^\top Y) = b\,\|Y\|_F^2 - a\left\|X\overline{Y}^\top\right\|_F^2.$$

For a given received message Y the right hand side depends only on $-\left\|X\overline{Y}^\top\right\|_F^2$ and in conclusion the maximum likelihood decoder decision from corollary 5.2 can be rewritten as

$$\operatorname{argmax}_{X\in\mathcal{C}} P(Y|X) = \operatorname{argmin}_{X\in\mathcal{C}} \operatorname{Tr}(\Lambda_X^{-1}\overline{Y}^\top Y) = \operatorname{argmax}_{X\in\mathcal{C}} \left\|X\overline{Y}^\top\right\|_F. \qquad \square$$

Altogether the results of this section extend the class of codes which may be optimized with respect to the GLRT-distance and decoded by the GLRT-criterion from unitary codes to codebooks built from the set $C_{\mathcal{S}}$.

In particular this includes sets of square matrices which opens the possibility to design square noncoherent STBC that may be decoded by the GLRT decoder. The construction and performance of such codes will be discussed in the following.

That process will be guided by the design criterion 5.5 which can be further simplified for this specific case.

(6.11) Design Criteria for noncoherent STBC (3)
Let $\mathcal{C} \subseteq C_{\mathcal{S}}$ be a finite set of complex matrices. For $\mathcal{C}$ to be used as a codebook for communicating over a channel with an SNR of $\rho \in \mathbb{R}_{>0}$ the following conditions shall be satisfied:

1) $\|X\|_F = \rho$ for $X \in \mathcal{C}_{\mathcal{S}}$.

2) $\min_{X_1,X_2\in\mathcal{C}_{\mathcal{S}}, X_1\neq X_2} d(X_1, X_2)$ is maximized, where d is either the GLRT-distance d_{GLRT} or one of the other equivalent distance functions from proposition 6.9. ◇

6.2 Parametrization of codewords via the exponential map

The previous section dealt with distance functions on a coding space $C_{\mathcal{S}}$ for a spectrum $\mathcal{S}$ containing exactly two eigenvalues. In particular, it was found that on these coding spaces the distance function Δ is equivalent to the GLRT-distance. By theorem 5.18 any codebook is Λ-equivalent to a codebook consisting of square positive semidefinite Hermitian matrices. Therefore, the coding spaces that will be considered from here on will consist of positive semidefinite Hermitian matrices which possess a fixed spectrum of exactly two distinct eigenvalues. The aim of this section is to give a parametrization of these coding spaces and to determine how that parametrization can be used to construct codebooks with large minimal distances. At the end of the section an example codebook is constructed by means of this parametrization and corresponding simulation results are presented.

To begin with, a notation for the set of Hermitian matrices with a fixed spectrum is introduced.

(6.12) Definition
For a spectrum $\mathcal{S}$ and $T \in \mathbb{N}$ such that $\mathcal{H}_{\mathcal{S}} \subseteq \mathcal{H}(T)$, we define the set of Hermitian matrices with that spectrum:

$$\mathcal{H}_{\mathcal{S}} := \{X \in \mathbb{C}^{T \times T} \mid X = \overline{X}^{\top}, \operatorname{spec}(X) = \mathcal{S}\}.$$

◇

The requirement that a codeword with two eigenvalues is positive semidefinite implies that both eigenvalues are nonnegative. In the following, only the slightly more special case of positive definite matrices will be considered. This excludes the case that one of the eigenvalues of the codewords is zero, which will be revisited later. Making this restriction allows for the parametrization of codewords via the exponential map introduced in section 2.3.

(6.13) Lemma
Consider the spectra $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ and $\mathcal{S} = \{(\exp(\lambda_1), \mu_1), (\exp(\lambda_2), \mu_2)\}$ for $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mu_1, \mu_2 \in \mathbb{N}$.

The restricted exponential map $\exp : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}}$ is bijective. ◇

Proof. Any $X \in \mathcal{H}_{\mathcal{S}_0}$ can be written as $UD\overline{U}^{\top}$ for $U \in U(T)$ and a real diagonal matrix D. By lemma 2.26 the equality $\exp(UD\overline{U}^{\top}) = U \exp(D)\overline{U}^{\top}$ holds. Note that the exponential map acts on the diagonal entries of D as the real exponential map. Therefore, an inverse map is given by $UD\overline{U}^{\top} \mapsto U \ln(D)\overline{U}^{\top}$ where ln acts on the diagonal entries of D as the real natural logarithm. □

Using this parametrization, the design criteria can be reformulated in terms of Hermitian matrices from $\mathcal{H}_{\mathcal{S}_0}$.

(6.14) Design Criteria for noncoherent STBC (4)
Consider a spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ for two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mu_1, \mu_2 \in \mathbb{N}$ and let $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ be a finite set of complex matrices.

For $\mathcal{C} := \{\exp(X) \mid X \in \mathcal{C}_0\}$ to be used as a codebook for communicating over a channel with an SNR of $\rho \in \mathbb{R}_{>0}$ the following conditions shall be satisfied:

1) $\|\exp(X)\|_F = \rho$ for $X \in \mathcal{C}_0$.

2) $\max_{X_1, X_2 \in \mathcal{C}_0, X_1 \neq X_2} \|\exp(X_1)\exp(X_2)\|_F$ is minimized. ◇

This point of view will yield a helpful connection between the GLRT-distance of two codewords and the euclidean distance in a certain real vector space. Before this connection is developed, the following technical lemma is required.

(6.15) Lemma
Consider a diagonalizable matrix $A \in \mathbb{C}^{T \times T}$ with $\operatorname{spec}(A) = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ for two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{C}$ and $\mu_1, \mu_2 \in \mathbb{N}$. For any $n \in \mathbb{N}_0$ the following equation holds:

$$A^n = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} A + \frac{\lambda_1 \lambda_2^n - \lambda_2 \lambda_1^n}{\lambda_1 - \lambda_2} I_T.$$

Proof. Since A is diagonalizable, there is a basis of $\mathbb{C}^T$ consisting of eigenvectors of A. Therefore, it suffices to show that left multiplying both sides of the equation with an eigenvector of A yields the same result.

Now suppose that $v \in \mathbb{C}^T$ is an eigenvector of A with respect to the eigenvalue $\lambda \in \{\lambda_1, \lambda_2\}$. Concerning the left hand side, it is easily seen that $A^n v = \lambda^n v$ holds for any $n \in \mathbb{N}_0$.

Evaluating the right hand side for $n \in \mathbb{N}_0$ yields

$$\begin{aligned}
\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} A v + \frac{\lambda_1 \lambda_2^n - \lambda_2 \lambda_1^n}{\lambda_1 - \lambda_2} I_T v &= \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \lambda v + \frac{\lambda_1 \lambda_2^n - \lambda_2 \lambda_1^n}{\lambda_1 - \lambda_2} v \\
&= \frac{\lambda \lambda_1^n - \lambda \lambda_2^n}{\lambda_1 - \lambda_2} v + \frac{\lambda_1 \lambda_2^n - \lambda_2 \lambda_1^n}{\lambda_1 - \lambda_2} v \\
&= \frac{(\lambda - \lambda_2)\lambda_1^n + (\lambda_1 - \lambda)\lambda_2^n}{\lambda_1 - \lambda_2} v \\
&= \lambda^n v.
\end{aligned}$$

The last equality is easily verified by distinguishing the two possible cases $\lambda = \lambda_1$ and $\lambda = \lambda_2$. □

The lemma above may be used to connect the GLRT-distance of two codewords parametrized by the exponential map to another map, which can be identified with an inner product on a Euclidean vector space.

(6.16) Lemma
Consider a spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ for two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mu_1, \mu_2 \in \mathbb{N}$.

There is a strictly monotonously growing map $\varphi : \mathbb{R} \to \mathbb{R}$, $x \mapsto A^2 x + 2ABC + B^2 T$ satisfying

$$\|\exp(X_1)\exp(X_2)\|_F = \varphi(\operatorname{Tr}(X_1 X_2))$$

for all $X_1, X_2 \in \mathcal{H}_{\mathcal{S}_0}$ and $A, B, C \in \mathbb{R}$ given by

$$\begin{aligned} A &= \frac{e^{2\lambda_1} - e^{2\lambda_2}}{\lambda_1 - \lambda_2}, \\ B &= \frac{\lambda_1 e^{2\lambda_2} - \lambda_2 e^{2\lambda_1}}{\lambda_1 - \lambda_2}, \\ C &= \mu_1 \lambda_1 + \mu_2 \lambda_2. \end{aligned} \qquad \diamond$$

Proof. Consider two matrices $X_1, X_2 \in \mathcal{H}_{\mathcal{S}_0}$. Corollary 2.27 implies that the matrices $\exp(X_1)$ and $\exp(X_2)$ are also Hermitian. Using this fact and basic properties of the exponential map (lemma 2.26), the following equation is obtained:

$$\begin{aligned} \|\exp(X_1)\exp(X_2)\|_F &= \operatorname{Tr}\left(\overline{\exp(X_1)\exp(X_2)}^{\top} \exp(X_1)\exp(X_2)\right) \\ &= \operatorname{Tr}(\exp(X_1)^2 \exp(X_2)^2) \\ &= \operatorname{Tr}(\exp(2X_1)\exp(2X_2)). \end{aligned}$$

The right hand side of this equation can be further simplified by means of corollary 2.30, which yields

$$\exp(2X_1)\exp(2X_2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{i!j!}(2X_1)^i (2X_2)^j.$$

All series on the right hand side are convergent and therefore corollary 2.32 may be applied to obtain

$$\operatorname{Tr}(\exp(2X_1)\exp(2X_2)) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{i!j!} \operatorname{Tr}((2X_1)^i (2X_2)^j).$$

To simplify this further, the preceding lemma 6.15 may be applied to the matrices X_1 and X_2. Multiplying the equation in the lemma by 2^i yields coefficients $a_i = 2^i \frac{\lambda_1^i - \lambda_2^i}{\lambda_1 - \lambda_2}$ and $b_i = 2^i \frac{\lambda_1 \lambda_2^i - \lambda_2 \lambda_1^i}{\lambda_1 - \lambda_2}$, such that $(2X_1)^i = a_i X_1 + b_i I_T$ and $(2X_2)^i = a_i X_2 + b_i I_T$ hold for all $i \in \mathbb{N}_0$. Therefore, for any $i, j \in \mathbb{N}_0$ one obtains

$$\begin{aligned}\operatorname{Tr}((2X_1)^i (2X_2)^j) &= \operatorname{Tr}(a_i a_j X_1 X_2 + a_i b_j X_1 + b_i a_j X_2 + b_i b_j I_T)\\ &= a_i a_j \operatorname{Tr}(X_1 X_2) + a_i b_j \operatorname{Tr}(X_1) + b_i a_j \operatorname{Tr}(X_2) + b_i b_j T.\end{aligned}$$

Combining all these simplifications yields the equation

$$\|\exp(X_1)\exp(X_2)\|_F = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{i!j!}\left(a_i a_j \operatorname{Tr}(X_1 X_2) + a_i b_j \operatorname{Tr}(X_1) + b_i a_j \operatorname{Tr}(X_2) + b_i b_j T\right).$$

The assertion of the lemma can now be shown by splitting the infinite series into a sum of products of several convergent series. In order to do this, the corresponding series are defined and it is checked that they are in fact convergent.

To start with, the series $\sum_{i=0}^{\infty} \frac{(2\lambda_1)^i}{i!} = \exp(2\lambda_1)$ and $\sum_{i=0}^{\infty} \frac{(2\lambda_2)^i}{i!} = \exp(2\lambda_2)$ both correspond to the series expansion of the real exponential map and are consequently convergent. Therefore, the equation

$$\sum_{i=0}^{\infty} \frac{a_i}{i!} = \sum_{i=0}^{\infty} \frac{2^i}{i!}\frac{\lambda_1^i - \lambda_2^i}{\lambda_1 - \lambda_2} = \frac{1}{\lambda_1 - \lambda_2}\left(\sum_{i=0}^{\infty}\frac{(2\lambda_1)^i}{i!} - \sum_{i=0}^{\infty}\frac{(2\lambda_2)^i}{i!}\right)$$

holds and the series on the left hand side of the equation is also convergent. By the same argument also the following equation holds and the series on its left hand side is convergent as well:

$$\sum_{i=0}^{\infty} \frac{b_i}{i!} = \sum_{i=0}^{\infty} \frac{2^i}{i!}\frac{\lambda_1\lambda_2^i - \lambda_2\lambda_1^i}{\lambda_1 - \lambda_2} = \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1\sum_{i=0}^{\infty}\frac{(2\lambda_2)^i}{i!} - \lambda_2\sum_{i=0}^{\infty}\frac{(2\lambda_1)^i}{i!}\right).$$

Now define $A := \sum_{i=0}^{\infty} \frac{a_i}{i!}$ and $B := \sum_{i=0}^{\infty} \frac{b_i}{i!}$. Furthermore, set $C := \lambda_1\mu_1 + \lambda_2\mu_2 \in \mathbb{R}$. This is chosen so that for any $X \in \mathcal{H}_{S_0}$ the equality $C = \operatorname{Tr}(X)$ holds. Finally, one obtains

$$\begin{aligned}
&A^2 \operatorname{Tr}(X_1 X_2) + 2ABC + B^2 T \\
&= A^2 \operatorname{Tr}(X_1 X_2) + AB \operatorname{Tr}(X_1) + AB \operatorname{Tr}(X_2) + B^2 T \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{a_i a_j}{i!j!} \operatorname{Tr}(X_1 X_2) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{a_i b_j}{i!j!} \operatorname{Tr}(X_1) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{b_i a_j}{i!j!} \operatorname{Tr}(X_2) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{b_i b_j}{i!j!} \operatorname{Tr}(I_T) \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{a_i a_j}{i!j!} \operatorname{Tr}(X_1 X_2) + \frac{a_i b_j}{i!j!} \operatorname{Tr}(X_1) + \frac{b_i a_j}{i!j!} \operatorname{Tr}(X_2) + \frac{b_i b_j}{i!j!} \operatorname{Tr}(I_T) \\
&= \|\exp(X_1) \exp(X_2)\|_F .
\end{aligned}$$

Therefore, the map $\varphi : \mathbb{R} \to \mathbb{R}$, $x \mapsto A^2 x + 2ABC + B^2 T$ satisfies the equality $\|\exp(X_1)\exp(X_2))\|_F = \varphi(\operatorname{Tr}(X_1 X_2))$ for any $X_1, X_2 \in \mathcal{H}_{\mathcal{S}_0}$. The values for A, B, C given in the lemma are apparent from the definitions of A, B and C in this proof and since $\lambda_1 \neq \lambda_2$ also $A > 0$ holds. In particular, A is not zero and hence φ is strictly monotonously growing. □

The map $(X_1, X_2) \mapsto \operatorname{Tr}(\overline{X_1}^\top X_2)$ is an inner product on $\mathbb{C}^{T \times T}$ which induces the Frobenius norm. Therefore, the GLRT distance can be connected with the distance induced by the Frobenius norm as follows:

(6.17) Theorem

Consider a spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ for two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mu_1, \mu_2 \in \mathbb{N}$.

The following two optimization problems are equivalent for $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$:

1) $\max_{X_1, X_2 \in \mathcal{C}_0, X_1 \neq X_2} \|\exp(X_1)\exp(X_2)\|_F$ is minimized.

2) $\min_{X_1, X_2 \in \mathcal{C}_0, X_1 \neq X_2} \|X_1 - X_2\|_F$ is maximized. ◇

Proof. It is sufficient to show that there is a strictly monotonously decreasing function $\psi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfying $\|\exp(X_1)\exp(X_2)\|_F = \psi(\|X_1 - X_2\|_F)$ for any matrices $X_1, X_2 \in \mathcal{H}_{\mathcal{S}_0}$.

By the previous lemma there is a strictly monotonously growing map $\varphi : \mathbb{R} \to \mathbb{R}$ for which $\|\exp(X_1)\exp(X_2))\|_F = \varphi(\operatorname{Tr}(X_1 X_2))$ holds for any $X_1, X_2 \in \mathcal{H}_{\mathcal{S}_0}$.

$$\begin{aligned}
\|X_1 - X_2\|_F^2 &= \operatorname{Tr}(\overline{(X_1 - X_2)}^\top (X_1 - X_2)) \\
&= \operatorname{Tr}(\overline{X_1}^\top X_1 - X_1 X_2 - X_2 X_1 + \overline{X_2}^\top X_2) \\
&= \|X_1\|_F^2 + \|X_2\|_F^2 - 2 \operatorname{Tr}(X_1 X_2)
\end{aligned}$$

The norm of any $X_1, X_2 \in \mathcal{H}_{S_0}$ is given by $C := \|X_1\|_F^2 = \|X_2\|_F^2 = \mu_1\lambda_1^2 + \mu_2\lambda_2^2$. Therefore, $\psi(x) := \varphi(C - \frac{1}{2}x^2)$ defines a strictly monotonously decreasing map satisfying the desired equation. □

This optimization problem may even be transferred to Euclidean space. More precisely, there is an isometry from n^2 dimensional Euclidean space to the space of Hermitian $n \times n$ matrices with the distance induced by the Frobenius norm.

(6.18) Remark
There are isometries of the Euclidean space $\mathbb{R}^{n^2}$ to the following inner product spaces:

1) The real $n \times n$ matrices $\mathbb{R}^{n\times n}$ with inner product $(X_1, X_2) \mapsto \operatorname{Tr}(X_1 X_2^\top)$.

2) The Hermitian $n \times n$ matrices $\mathcal{H}(n)$ with inner product $(X_1, X_2) \mapsto \operatorname{Tr}(X_1 \overline{X_2}^\top)$.

An isometry $\mathbb{R}^{n^2} \to \mathbb{R}^{n\times n}$ may be obtained by arranging the entries of a vector into a matrix.

A simple isometry $\mathbb{R}^{n\times n} \to \mathcal{H}(n)$ is given by $X \mapsto \frac{1}{2}((X + X^\top) + i(X - X^\top))$. ◇

Consequently, Hermitian matrices can be isometrically parametrized by elements of Euclidean space. Note that the Hermitian matrices with a specified fixed spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ are exactly the Hermitian matrices of which the characteristic polynomial is exactly the polynomial of which the zeros are exactly the eigenvalues of given multiplicities. Therefore, good codebooks may be obtained by regarding the elements of $\mathcal{H}_{\mathcal{S}_0}$ as elements of Euclidean space which additionally satisfy a given polynomial equation.

(6.19) Corollary
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ for $\lambda_1 \neq \lambda_2$, the corresponding polynomial $P = (x - \lambda_1)^{\mu_1}(x - \lambda_2)^{\mu_2} \in \mathbb{R}[x]$ and an isometry $\varphi : \mathbb{R}^{T^2} \to \mathcal{H}(T)$ from Euclidean T^2-dimensional space into the Hermitian $T \times T$ matrices with inner product given by $(X_1, X_2) \mapsto \operatorname{Tr}(X_1 \overline{X_2}^\top)$. Moreover, denote the characteristic polynomial of $X \in \mathcal{H}(T)$ by $P_X \in \mathbb{R}[x]$.

If $\mathcal{D} \subseteq \{v \in \mathbb{R}^{T^2} \mid P_{\varphi(v)} = P\}$ is a finite set such that the minimal euclidean distance $\min_{v_1, v_2 \in \mathcal{D}, v_1 \neq v_2} \|v_1 - v_2\|_F$ is maximized, $\mathcal{C}_0 := \varphi(\mathcal{D}) \subseteq \mathcal{H}_{\mathcal{S}_0}$ is a finite set such that $\max_{X_1, X_2 \in \mathcal{C}_0, X_1 \neq X_2} \operatorname{Tr}(\exp(X_1)\exp(X_2))$ is minimized. ◇

For 2×2 matrices this problem turns out to be particularly simple. As a result the optimal 2×2 codes for a given spectrum $\mathcal{S}$ can be found.

(6.20) Example (Noncoherent 2 × 2 STBC for the GLRT-Receiver)
Consider the spectrum $\mathcal{S}_0 = \{(1,1), (-1,1)\}$. In the following, a noncoherent STBC $\mathcal{C} \subseteq \mathcal{H}(2)$ will be constructed by means of the map $\exp : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}(2)$. To that end, a good packing $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ has to be found. According to theorem 6.17 a good codebook with respect to the GLRT distance may then be chosen as $\mathcal{C} = \{\exp(X) \mid X \in \mathcal{C}_0\}$.

To start with, a connection of $\mathcal{H}_{\mathcal{S}_0}$ to Euclidean space is developed. An isometry from 4-dimensional Euclidean space into $\mathcal{H}(2)$ with inner product as in the preceding corollary is given by

$$\mathbb{R}^4 \to \mathcal{H}(2), \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \mapsto \begin{pmatrix} a & \frac{1}{\sqrt{2}}(b+ci) \\ \frac{1}{\sqrt{2}}(b-ci) & d \end{pmatrix}.$$

The characteristic polynomial of all elements of $\mathcal{H}_{\mathcal{S}_0}$ is given by $x^2 - 1 \in \mathbb{R}[x]$. Equivalently all matrices $X = \begin{pmatrix} a & b+ci \\ b-ci & d \end{pmatrix} \in \mathcal{H}_{\mathcal{S}_0}$ satisfy $\mathrm{Tr}(X) = a + d = 0$ and $\det(X) = ad - b^2 - c^2 = -1$ and hence $d = -a$ and $a^2 + b^2 + c^2 = 1$.

That yields $\mathcal{H}_{\mathcal{S}_0} = \left\{ \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix} \middle| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in S^3 \subseteq \mathbb{R}^3 \right\}$, where S^3 denotes the unit sphere in $\mathbb{R}^3$. It can also be seen that $\mathbb{R}^3 \to \mathcal{H}(2), \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \frac{1}{\sqrt{2}} \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix}$ is an isometrical embedding. Therefore, packing a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ with respect to the canonical Hermitian distance is equivalent to packing a finite set $\mathcal{D} \subseteq S^3$ with respect to the euclidean distance and writing

$$\mathcal{C}_0 = \left\{ \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix} \middle| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathcal{D} \right\}.$$

Finding good euclidean packings on the n-dimensional unit sphere is a well known problem and corresponding sets are known as *spherical codes*. Spherical codes are a well investigated subject. For example, they are extensively treated in the book of Conway and Sloane [CS98]. Good spherical codes in high dimensions may be obtained from lattices and for lower dimensions optimal codes can be computed by optimization algorithms. A list of putatively optimal spherical codes in $\mathbb{R}^3, \mathbb{R}^4$ and

$\mathbb{R}^5$ for codebooks consisting of $m = 4, \ldots, 130$ points can be found on the web page of Neil Sloane [Slo].

Consider a spherical code $\mathcal{D} \subseteq S^3$ and a finite subset $\mathcal{C}_0 \subset \mathcal{H}_{\mathcal{S}_0}$.

$$\mathcal{C}_0 := \left\{ \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix} \,\middle|\, \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathcal{D} \right\}.$$

Following the design criteria, a square noncoherent STBC may now be constructed as

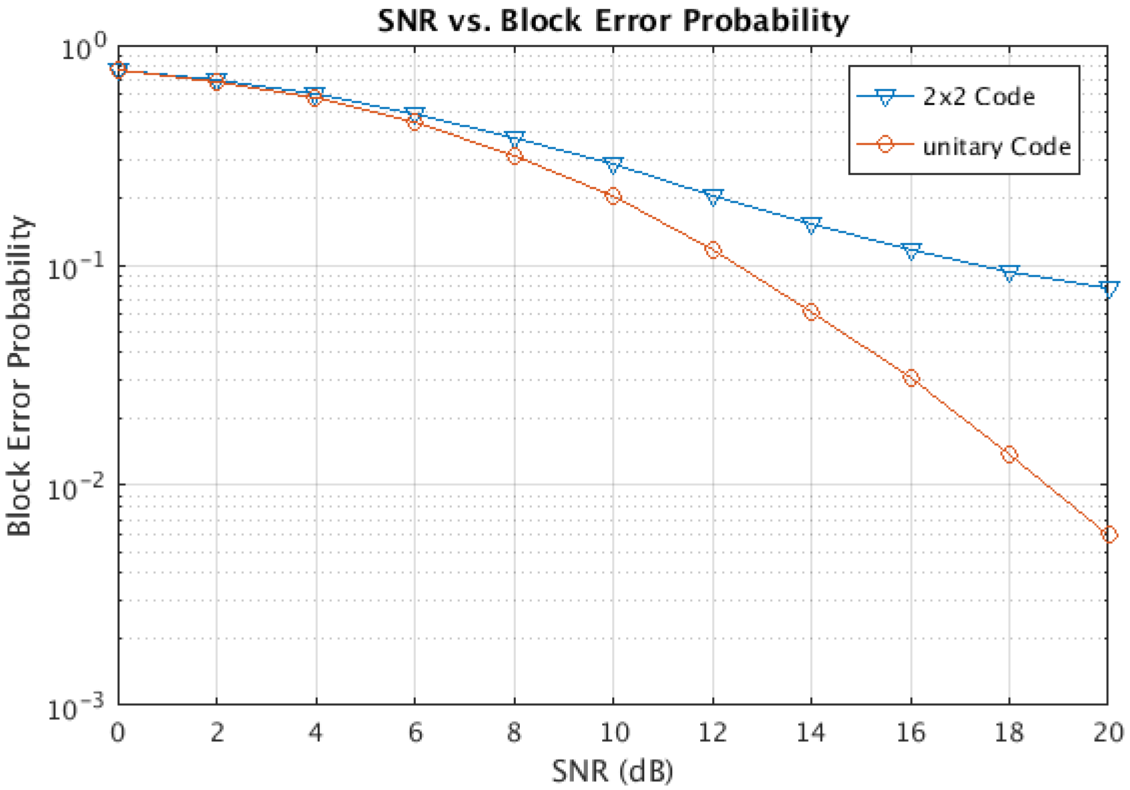
$$\mathcal{C} = \{\exp(X) \mid X \in \mathcal{C}_0\}.$$

The performance of a code $\mathcal{C}$ constructed in the manner that was described above shall now be compared to the performance of an optimal unitary code. In particular, $\mathcal{C}$ will be designed to yield a rate of $\operatorname{rate}(\mathcal{C}) = \frac{\log_2(|\mathcal{C}|)}{T} = 2$ information symbols per time step. Since 2×2 matrices are considered, $T = 2$ holds, which implies $|\mathcal{C}| = 2^4 = 16$.

Consequently, a spherical code in $\mathbb{R}^3$ with 16 elements has to be used to construct $\mathcal{C}$. For the following comparison, a putatively optimal spherical code $\mathcal{D}$ listed in [Slo] with $|\mathcal{D}| = 16$ and a minimal angle between the points of $\mathcal{D}$ of 52.2444 degrees is used.

The code will be compared with a unitary codebook $\mathcal{C}^u \subseteq \mathbb{C}^{1\times 2}$ obtained by the optimization algorithm presented in section 4.4. The unitary codebook also satisfies $|\mathcal{C}^u| = 16$ and, therefore, yields a rate of 2 bit per time step. Moreover, it realizes a minimal distance of $\min_{X_1 \neq X_2, X_1, X_2 \in \mathcal{C}^u} - \left\| \overline{X_1}^\top X_2 \right\|_F = -0.9059$, when all codewords $X \in \mathcal{C}^u$ are normalized to satisfy $\|X\|_F = 1$.

All simulations were run under the assumption that two receive antennas are employed. The error rate that was obtained for each code at different signal-to-noise ratios is plotted below.



It can be seen that the unitary code clearly outperforms the newly constructed 2 × 2 code. However, the spectrum $\mathcal{S}_0$ used to construct the 2 × 2 code was chosen solely for simplicity. Therefore, it has to be investigated how the spectrum should be chosen in order to obtain a better codebook. This will be done in the following section. ◇

6.3 Optimal spectra for the parametrization

In this section, the question how the eigenvalues λ_1 and λ_2 are chosen optimally will be investigated. The most important tool to answer this question is an improved parametrization of the codewords. It is introduced in the following definition.

(6.21) Definition
For $\alpha, \beta \in \mathbb{R}$, $\beta > 0$ and $n \in \mathbb{N}$ define the map

$$\exp_{\alpha,\beta} : \mathcal{H}(n) \to \mathcal{H}^+(n),\ X \mapsto \beta \exp(\alpha X).$$

◇

The map $\exp_{\alpha,\beta}$ yields a useful parametrization of positive definite Hermitian matrices. It will ultimately allow the comparison of different codebooks $\mathcal{C} \subseteq \mathcal{H}_{\mathcal{S}}$ and $\mathcal{C}' \subseteq \mathcal{H}_{\mathcal{S}'}$ for $\mathcal{S} = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ and $\mathcal{S}' = \{(\lambda_1', \mu_1), (\lambda_2', \mu_2)\}$ with real numbers $\lambda_1, \lambda_2, \lambda_1', \lambda_2' > 0$ and $\mu_1, \mu_2 \in \mathbb{N}$. The following lemmata prepare this comparison. As a first step it will be shown that any such codebook can be parametrized by the map $\exp_{\alpha,\beta}$ for some $\alpha, \beta \in \mathbb{R}$.

(6.22) Lemma

For two given spectra $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ and $\mathcal{S} = \{(\lambda_1', \mu_1), (\lambda_2', \mu_2)\}$ with $\lambda_1, \lambda_2, \lambda_1', \lambda_2' \in \mathbb{R}$, $\lambda_1 \neq \lambda_2$, $\lambda_1' \neq \lambda_2'$, $\lambda_1', \lambda_2' > 0$ and $\mu_1, \mu_2 \in \mathbb{N}$, there are $\alpha, \beta \in \mathbb{R}$, $\alpha \neq 0$, $\beta > 0$ such that the restricted map $\exp_{\alpha,\beta} : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}}$ is well defined and bijective. ◇

Proof. First suppose that $\alpha, \beta \in \mathbb{R}$ with $\alpha \neq 0$ and $\beta > 0$ are given and define the spectra

$$\begin{aligned} \mathcal{S}_1 &= \{(\alpha\lambda_1, \mu_1), (\alpha\lambda_2, \mu_2)\}, \\ \mathcal{S}_2 &= \{(\exp(\alpha\lambda_1), \mu_1), (\exp(\alpha\lambda_2), \mu_2)\} \text{ and} \\ \mathcal{S}_3 &= \{(\beta\exp(\alpha\lambda_1), \mu_1), (\beta\exp(\alpha\lambda_2), \mu_2)\}. \end{aligned}$$

According to lemma 6.13 the restricted map $\exp : \mathcal{H}_{\mathcal{S}_1} \to \mathcal{H}_{\mathcal{S}_2}$ is bijective. Also, the maps $l_\alpha : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}_1}$, $X \mapsto \alpha X$ and $l_\beta : \mathcal{H}_{\mathcal{S}_2} \to \mathcal{H}_{\mathcal{S}_3}$, $X \mapsto \beta X$ are bijective and hence $\exp_{\alpha,\beta} = l_\beta \circ \exp \circ l_\alpha : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}_3}$ is a bijection.

By that observation it is sufficient to find suitable parameters α, β such that $\mathcal{S}_3 = \mathcal{S}$ holds in order to obtain the assertion of the lemma. In particular, it remains to be shown that there are $\alpha, \beta \in \mathbb{R}$ with $\alpha \neq 0$ and $\beta > 0$ such that $\lambda_1' = \beta\exp(\alpha\lambda_1)$ and $\lambda_2' = \beta\exp(\alpha\lambda_2)$ hold.

To that end, set $\alpha := \frac{\ln(\lambda_1') - \ln(\lambda_2')}{\lambda_1 - \lambda_2}$ and $\beta := \frac{\lambda_1'}{\exp(\alpha\lambda_1)}$. This implies

$$\frac{\lambda_1'}{\lambda_2'} = \exp(\alpha(\lambda_1 - \lambda_2)) \quad \text{and} \quad \lambda_1' = \beta\exp(\alpha\lambda_1).$$

By plugging the second equation into the first one, one also obtains

$$\lambda_2' = \frac{\lambda_2'}{\lambda_1'}\lambda_1' = \exp(\alpha(\lambda_2 - \lambda_1)) \cdot \beta\exp(\alpha\lambda_1) = \beta\exp(\alpha\lambda_2).$$

So, the defined α and β satisfy the required condition and hence the assertion follows. □

This lemma simplifies the study of codebooks made up of codewords with one fixed spectrum. It implies that for given multiplicities μ_1 and μ_2 it is sufficient to study codes obtained from the sets $\exp_{\alpha,\beta}(\mathcal{H}_{\mathcal{S}_0})$ for $\alpha, \beta \in \mathbb{R}$, $\alpha \neq 0$, $\beta > 0$ and any suitable spectrum $\mathcal{S}_0$. Therefore, it will be investigated how the values α and β influence the performance of resulting codes. In order to compare the performance of two codes in a reasonable manner, they have to provide the same signal-to-noise ratio when transmitting information across the same channel. For the class of codewords considered here this is equivalent to restricting the Frobenius norm of every codeword of both codebooks in question to an identical value $\rho \in \mathbb{R}$.

Suppose that such a value ρ is fixed. It will now be shown that all corresponding codebooks can be parametrized by the map $\exp_{\alpha,\beta}$ for some $\alpha \in \mathbb{R}$ and an element $\beta \in \mathbb{R}$ completely determined by α and ρ.

(6.23) Lemma
Consider two given spectra $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ and $\mathcal{S} = \{(\lambda'_1, \mu_1), (\lambda'_2, \mu_2)\}$ with $\lambda_1, \lambda_2, \lambda'_1, \lambda'_2 \in \mathbb{R}$, $\lambda_1 \neq \lambda_2$, $\lambda'_1 \neq \lambda'_2$, $\lambda'_1, \lambda'_2 > 0$, $\mu_1, \mu_2 \in \mathbb{N}$ and $\rho \in \mathbb{R}_{>0}$, such that $\|X\|_F = \rho$ holds for all $X \in \mathcal{H}_{\mathcal{S}}$.

Moreover, define the map $\beta : \mathbb{R} \to \mathbb{R}$, $\alpha \mapsto \frac{\rho}{\|\exp(\alpha X)\|_F}$ for an arbitrary $X \in \mathcal{H}_{\mathcal{S}}$.

There is $\alpha \in \mathbb{R}$ such that the restricted map $\exp_{\alpha,\beta(\alpha)} : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}}$ is well defined and bijective. ◇

Proof. By lemma 6.22 there are $\alpha, \hat{\beta} \in \mathbb{R}$, $\alpha \neq 0$, $\hat{\beta} > 0$ such that $\exp_{\alpha,\hat{\beta}} : \mathcal{H}_{\mathcal{S}_0} \to \mathcal{H}_{\mathcal{S}}$ is well defined and bijective. This α satisfies the assertion of the lemma. In order to see that, it needs to be shown that $\hat{\beta} = \beta(\alpha)$ holds, where β is the map defined in the lemma.

As the map is well defined, $\exp_{\alpha,\hat{\beta}}(X) \in \mathcal{H}_{\mathcal{S}}$ holds for any $X \in \mathcal{H}_{\mathcal{S}_0}$. By the assumption on $\mathcal{S}$ in the lemma this implies $\left\|\exp_{\alpha,\hat{\beta}}(X)\right\|_F = \|\hat{\beta}\exp(\alpha X)\|_F = \rho$ and hence $\hat{\beta} = \frac{\rho}{\|\exp(\alpha X)\|_F}$. □

So far the code $\mathcal{C} = \{\exp(X) \mid X \in \mathcal{C}_0\}$ has been considered for a finite subset $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$. Using the newly introduced parametrization, a new class of codes can be constructed from $\mathcal{C}_0$.

(6.24) Definition
Consider a real spectrum $\mathcal{S}_0$, a finite subset $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ and the signal-to-noise ratio $\rho \in \mathbb{R}_{>0}$.

For $\alpha \in \mathbb{R}$ define $\mathcal{C}_0^\alpha := \left\{ \frac{\rho}{\|\exp(\alpha X)\|_F} \exp(\alpha X) \,\middle|\, X \in \mathcal{C}_0 \right\}$. ◇

(6.25) Remark
The significance of the previous definition is that for any positive definite spectrum $\mathcal{S} = \{(\lambda'_1, \mu_1), (\lambda'_2, \mu_2)\}$ an arbitrary codebook $\mathcal{C} \subseteq \mathcal{H}_{\mathcal{S}}$ that yields the signal-to-noise ratio ρ can be described using this definition.

In particular, for any spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$, there is a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ and $\alpha \in \mathbb{R}$ such that $\mathcal{C} = \mathcal{C}_0^\alpha$. ◇

It will now be studied how α may be chosen for a given set $\mathcal{C}_0$, in order for the code $\mathcal{C}_0^\alpha$ to perform as good as possible. Before precise results are presented, some example codes are constructed and compared in simulations.

(6.26) Example (Noncoherent 2 × 2 STBC for the GLRT-Receiver)
Consider the spectrum $\mathcal{S}_0 = \{(1,1), (-1,1)\}$. In the following, the noncoherent codebooks $\mathcal{C}_0^\alpha \subseteq \mathcal{H}(2)$ will be constructed for a packing $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ and several $\alpha \in \mathbb{R}$. Analogously to example 6.20, one obtains

$$\mathcal{H}_{\mathcal{S}_0} = \left\{ \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix} \middle| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in S^3 \subseteq \mathbb{R}^3 \right\}.$$

Moreover, finding a good packing $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ is, again, equivalent to finding a good spherical code $\mathcal{D} \subseteq S^3 \subseteq \mathbb{R}^3$ and setting

$$\mathcal{C}_0 := \left\{ \begin{pmatrix} a & b+ci \\ b-ci & -a \end{pmatrix} \middle| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathcal{D} \right\}.$$

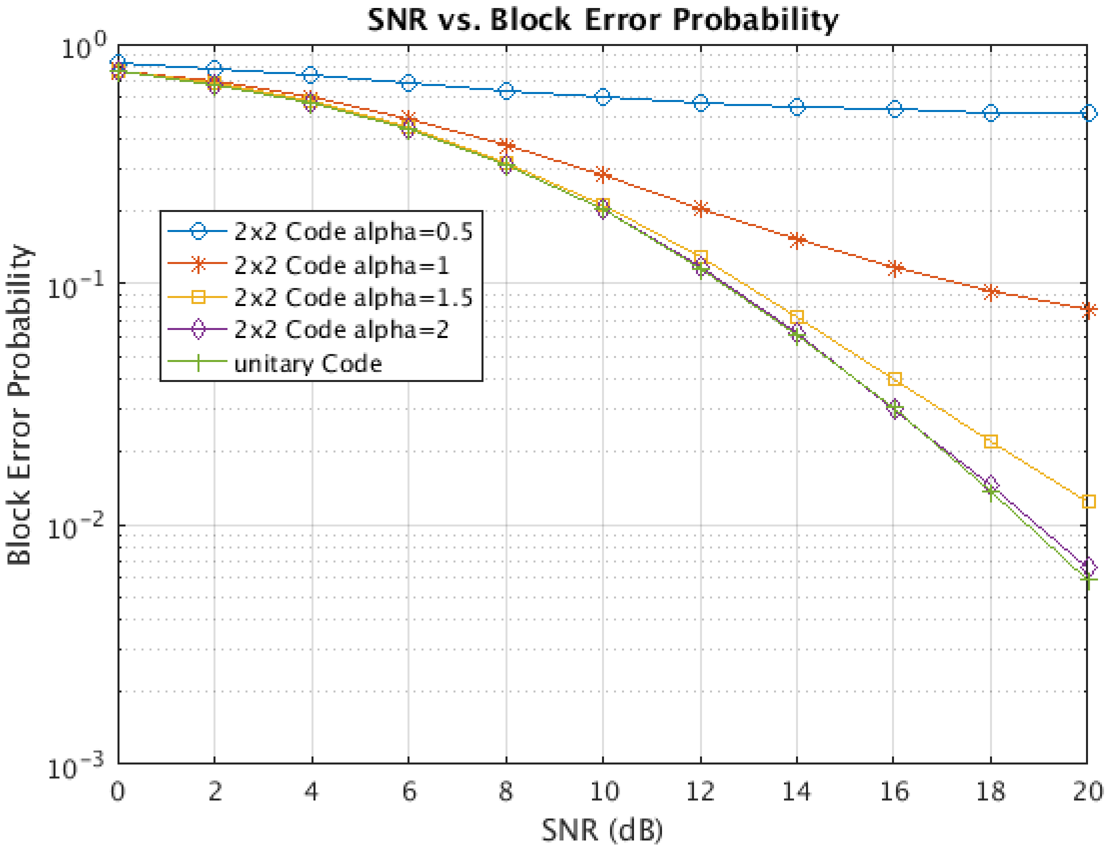
Also as in example 6.20, a spherical code $\mathcal{D}$ with $|\mathcal{D}| = 16$ and minimal angle of 52.2444 degrees is used to obtain $\mathcal{S}_0$. Therefore, the codebooks

$$\mathcal{C}_0^\alpha := \left\{ \frac{\rho}{\|\exp(\alpha X)\|_F} \exp(\alpha X) \middle| X \in \mathcal{C}_0 \right\}$$

yield a rate of 2 bit per time step. The codes will be compared with a unitary codebook $\mathcal{C}^u \subseteq \mathbb{C}^{1\times 2}$ obtained by the optimization algorithm presented in section 4.4. The unitary codebook also satisfies $|\mathcal{C}^u| = 16$ and, therefore, yields a rate of 2 bit per time step. Moreover, assuming all codewords $X \in \mathcal{C}^u$ are normalized to satisfy $\|X\|_F = 1$, its minimal distance with respect to d_{GLRT} is given by

$$\min_{X_1 \neq X_2, X_1, X_2 \in \mathcal{C}^u} d_{\text{GLRT}}(X_1, X_2) = \min_{X_1 \neq X_2, X_1, X_2 \in \mathcal{C}^u} -\left\| \overline{X_1}^\top X_2 \right\|_F = -0.9059.$$

All simulations were run under the assumption that two receive antennas are employed. The error rate that was obtained for each code at different signal-to-noise ratios is plotted below.



The error rate of the codebooks $\mathcal{C}_0^\alpha$ improves for larger values of α. However, for all values of α it is larger than the error rate of the unitary code. In fact, the performance of the codebooks $\mathcal{C}_0^\alpha$ appears to approach the performance of $\mathcal{C}^u$ for growing α. Already for $\alpha = 2$ it can hardly be distinguished from the performance of the unitary codebook. Simulations have also been run for larger values of α, reinforcing the conjecture that the error rate of $\mathcal{C}_0^\alpha$ converges to the error rate of $\mathcal{C}^u$ for $\alpha \to \infty$. $\diamond$

The example suggests a specific behavior of the performance of codebooks of the form $\mathcal{C}_0^\alpha$ depending on the parameter α. Consequently, the performance of such codebooks and its dependence on α will be studied in the following.

One should bear in mind that the performance of two codebooks $\mathcal{C}^{\alpha_1}$ and $\mathcal{C}^{\alpha_2}$ for $\alpha_1 \neq \alpha_2$ cannot be compared by means of the distance functions d_{GLRT} or d_5 from proposition 6.9. These codebooks are subsets of distinct coding spaces $\mathcal{H}_{\mathcal{S}_1}$ and $\mathcal{H}_{\mathcal{S}_2}$ for distinct Spectra $\mathcal{S}_1$ and $\mathcal{S}_2$, and the mentioned distance functions have only been shown to be equivalent within one such coding space. Therefore, in order to compare the two codebooks, the distance function Δ which was derived directly from the decoding probability has to be considered. As a first step, the value of the distance function between codewords $X_1, X_2 \in \mathcal{C}_0^\alpha$ for $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ built from a specific spectrum $\mathcal{S}_0$, can be shown to be monotonously increasing in α.

(6.27) Proposition
Consider the real spectrum $\mathcal{S} = \{(1, \mu_1), (0, \mu_2)\}$ for $\mu_1, \mu_2 \in \mathbb{N}$, the signal-to-noise ratio $\rho \in \mathbb{R}_{>0}$ and $X_1, X_2 \in \mathcal{H}_\mathcal{S}$.

Moreover, define $\beta : \mathbb{R} \to \mathbb{R}$, $\alpha \mapsto \frac{\rho}{\|\exp(\alpha X)\|_F}$ for an arbitrary $X \in \mathcal{H}_\mathcal{S}$.

Then, the map $\mathbb{R}_{>0} \to \mathbb{R}$, $\alpha \mapsto \Delta\left(\exp_{\alpha,\beta(\alpha)}(X_1), \exp_{\alpha,\beta(\alpha)}(X_2)\right)$ is continuous and strictly monotonously increasing. ◇

Proof. The assertion will be shown by computing the derivative of the map in question and by showing that it is strictly positive. In particular, the map is differentiable and therefore continuous, as claimed.

To start with, abbreviate $X^{(\alpha)} := \exp_{\alpha,\beta(\alpha)}(X)$ for any $X \in \mathcal{H}_\mathcal{S}$. It will be shown how $\Delta\left(\exp_{\alpha,\beta(\alpha)}(X_1), \exp_{\alpha,\beta(\alpha)}(X_2)\right)$ can be completely parametrized by elementary functions in α.

The eigenvalues of $\exp(\alpha X)$ for $X \in \mathcal{H}_\mathcal{S}$ are given by e^α of multiplicity μ_1 and $1 = e^0$ of multiplicity μ_2. Therefore, $\|\exp(\alpha X)\|_F^2 = \mu_1 e^{2\alpha} + \mu_2$. Consequently, the eigenvalues of $X^{(\alpha)} = \beta(\alpha)\exp(\alpha X)$ for $X \in \mathcal{H}_\mathcal{S}$ are given by

$$\begin{aligned} \lambda_1^{(\alpha)} &:= \frac{\rho e^\alpha}{\sqrt{\mu_1 e^{2\alpha} + \mu_2}} \text{ of multiplicity } \mu_1 \text{ and} \\ \lambda_2^{(\alpha)} &:= \frac{\rho}{\sqrt{\mu_1 e^{2\alpha} + \mu_2}} \text{ of multiplicity } \mu_2. \end{aligned}$$

Now, denote

$$f : \mathbb{R} \to \mathbb{R},\ \alpha \mapsto \frac{-1}{\left(1 + \left(\lambda_1^{(\alpha)}\right)^2\right)\left(1 + \left(\lambda_2^{(\alpha)}\right)^2\right)}$$

and

$$g : \mathbb{R} \to \mathbb{R},\ \alpha \mapsto 1 + \left(\lambda_1^{(\alpha)}\right)^2 + \left(\lambda_2^{(\alpha)}\right)^2.$$

Lemma 6.7 implies the following:

$$\begin{aligned}\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right) &= \operatorname{Tr}\left(\Lambda^{-1}_{X_1^{(\alpha)}}\Lambda_{X_2^{(\alpha)}}\right)\\ &= \operatorname{Tr}\left(f(\alpha)\left(\left(X_1^{(\alpha)}\right)^2 - g(\alpha)I_T\right)\left(\left(X_2^{(\alpha)}\right)^2 + I_T\right)\right)\\ &= f(\alpha)\operatorname{Tr}\left(\left(X_1^{(\alpha)}\right)^2\left(X_2^{(\alpha)}\right)^2 + \left(X_1^{(\alpha)}\right)^2 - g(\alpha)\left(\left(X_2^{(\alpha)}\right)^2 + I_T\right)\right)\\ &= f(\alpha)\left(\left\|X_1^{(\alpha)}X_2^{(\alpha)}\right\|_F^2 + \left\|X_1^{(\alpha)}\right\|_F^2 - g(\alpha)\left(\left\|X_2^{(\alpha)}\right\|_F^2 + T\right)\right)\\ &= f(\alpha)\left(\left\|X_1^{(\alpha)}X_2^{(\alpha)}\right\|_F^2 + \rho^2 - g(\alpha)(\rho^2 + T)\right).\end{aligned}$$

Concerning the last equality, note that by definition of the map β all $X \in \mathcal{H}_{\mathcal{S}}$ satisfy $\left\|X^{(\alpha)}\right\|_F = \|\beta(\alpha)\exp(\alpha X)\|_F = \rho$.

In order to compute the derivative of the obtained expression with respect to α, it remains to find an explicit expression of $\left\|X_1^{(\alpha)}X_2^{(\alpha)}\right\|_F^2$ in terms of α.

To that end, note that by lemma 6.16 we have the following:

$$\|\exp(\alpha X_1)\exp(\alpha X_2)\|_F^2 = A^2\operatorname{Tr}(\alpha X_1 \alpha X_2) + 2ABC + B^2 T,$$

where A, B, C are real numbers given by

$$\begin{aligned}A &= \frac{e^{2\alpha} - e^0}{\alpha - 0} = \frac{e^{2\alpha} - 1}{\alpha},\\ B &= \frac{\alpha e^0 + 0e^{2\alpha}}{\alpha - 0} = 1,\\ C &= \operatorname{Tr}(\alpha X_1) = \operatorname{Tr}(\alpha X_2) = \alpha\mu_1.\end{aligned}$$

Multiplying both sides of this equation by $\beta(\alpha)^4$ yields

$$\begin{aligned}\left\|X_1^{(\alpha)}X_2^{(\alpha)}\right\|_F^2 &= \beta(\alpha)^4\|\exp(\alpha X_1)\exp(\alpha X_2)\|_F^2\\ &= \frac{\rho^4}{(\mu_1 e^{2\alpha} + \mu_2)^2}\left(\left(\frac{e^{2\alpha} - 1}{\alpha}\right)^2\operatorname{Tr}(\alpha X_1 \alpha X_2) + 2\frac{e^{2\alpha} - 1}{\alpha}\mu_1\alpha + T\right)\\ &= \frac{\rho^4}{(\mu_1 e^{2\alpha} + \mu_2)^2}\left((e^{2\alpha} - 1)^2\operatorname{Tr}(X_1X_2) + 2\mu_1(e^{2\alpha} - 1) + T\right).\end{aligned}$$

Putting all the above calculations together, $\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right)$ can be expressed as a rational function in e^{α}. The computation of its derivative is elementary, but lengthy when carried out by hand. It was checked in matlab that it is given by

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\,\alpha}\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right)\\&= \frac{2\rho^2 e^{2\alpha}(e^{2\alpha}-1)(\mu_1 - \operatorname{Tr}(X_1X_2))(\rho^2+\mu_1+\mu_2)((\rho^2+2\mu_1)e^{2\alpha}+\rho^2+2\mu_2)}{((\mu_1^2+\rho^2\mu_1)e^{4\alpha}+(\rho^4+2\mu_1\mu_2+\rho^2(\mu_1+\mu_2))e^{2\alpha}+\rho^2\mu_2+\mu_2^2)^2}.\end{aligned}$$

In order to show that this is strictly positive, all factors involved have to be considered. Note first that ρ, μ_1, μ_2 and all values of the exponential map are strictly positive. This leaves two remaining factors. For one, $(e^{2\alpha}-1)$ is easily seen to be strictly positive for $\alpha > 0$. Lastly, $(\mu_1 - \operatorname{Tr}(X_1X_2))$ has to be considered. As $X_1 \neq X_2$ and $\|X_1\|_F^2 = \|X_2\|_F^2 = \mu_1$ hold, the desired inequality is obtained as follows:

$$0 < \frac{1}{2}\|X_1 - X_2\|_F^2 = \frac{1}{2}\left(\|X_1\|_F^2 + \|X_1\|_F^2 - 2\operatorname{Tr}(X_1X_2)\right) = \mu_1 - \operatorname{Tr}(X_1X_2).$$

Therefore, $\frac{\mathrm{d}}{\mathrm{d}\,\alpha}\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right)$ is strictly positive for $\alpha > 0$ and the assertion follows. □

The assertion of the proposition describes the dependence of the value $\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right)$ on α for two fixed codewords X_1, X_2 possessing one specific spectrum. This can be generalized to the minimal distance of a codebook with respect to Δ. To this end, two technical lemmas are required.

(6.28) Lemma
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$, $\lambda_1 \neq \lambda_2$, a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$, $\rho \in \mathbb{R}_{>0}$ and denote $\beta : \mathbb{R} \to \mathbb{R}$, $\alpha \mapsto \frac{\rho}{\|\exp(\alpha X)\|_F}$ for an arbitrary $X \in \mathcal{H}_{\mathcal{S}_0}$.

There are $X_1^*, X_2^* \in \mathcal{C}_0$ such that for any $\alpha \in \mathbb{R}$ the minimal distance in $\mathcal{C}_0^{\alpha}$ with respect to Δ is attained by

$$(X_1^*)^{(\alpha)} := \beta(\alpha)\exp(\alpha X_1^*) \quad \text{and} \quad (X_2^*)^{(\alpha)} := \beta(\alpha)\exp(\alpha X_2^*).$$

That is

$$\Delta\left((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)}\right) = \min_{X_1, X_2 \in \mathcal{C}_0^{\alpha}, X_1 \neq X_2} \Delta(X_1, X_2).$$

◇

Proof. Choose $X_1^*, X_2^* \in \mathcal{C}_0$ which satisfy $\operatorname{Tr}(X_1^* X_2^*) = \max_{X_1,X_2\in\mathcal{C}_0, X_1\neq X_2} \operatorname{Tr}(X_1X_2)$. Equivalently, $\operatorname{Tr}(\alpha X_1^* \alpha X_2^*) = \max_{X_1,X_2\in\mathcal{C}_0, X_1\neq X_2} \operatorname{Tr}(\alpha X_1 \alpha X_2)$ holds, which by lemma 6.16 is equivalent to

$$\|\exp(\alpha X_1^*)\exp(\alpha X_2^*)\|_F = \max_{X_1,X_2\in\mathcal{C}_0,X_1\neq X_2} \|\exp(\alpha X_1)\exp(\alpha X_2)\|_F .$$

By definition of the distance function $d_{\mathrm{GLRT}} : (X_1, X_2) \mapsto -\left\|\overline{X_1}^\top X_2\right\|_F$, this is equivalent to

$$d_{\mathrm{GLRT}}(\exp(\alpha X_1^*), \exp(\alpha X_2^*)) = \min_{X_1,X_2\in\mathcal{C}_0,X_1\neq X_2} d_{\mathrm{GLRT}}(\exp(\alpha X_1), \exp(\alpha X_2)).$$

Multiplying both sides of this equation by $\beta(\alpha)^2$ yields

$$\begin{aligned} d_{\mathrm{GLRT}}((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)}) &= \min_{X_1,X_2\in\mathcal{C}_0,X_1\neq X_2} d_{\mathrm{GLRT}}(\beta(\alpha)\exp(\alpha X_1), \beta(\alpha)\exp(\alpha X_2)) \\ &= \min_{X_1',X_2'\in\mathcal{C}_0^\alpha,X_1'\neq X_2'} d_{\mathrm{GLRT}}(X_1', X_2'). \end{aligned}$$

For any spectrum $\mathcal{S}$ which contains exactly two distinct eigenvalues the distance functions Δ and d_{GLRT} are equivalent on $\mathcal{H}_\mathcal{S}$ by proposition 6.9. Because $\mathcal{C}_0^\alpha \subseteq \mathcal{H}_\mathcal{S}$ holds for $\mathcal{S} = \{(\beta(\alpha)\exp(\alpha\lambda_1), \mu_1), (\beta(\alpha)\exp(\alpha\lambda_2), \mu_2)\}$, also the following equality is satisfied:

$$\Delta\left((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)}\right) = \min_{X_1,X_2\in\mathcal{C}_0^\alpha,X_1\neq X_2} \Delta(X_1, X_2).$$

This is the assertion of the lemma. □

(6.29) Lemma
Consider two real spectra $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}, \mathcal{S}_0' = \{(\lambda_1', \mu_1), (\lambda_2', \mu_2)\}$ satisfying $\lambda_1 \neq \lambda_2$ as well as $\lambda_1' \neq \lambda_2'$ and consider $X \in \mathcal{H}_{\mathcal{S}_0}$.

There are $X' \in \mathcal{H}_{\mathcal{S}_0'}$ and $\alpha' \in \mathbb{R}_{>0}$ such that for any $\alpha, \rho \in \mathbb{R}$ the following is true

$$\frac{\rho}{\|\exp(\alpha X)\|_F}\exp(\alpha X) = \frac{\rho}{\|\exp((\alpha\alpha')X')\|_F}\exp((\alpha\alpha')X').$$

Proof. Without loss of generality, assume $\lambda_1 > \lambda_2$ and $\lambda_1' > \lambda_2'$. Otherwise, the variables can be renamed accordingly throughout this proof.

Suppose X is given by $X = UD\overline{U}^\top$ for $U \in U(T)$ and

$$D = \operatorname{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{\mu_1 \text{ times}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{\mu_2 \text{ times}}).$$

Then, by lemma 2.26, $\exp(\alpha X)$ is given by $U \exp(\alpha D)\overline{U}^\top$.

Define $\alpha' := \frac{\lambda_1-\lambda_2}{\lambda_1'-\lambda_2'} > 0$ and $X' := UD'\overline{U}^\top \in \mathcal{H}_{\mathcal{S}_0'}$ for

$$D' = \operatorname{diag}(\underbrace{\lambda_1', \dots, \lambda_1'}_{\mu_1 \text{ times}}, \underbrace{\lambda_2', \dots, \lambda_2'}_{\mu_2 \text{ times}}).$$

Then, the eigenvalues of $\exp((\alpha \cdot \alpha')X')$ are given by

$$\begin{aligned}
\exp(\alpha\alpha'\lambda_1') &= \exp\left(\alpha\lambda_1' \frac{\lambda_1-\lambda_2}{\lambda_1'-\lambda_2'}\right) = \exp\left(\alpha\lambda_1 - \alpha\lambda_1\frac{\lambda_1'-\lambda_2'}{\lambda_1'-\lambda_2'} + \alpha\lambda_1'\frac{\lambda_1-\lambda_2}{\lambda_1'-\lambda_2'}\right) \\
&= \exp\left(\alpha\lambda_1 + \alpha\frac{\lambda_1'(\lambda_1-\lambda_2) - \lambda_1(\lambda_1'-\lambda_2')}{\lambda_1'-\lambda_2'}\right) \\
&= \exp(\alpha\lambda_1)\exp\left(\alpha\frac{\lambda_1\lambda_2' - \lambda_1'\lambda_2}{\lambda_1'-\lambda_2'}\right)
\end{aligned}$$

and

$$\begin{aligned}
\exp(\alpha\alpha'\lambda_2') &= \exp\left(\alpha\lambda_2' \frac{\lambda_1-\lambda_2}{\lambda_1'-\lambda_2'}\right) = \exp\left(\alpha\lambda_2 + \alpha\frac{\lambda_2'(\lambda_1-\lambda_2) - \lambda_2(\lambda_1'-\lambda_2')}{\lambda_1'-\lambda_2'}\right) \\
&= \exp(\alpha\lambda_2)\exp\left(\alpha\frac{\lambda_1\lambda_2' - \lambda_1'\lambda_2}{\lambda_1'-\lambda_2'}\right).
\end{aligned}$$

Denote $c := \exp\left(\alpha\frac{\lambda_1\lambda_2'-\lambda_1'\lambda_2}{\lambda_1'-\lambda_2'}\right) \in \mathbb{R}_{>0}$. By the above calculations, $\exp(D') = c\exp(D)$ holds and consequently $\exp(\alpha\alpha' X') = c\exp(\alpha X)$. This also implies $c = \frac{\|\exp(\alpha\alpha' X')\|_F}{\|\exp(\alpha X)\|_F}$ which yields the desired equation. □

The two preceding lemmata may now be applied to analyze to what extent the minimal distance of a codebook $\mathcal{C}_0^\alpha$ depends on α.

(6.30) Corollary
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ satisfying $\lambda_1 \neq \lambda_2$ and a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$.

Then

$$\mathbb{R}_{>0} \to \mathbb{R},\ \alpha \mapsto \min_{X_1, X_2 \in \mathcal{C}_0^\alpha, X_1 \neq X_2} \Delta(X_1, X_2)$$

is continuous and strictly monotonously increasing. ◇

Proof. Choose $X_1^*, X_2^* \in \mathcal{C}_0$ according to lemma 6.28, such that for all $\alpha \in \mathbb{R}$

$$\Delta\left((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)}\right) = \min_{X_1,X_2 \in \mathcal{C}_0^\alpha, X_1 \neq X_2} \Delta(X_1, X_2).$$

It will be shown that the left hand side is continuous and strictly monotonously increasing in α.

To that end, define $\mathcal{S}_0' = \{(1,\mu_1),(0,\mu_2)\}$ and according to lemma 6.29 choose $\alpha' \in \mathbb{R}$ and $X_1', X_2' \in \mathcal{H}_{\mathcal{S}_0'}$ satisfying

$$(X_i^*)^{(\alpha)} = \frac{\rho}{\left\|\exp(\alpha X_i^*)\right\|_F} \exp(\alpha X_i^*) = \frac{\rho}{\left\|\exp((\alpha\alpha')X_i')\right\|_F} \exp((\alpha\alpha')X_i')$$

for $i = 1, 2$ and all $\alpha \in \mathbb{R}$. By proposition 6.27 the map

$$\varphi : \mathbb{R}_{>0} \to \mathbb{R},\ \alpha \mapsto \Delta\left(\frac{\rho}{\left\|\exp(\alpha X_1')\right\|_F}\exp(\alpha X_1'), \frac{\rho}{\left\|\exp(\alpha X_2')\right\|_F}\exp(\alpha X_2')\right)$$

is strictly monotonously increasing. Since $\alpha' > 0$, $l_{\alpha'} : \mathbb{R} \to \mathbb{R}, \alpha \mapsto \alpha'\alpha$ is continuous and strictly monotonously increasing as well and consequently $\varphi \circ l_{\alpha'}$ is. The latter map is identical to $\alpha \mapsto \Delta((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)})$, as for any $\alpha \in \mathbb{R}$ one obtains

$$\begin{aligned}\varphi \circ l_{\alpha'}(\alpha) &= \Delta\left(\frac{\rho}{\left\|\exp(\alpha'\alpha X_1')\right\|_F}\exp(\alpha'\alpha X_1'), \frac{\rho}{\left\|\exp(\alpha'\alpha X_2')\right\|_F}\exp(\alpha'\alpha X_2')\right)\\ &= \Delta\left(\frac{\rho}{\left\|\exp(\alpha X_1^*)\right\|_F}\exp(\alpha X_1^*), \frac{\rho}{\left\|\exp(\alpha X_2^*)\right\|_F}\exp(\alpha X_2^*)\right)\\ &= \Delta\left((X_1^*)^{(\alpha)}, (X_2^*)^{(\alpha)}\right).\end{aligned}$$

Therefore, the assertion of the corollary follows. □

The preceding corollary establishes that for a given set $\mathcal{C}_0$, the minimal distance of the codebooks $\mathcal{C}_0^\alpha$ increases strictly monotonously in α. Next, it will be shown that the limit of the codebooks $\mathcal{C}_0^\alpha$ for $\alpha \to \infty$ exists and consequently provides a codebook with strictly larger minimal distance than any of the $\mathcal{C}_0^\alpha$. The fact that the limit exists is a direct result of the following proposition.

(6.31) Proposition
Let $A \in \mathcal{H}(T)$ be a Hermitian matrix with maximal eigenvalue $\lambda_{\max}$ of multiplicity μ. Furthermore suppose $u_1, \ldots, u_T$ to be linearly independent eigenvectors of A such that $u_1, \ldots, u_\mu$ are eigenvectors with respect to $\lambda_{\max}$.

The following holds:

$$\lim_{\alpha\to\infty}\left(\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\exp(\alpha A)\right) = U\begin{pmatrix} I_\mu & 0 \\ 0 & 0 \end{pmatrix}\overline{U}^\top.$$

This is the projection matrix onto the eigenspace of A with respect to $\lambda_{\max}$. ◇

Proof. Denote by $\lambda_1, \dots, \lambda_T$ the eigenvalues of A such that $\lambda_i = \lambda_{\max}$ for $i = 1, \dots, \mu$. By lemma (2.27), an eigenvalue decomposition of $\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\exp(\alpha A)$ is given by

$$\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\exp(\alpha A) = U\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\begin{pmatrix} \exp(\alpha\lambda_1) & & \\ & \ddots & \\ & & \exp(\alpha\lambda_T) \end{pmatrix}\overline{U}^\top.$$

It will be shown that the limit of the corresponding diagonal matrix exists and is given by:

$$\lim_{\alpha\to\infty}\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\begin{pmatrix} \exp(\alpha\lambda_1) & & \\ & \ddots & \\ & & \exp(\alpha\lambda_T) \end{pmatrix} = \begin{pmatrix} I_\mu & 0 \\ 0 & 0 \end{pmatrix}.$$

The assertion of the proposition follows directly from this equality.

To check the statement made on the diagonal matrix above, the eigenvalues of $\frac{\sqrt{\mu}}{\|\exp(\alpha A)\|_F}\exp(\alpha A)$ have to be examined. They are given by $\lambda_i^{(\alpha)} := \frac{\sqrt{\mu}\exp(\alpha\lambda_i)}{\sqrt{\sum_{j=1}^T \exp(\alpha\lambda_j)^2}}$ for $i = 1, \dots, T$. In order to study their behavior for $\alpha \to \infty$, the maximal eigenvalue has to be treated seperately. Therefore, two cases are distinguished.

1) For $i \in \{\mu+1, \dots, T\}$, that is $\lambda_i < \lambda_{\max}$, one obtains

$$0 \leq \lambda_i^{(\alpha)} \leq \frac{\sqrt{\mu}\exp(\alpha\lambda_i)}{\sqrt{\sum_{j=1}^{\mu}\exp(\alpha\lambda_j)^2}} = \frac{\exp(\alpha\lambda_i)}{\exp(\alpha\lambda_{\max})} = \left(\frac{\exp(\lambda_i)}{\exp(\lambda_{\max})}\right)^\alpha.$$

The right hand side of the inequality above converges towards zero for $\alpha \to \infty$ and consequently $\lim_{\alpha\to\infty}\left(\lambda_i^{(\alpha)}\right) = 0$ can be concluded.

2) For $i \in \{1, \dots, \mu\}$, that is $\lambda_i = \lambda_{\max}$, one obtains

$$\begin{aligned}\left(\lambda_i^{(\alpha)}\right)^2 &= \frac{\mu \exp(\alpha \lambda_{\max})^2}{\sum_{j=1}^{T} \exp(\alpha \lambda_j)^2} \\ &= \frac{\mu \exp(\alpha \lambda_{\max})^2 + \sum_{k=\mu+1}^{T} \exp(\alpha \lambda_k)^2 - \sum_{k=\mu+1}^{T} \exp(\alpha \lambda_k)^2}{\sum_{j=1}^{T} \exp(\alpha \lambda_j)^2} \\ &= 1 - \sum_{k=\mu+1}^{T} \frac{\exp(\alpha \lambda_k)^2}{\sum_{j=1}^{T} \exp(\alpha \lambda_j)^2} = 1 - \frac{1}{\mu} \sum_{k=\mu+1}^{T} \left(\lambda_k^{(\alpha)}\right)^2.\end{aligned}$$

By case 1) all the summands converge toward zero as α goes towards infinity and hence $\lim_{\alpha\to\infty} \left(\lambda_i^{(\alpha)}\right) = 1$ follows. □

The proposition provides the limit for $\alpha \to \infty$ for any codeword of $\mathcal{C}_0^\alpha$. Therefore, a limit of the whole codebook $\mathcal{C}_0^\alpha$ for $\alpha \to \infty$ exists.

(6.32) Definition
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ satisfying $\lambda_1 > \lambda_2$, a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ and the signal-to-noise ratio $\rho \in \mathbb{R}_{>0}$.

Define

$$\begin{aligned}\mathcal{C}_0^\infty := &\left\{ \lim_{\alpha\to\infty} \frac{\rho}{\|\exp(\alpha X)\|_F} \exp(\alpha X) \,\middle|\, X \in \mathcal{C}_0 \right\} \\ = &\left\{ \frac{\rho}{\sqrt{\mu_1}} U \begin{pmatrix} I_{\mu_1} & 0 \\ 0 & 0 \end{pmatrix} \overline{U}^\top \,\middle|\, UD\overline{U}^\top \in \mathcal{C}_0 \right\}\end{aligned}$$ ◇

Note that codebooks of the form $\mathcal{C}_0^\infty$ correspond to the case of positive semidefinite, but not positive definite spectra which was not covered by the parametrization $\exp_{\alpha,\beta}$. Considering the codebooks $\mathcal{C}_0^\alpha$ and $\mathcal{C}_0^\infty$ covers all possible codebooks from coding spaces $\mathcal{H}_\mathcal{S}$ for a positive semidefinite spectrum $\mathcal{S}$ containing exactly two distinct eigenvalues.

It can now be formally shown that the codebook $\mathcal{C}_0^\infty$ is expected to outperform all codebooks $\mathcal{C}^\alpha$ as was conjectured in example 6.26.

(6.33) Lemma
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ satisfying $\lambda_1 > \lambda_2$ and a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$.

The code $\mathcal{C}_0^\infty$ has a larger minimal distance than all the $\mathcal{C}_0^\alpha$:

$$\min_{X_1, X_2 \in \mathcal{C}_0^\infty, X_1 \neq X_2} \Delta(X_1, X_2) > \min_{X_1, X_2 \in \mathcal{C}_0^\alpha, X_1 \neq X_2} \Delta(X_1, X_2) \text{ for all } \alpha \in \mathbb{R}.$$ ◇

Proof. Consider $X_1, X_2 \in \mathcal{C}_0$ and for $\alpha \in \mathbb{R}$ define $\beta : \mathbb{R} \to \mathbb{R}$, $\alpha \mapsto \frac{\rho}{\|\exp(\alpha X)\|_F}$ for some $X \in \mathcal{H}_{\mathcal{S}_0}$. Moreover, denote $X_1^{(\alpha)} := \beta(\alpha)\exp(\alpha X_1)$ and $X_2^{(\alpha)} := \beta(\alpha)\exp(\alpha X_2)$ such that $X_1^{(\alpha)}, X_2^{(\alpha)} \in \mathcal{C}_0^\alpha$.

The map $\mathcal{H}(T) \times \mathcal{H}(T) \to \mathbb{R}$, $(X_1, X_2) \mapsto \Delta(X_1, X_2) = \mathrm{Tr}((\Lambda_{X_2})^{-1}\Lambda_{X_1})$ is continuous since trace, matrix multiplication and matrix inversion are. Furthermore, the limit $(X_1^\infty, X_2^\infty) := \lim_{\alpha\to\infty}\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right)$ exists by proposition 6.31 and, therefore, $\lim_{\alpha\to\infty}\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right) = \Delta(X_1^\infty, X_2^\infty)$ follows.

Consequently, for any two elements $X_1^\infty, X_2^\infty \in \mathcal{C}_0^\infty$ and any $\alpha \in \mathbb{R}$, there are elements $X_1^{(\alpha)}, X_2^{(\alpha)} \in \mathcal{C}^{(\alpha)}$ such that $\Delta\left(X_1^{(\alpha)}, X_2^{(\alpha)}\right) < \Delta(X_1^\infty, X_2^\infty)$. This is in particular true for the two elements yielding the minimal distance in $\mathcal{C}_0^\infty$ which implies the assertion of the lemma. □

In example 6.26 it was seen that enlarging α did only yield codes $\mathcal{C}_0^\alpha$ that performed at most as good as optimal unitary codes. This observation will now be substantiated.

(6.34) Definition
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ satisfying $\lambda_1 > \lambda_2$, a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$ and the signal-to-noise ratio $\rho \in \mathbb{R}_{>0}$. Moreover, for any $X \in \mathcal{C}_0$ denote by U_X a unitary matrix satisfying $X = U_X D \overline{U_X}^\top$ for a real diagonal matrix D.

Define the unitary code $\mathcal{C}_0^u := \left\{ \frac{\rho}{\sqrt{\mu_1}} \overline{U_X \cdot I_{T\times\mu_1}}^\top \in \mathbb{C}^{\mu\times T} \,\middle|\, X \in \mathcal{C}_0 \right\}$. ◇

This unitary code is in fact Λ-equivalent to $\mathcal{C}_0^\infty$ which explains that the codebooks $\mathcal{C}_0^\alpha$ did perform at most equally good as an optimal unitary code in the simulation results presented in example 6.26.

(6.35) Lemma
Consider a real spectrum $\mathcal{S}_0 = \{(\lambda_1, \mu_1), (\lambda_2, \mu_2)\}$ satisfying $\lambda_1 > \lambda_2$ and a finite set $\mathcal{C}_0 \subseteq \mathcal{H}_{\mathcal{S}_0}$.

$\mathcal{C}_0^\infty$ is Λ-equivalent to $\mathcal{C}_0^u$. ◇

Proof. Denote by $\mathcal{U} = \{U_X \mid X \in \mathcal{C}_0\}$ the finite set of unitary matrices used to define $\mathcal{C}_0^u = \left\{ \frac{\rho}{\sqrt{\mu_1}} \overline{U_X \cdot I_{T\times\mu_1}}^\top \in \mathbb{C}^{\mu\times T} \,\middle|\, X \in \mathcal{C}_0 \right\}$. Then, $\mathcal{C}_0^\infty$ can be written as

$$\mathcal{C}_0^\infty = \left\{ \frac{\rho}{\sqrt{\mu_1}} U_X \begin{pmatrix} I_{\mu_1} & 0 \\ 0 & 0 \end{pmatrix} \overline{U_X}^\top \,\middle|\, U_X \in \mathcal{U} \right\}.$$

Now note that $U \begin{pmatrix} I_{\mu_1} & 0 \\ 0 & 0 \end{pmatrix} \overline{U}^{\top} = U'\overline{U'}^{\top}$ holds for any $U \in U(T)$ with $U' = U \cdot I_{T\times\mu_1}$. The matrix U' satisfies $\overline{U'}^{\top} U' = I_{\mu_1}$. This yields

$$\begin{aligned}
\Lambda_{\frac{\rho}{\sqrt{\mu_1}}U'\overline{U'}^{\top}} &= I_T + \frac{\rho^2}{\mu_1}\overline{(U'\overline{U'}^{\top})}^{\top}(U'\overline{U'}^{\top}) \\
&= I_T + \frac{\rho^2}{\mu_1}U'\overline{U'}^{\top}U'\overline{U'}^{\top} \\
&= I_T + \frac{\rho^2}{\mu_1}U'\overline{U'}^{\top}.
\end{aligned}$$

Altogether, one obtains the Λ-equivalency of $\mathcal{C}_0^{\infty}$ and $\mathcal{C}_0^{u}$:

$$\{\Lambda_X \mid X \in \mathcal{C}_0^{\infty}\} = \left\{ I_T + \frac{\rho^2}{\mu_1}U'\overline{U'}^{\top} \mid U' = U_X \cdot I_{T\times\mu_1}, U_X \in \mathcal{U} \right\} = \{\Lambda_X \mid X \in \mathcal{C}_0^{u}\}.$$

□

The results on codebooks constructed from a set $\mathcal{H}_{\mathcal{S}}$ for a fixed Spectrum $\mathcal{S}$ containing exactly two distinct eigenvalues are summarized in the following theorem.

(6.36) Theorem

Consider a noncoherent STBC $\mathcal{C} \subseteq \mathbb{C}^{T\times T}$ such that $\operatorname{spec}(X) = \{(\lambda_1, M), (\lambda_2, T-M)\}$ holds for all $X \in \mathcal{C}$ and distinct real numbers $\lambda_1, \lambda_2 > 0$.

1) There is a code $\mathcal{C}^{\infty} \subseteq \mathbb{C}^{T\times T}$ with strictly larger minimal distance with respect to the distance function Δ than $\mathcal{C}$.

2) There is a unitary code $\mathcal{C}^{u} \subseteq \mathbb{C}^{M\times T}$ which is Λ-equivalent to $\mathcal{C}^{\infty}$. ◇

In conclusion, this means that it is not possible to find codebooks which outperform unitary codebooks within the considered class of codes. To that end, in the future more general classes of codebooks should be investigated.

§ 7 Conclusion and further thoughts

Within this work a new approach to noncoherent space-time block codes has been developed. One restriction has been imposed on all considered codebooks. Namely, it has been assumed that the determinants of the matrices $\Lambda_X = I_T + \overline{X}^\top X$ are identical for all codewords X throughout the codebooks. Under this restriction, criteria to aid the design of good codebooks were formulated. Subsequently, the problem of finding good codebooks was further formalized by introducing notions of reduced and equivalent codebooks. It was found that any reduced noncoherent codebook is equivalent to a codebook which is made up of square positive semidefinite Hermitian matrices. In practice, this means that in order to find good or optimal codes, it suffices to consider positive semidefinite Hermitian matrices.

However, the problem of designing good noncoherent codebooks remains difficult. Therefore, special cases for which the design problem can be further simplified are of interest. In order to provide a systematical approach to simplify the design criteria for special cases, the notion of a distance function was introduced and studied.

The results summarized above were then applied in the fifth chapter where such a special case was considered. In particular, the class of codewords for which the matrices $\Lambda_X = I_T + \overline{X}^\top X$ for all codewords X possess the same two distinct eigenvalues of the same respective multiplicities was considered. This class has several useful properties. To begin with, it contains the class of unitary codes and it was shown that corresponding codebooks can be decoded by means of the GLRT criterion, which is also the main decoding criterion for unitary codes. Moreover, it was investigated which distance functions can be used on the considered class to design good codebooks. It was found that amongst others, the GLRT distance can be used which again can also be used for unitary codes.

In order to further study codebooks which can be obtained from the considered class, a parametrization of codewords by means of the exponential map was introduced. With the help of this parametrization it was studied how the choice of the eigenvalues of the matrices Λ_X influences the performance of corresponding codebooks. The optimal choice of the eigenvalues was determined and it was shown that a codebook which corresponds to this optimal case is equivalent to a unitary codebook. This implies that unitary codes are optimal within the new class of codebooks which was considered.

The obtained results leave some obvious questions. First, the one general assumption to restrict the determinants of the matrices $\Lambda_X = I_T + \overline{X}^\top X$ for all codewords X to be identical was made purely for simplicity. The implications of not making this assumption should be investigated. In particular, the notions of Λ-equivalence and Λ-reducedness should be studied in this more general context.

Second, also within the considered class of codewords several restrictions can be relaxed. The restriction to two distinct eigenvalues per codeword was made in order to find a simpler distance function. For codewords possessing more eigenvalues, a new approach would be required. There appears to be no systematic way of finding a simpler distance function in this case. However, one may restrict the spectra of all codewords to a finite set of distinct spectra, each containing exactly two eigenvalues. This may allow to use similar methods as the ones presented in this work to find a distance function which allows the systematic design of codebooks in this case.

§ A Simulation details

This section provides a general explanation of simulations of wireless communications using space-time block codes. Whenever a choice specific to the implementation or specific to the communications channel has to be made, it is noted how this choice was made for the simulations presented in this work. These simulations were implemented and run in matlab.

To start with, the input parameters and the output values of a simulation are listed and each is discussed briefly.

Input parameters:

- The codebook $\mathcal{C} \subseteq \mathbb{C}^{M \times T}$ itself.

  This input implicitly defines the amount of transmit antennas M to be used, the amount of timesteps T over which the channel remains invariant and the rate of the transmission $\frac{log_2(|\mathcal{C}|)}{T}$.

- The target signal-to-noise ratio $\rho \in \mathbb{R}$ in dB.

  As discussed in section 4.1, it can be assumed that the background noise at any receive antenna is gaussian distributed with variance 1 and expected value 0. However, this means that the codebook has to be rescaled so that the target SNR will be achieved. This has to be done before the start of the actual simulation.

- The amount $N \in \mathbb{N}$ of receive antennas to be used.

  In order to compare different codebooks, one has to take care that the amount of transmit antennas in respective simulations is identical. For the noncoherent case, the probability of transmission errors was computed in proposition 4.5. It is clear that the probability becomes smaller, the larger N is. The same is true for the coherent case, for an explicit computation of the corresponding error probability see for example [TSC98].

- The number of iterations $n \in \mathbb{N}$ for which the simulation shall be run.

  It would be desirable to have a concrete bound on the number of iterations which implies that the results of the simulation are within a certain range of the expected value. This is, however, complicated to obtain in general and therefore a suitable number of iterations is commonly chosen high enough, such that the results of the simulation stays about the same when run multiple times for the same input parameters.

In each iteration, a codeword is randomly picked and its transmission across a channel with SNR ρ is simulated. A received message is obtained from this simulation and this message is then decoded by a suitable decoding algorithm. For each iteration it is noted if the decoding algorithm yields the codeword that was actually transmitted.

Output values:

- The actual signal-to-noise ratio of the simulated transmissions.

  In practice, it is important to check whether the actual SNR of the simulated channel coincides with the desired target SNR. Therefore, also the actual SNR of the simulated transmission is calculated.

- The error rate that was observed.

  The main objective of a simulation is to obtain an approximation of the probability of transmission errors for a given codebook and a given communications channel. Therefore, the actual error rate that was observed during the simulation is the main result of the simulation.

Each of the steps of the simulation will be discussed in the following.

A.1 Scaling the codebook to attain the desired SNR

As a first step, the given code needs to be scaled such that the target signal-to-noise ratio will be attained. Recall that the signal-to-noise ratio of a communications channel is given by

$$\frac{E[\|HX\|_F^2]}{E[\|V\|_F^2]},$$

The fading of the channel will now be assumed to be *Rayleigh fading* as described in section 4.1 and assumed throughout this work. That is, all fading coefficients are complex gaussian distributed with expected value 0 and variance 1. Also, the additive noise is assumed to be complex gaussian distributed with expected value 0 and variance 1. According to lemma 4.2, the above expression for the SNR can then be simplified to

$$\frac{E[\|X\|_F^2]}{T}.$$

Therefore, if the codebook is used unchanged to communicate over the channel, the SNR in dB is given by

$$10 \cdot \log_{10}\left(\frac{E[\|X\|_F^2]}{T}\right) \mathrm{dB}\,.$$

All codewords shall now be scaled by a factor $\sigma \in \mathbb{R}$ such that the scaled codebook $\mathcal{C}' = \{\sigma X \mid X \in \mathcal{C}\}$ attains the target SNR ρ given in dB.

To this end, set $\sigma = \sqrt{\frac{T}{E[\|X\|_F^2]} 10^{\frac{\rho}{20}}}$. Then the SNR of the channel, if the scaled codebook $\mathcal{C}'$ is used to transmit information, is given by

$$10 \cdot \log_{10}\left(\frac{E[\|\sigma X\|_F^2]}{T}\right) \mathrm{dB} = 10 \cdot \log_{10}\left(\sigma^2 \frac{E[\|X\|_F^2]}{T}\right) \mathrm{dB} = 10 \cdot \frac{\rho}{10}\,\mathrm{dB} = \rho\,\mathrm{dB}\,.$$

The scaling factor σ depends on the expected value of $\|X\|_F^2$. Therefore, a probability distribution needs to be chosen for the codebook $\mathcal{C}$. Practically speaking, this corresponds to the probabilities of one codeword to be transmitted in the practical situation which is to be modeled. It is usually assumed that all codewords are equally likely to be transmitted. This was also assumed for the simulations presented in this work and it implies $E[\|X\|_F^2] = \frac{1}{|\mathcal{C}|}\sum_{X \in \mathcal{C}} \|X\|_F^2$.

A.2 Modeling the transmission

During one iteration of the simulation, the transmission of one codeword over the channel is modeled. In order to do so, a codeword X is chosen from the codebook with respect to a predefined random distribution. As mentioned at the end of the previous section, usually all codewords are assumed to be transmitted with equal likelihood. That is, they are subject to a uniform random distribution.

In order to simulate the transmission of the chosen codeword X, a channel matrix $H \in \mathbb{C}^{N \times M}$ and a noise matrix $V \in \mathbb{C}^{N \times T}$ are generated as matrices with complex gaussian distributed entries each with variance 1 and expected value 0. From these parameters, the matrix $Y = HX + V$ can be computed. It represents the message that is received at the receiving antennas.

A.3 Decoding

Decoding describes the process of finding a codeword $X^* \in \mathcal{C}$ that coincides with high probability with the codeword $X \in \mathcal{C}$ that has actually been sent across the channel. In the case of noncoherent coding, all information that is available to this end is the received message Y. In the coherent case, also the channel matrix H is known to the decoder.

When it comes to decoding, often a tradeoff has to be made between computational efficiency and maximizing the probability of decoding the correct codeword that was actually sent. For all simulation results presented in this work, maximum likelihood decoding was used. That is, without taking the complexity into account, the decoder chooses the codeword that was sent with maximal probability. Note that for the relevant cases, it was shown that the GLRT criterion yields maxmimum likelihood decoding. Therefore, if a matrix $Y \in \mathbb{C}^{N \times T}$ is obtained as received message, it is decoded to $\operatorname{argmin}_{X \in \mathcal{C}} \left\| \overline{X}^\top Y \right\|_F$. This was implemented by exhaustively searching through the codebook.

A.4 Calculating the output values

To obtain the actual SNR of the simulated transmissions, in every iteration the norm of the faded signal and the norm of the background noise have to be recorded. Assume during the i-th iteration for $1 \leq i \leq n$, the codeword $X_i \in \mathcal{C}$ is transmitted across the channel and the corresponding fading and noise matrices are given by $H_i \in \mathbb{C}^{N \times M}$ and $V_i \in \mathbb{C}^{N \times T}$. The actual SNR of the simulated transmissions can then be computed as

$$\frac{\sum_{i=1}^{n} \|H_i X_i\|_F^2}{\|V_i\|_F^2}.$$

In order to obtain an error rate, during each iteration the decoded codeword X^* has to be compared with the transmitted codeword X. There are two different error rates which are commonly considered.

First, one can simply compare X and X^*. If the two matrices do not coincide, a transmission error has occured. Denote the number of transmission errors which occur over the course of the n iterations of the simulation by n_{err}. The *block error rate* of the simulation is then computed as $\frac{n_{\text{err}}}{n}$. This error rate was used for all simulations presented earlier.

Secondly, some simulations provide a *bit error rate* as result. To obtain this rate, all codewords $X \in \mathcal{C}$ have to be associated with a bit sequence $s_X \in \{0,1\}^r$ for some $r \in \mathbb{N}$ and instead of directly comparing X and X^*, their corresponding bit sequences are compared. Reasonably, the identification $X \mapsto s_X$ should be injective and r as small as possible. Under the first assumption, r may be chosen as $\lceil \log_2(|\mathcal{C}|) \rceil$, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. When following this approach, instead of the amount of incorrectly decoded matrices (blocks), the amount of incorrectly decoded bits is counted. To formalize this, denote the *Hamming distance* of two bit sequences $s, t \in \{0,1\}^r$ by

$$h(s,t) := |\{i \in \mathbb{N} \mid 1 \leq i \leq r, s_i \neq t_i\}| .$$

Assume during the i-th iteration for $1 \leq i \leq n$, the codeword $X_i \in \mathcal{C}$ is transmitted across the channel and the codeword $X_i^* \in \mathcal{C}$ is decoded. The amount of bit errors occuring during the simulation is then given by $n_{\text{err}} := \sum_{i=1}^{n} h(s_{X_i}, s_{X_i^*})$. Over the n iterations, n codewords associated each to a sequence of r bits are transmitted. That is, in total rn bits are transmitted. Consequently, the bit error rate of the simulation is defined as $\frac{n_{\text{err}}}{rn}$.

It is noteworthy that the codewords may be associated with corresponding bit sequences in a way that the bit error rate may be notably smaller than the block error rate. In any case, the bit error rate is less than or equal to the block error rate.

Nomenclature

$\mathbb{N}$	the natural numbers $\{1, 2, 3, 4, \dots\}$
$\mathbb{N}_0$	the nonnegative integers $\{0, 1, 2, 3, 4, \dots\}$
$\mathbb{R}$	the field of real numbers
$\mathbb{R}_{\geq 0}$	the nonnegative real numbers $\{r \in \mathbb{R} \mid r \geq 0\}$
$\mathbb{R}_{>0}$	the strictly positive real numbers $\{r \in \mathbb{R} \mid r > 0\}$
$\mathbb{C}$	the field of complex numbers
$K^{n \times m}$	the set of $n \times m$ matrices with entries from K
$\log_a$	the real logarithm with respect to base $a \in \mathbb{R}$
ln	the natural logarithm $\log_e$
Tr	the matrix trace $\mathrm{Tr}(A) = \sum_i A_{ii}$
argmax	An argument that maximizes a function $f : M \to \mathbb{R}$. $X^* = \mathrm{argmax}_{X \in M}(f(X)) \iff f(X^*) = \max_{X \in M} f(X)$
argmin	An argument that minimizes a function $f : M \to \mathbb{R}$. $X^* = \mathrm{argmin}_{X \in M}(f(X)) \iff f(X^*) = \min_{X \in M} f(X)$

Abbreviations

dB	Decibel, page 19
SNR	Signal-to-noise ratio (of a channel), page 22
SNR	Signal-to-noise ratio (of a transmission), page 19
STBC	Space-time block code, page 21

References

[ARU01] D. Agrawal, T. J. Richardson, and R. L. Urbanke. Multiple-antenna signal constellations for fading channels. *IEEE Transactions on Information Theory*, 47(6):2618–2626, Sep 2001.

[Bha96] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 1996.

[BO13] Gregory Berhuy and Frederique Oggier. *An Introduction to Central Simple Algebras and Their Applications to Wireless Communication*. American Mathematical Society, Boston, MA, USA, 2013.

[CHS96] John H. Conway, Ronald H. Hardin, and Neil J. A. Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental Mathematics*, 5:139–159, 1996.

[Cre07] Jean Creignou. Constructions of grassmannian simplices. *CoRR*, abs/cs/0703036, 2007.

[CS98] J. Conway and N.J.A. Sloane. *Sphere Packings, Lattices and Groups*. Grundlehren der mathematischen Wissenschaften. Springer New York, 1998.

[Gut95] A. Gut. *An intermediate course in probability*. Springer Texts in Statistics. Springer New York, 1995.

[Hen05] O. Henkel. Sphere-packing bounds in the grassmann and stiefel manifolds. *IEEE Transactions on Information Theory*, 51(10):3445–3456, Oct 2005.

[HM00] B.M. Hochwald and T.L. Marzetta. Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading. *Information Theory, IEEE Transactions on*, 46(2):543–564, Mar 2000.

[Hum72] James E Humphreys. *Introduction to Lie algebras and representation theory*, volume 9. Springer Science & Business Media, 1972.

[HV05] B. Hassibi and H. Vikalo. On the sphere-decoding algorithm i. expected complexity. *Signal Processing, IEEE Transactions on*, 53(8):2806 – 2818, aug. 2005.

[Kir08] Alexander A Kirillov. *An introduction to Lie groups and Lie algebras*, volume 113. Cambridge University Press Cambridge, 2008.

[Meg98] R.E. Megginson. *An Introduction to Banach Space Theory*. Graduate Texts in Mathematics. Springer New York, 1998.

[MH99] Thomas L. Marzetta and Bertrand M. Hochwald. Capacity of a mobile multiple-antenna communication link in rayleigh flat fading. *IEEE TRANS. INFORM. THEORY*, 45:139–157, 1999.

[RCC07] Daniel J. Ryan, Iain B. Collings, and I. Vaughan L. Clarkson. Glrt-optimal noncoherent lattice decoding. *CoRR*, abs/0704.1524, 2007.

[SK09] H.R. Schwarz and N. Köckler. *Numerische Mathematik*. Vieweg Studium. Vieweg+Teubner Verlag, 2009.

[Slo] N.J.A. Sloane. Spherical codes. `http://neilsloane.com/packings/`.

[TSC98] V. Tarokh, N. Seshadri, and A.R. Calderbank. Space-time codes for high data rate wireless communication: performance criterion and code construction. *Information Theory, IEEE Transactions on*, 44(2):744 –765, mar 1998.

[UCL08] Z. Utkovski, Pi-Chin Chen, and J. Lindner. Some geometric methods for construction of space-time codes in grassmann manifolds. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 111–118, Sept 2008.

[VB99] E. Viterbo and J. Boutros. A universal lattice code decoder for fading channels. *Information Theory, IEEE Transactions on*, 45(5):1639 –1642, jul 1999.

[ZT02] Lizhong Zheng and David N. C. Tse. Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Transactions on Information Theory*, 48(2):359–383, 2002.