

# Universidad de Guadalajara

## Centro Universitario de Ciencias Exactas e Ingenierías

División de Electrónica y Computación  
Departamento de Electrónica



**“Análisis de la Eficiencia Espectral Ergódica Asintótica de  
Sistemas MIMO con Correlación de Kronecker”**

Opción: Tesis

Para Obtener el Título de  
**Licenciado en Ingeniería en Comunicaciones y Electrónica**

Que Presenta **Mario Alberto Diaz Torres**  
Director de Tesis **Dr. Víctor Manuel Pérez Abreu Carrión**

**Guadalajara, Jalisco. Julio de 2011.**





# UNIVERSIDAD DE GUADALAJARA

Centro Universitario de Ciencias Exactas e Ingenierías  
Secretaría Académica / Coordinación de la Carrera de  
Ingeniería en Comunicaciones y Electrónica

CDT/CCYE/229/10  
Código:303521958

C. Mario Alberto Díaz Torres  
Pasante de la Carrera de la Licenciatura en Ingeniería  
en Comunicaciones y Electrónica  
Presente.-

Por este conducto le damos a conocer el dictamen emitido por el Comité de Titulación de la Licenciatura en Ingeniería en Comunicaciones y Electrónica con relación a su solicitud de aprobación de modalidad y opción de titulación, conforme al Reglamento General de Titulación de la Universidad de Guadalajara:

Artículo 13. Tesis, Tesina e Informes  
Opción I. Tesis

Y conforme al reglamento particular del Centro Universitario de Ciencias Exactas e Ingenierías.

Artículo 13. Tesis, Tesina e Informes  
Opción I. Tesis

Con el Título: "Análisis de la Eficiencia Espectral Ergódica Asintótica de Sistemas MIMO con Correlación de Kronecker".

Con base en lo anterior este Comité emite el siguiente

DICTAMEN APROBADO

Que queda asentado en el acta de sesión con fecha 23 de noviembre de 2010 con el folio No. 5735/10, este comité designa al Dr. Víctor Manuel Pérez Abreu Carrión como Director de Titulación, al Dr. Rubén Sánchez Gómez como asesor al Mtro. José Benjamín Mercado Sánchez como asesor, quienes darán seguimiento para su titulación.

Con base al procedimiento académico-administrativo de titulación del CUCEI se le otorga el plazo de un año a partir de la fecha de su dictaminación para la presentación del trabajo.

En cada uno de los ejemplares del trabajo de titulación que presente para su examen profesional, deberá aparecer una copia del presente dictamen.

Atentamente  
"Piensa y Trabaja"

**"2010 Bicentenario de la Independencia y Centenario de la Revolución Mexicana"**

Guadalajara, Jal., 23 de noviembre de 2010.



*Patricia Mendoza Sánchez*  
Ing. Patricia Mendoza Sánchez  
Presidenta del Comité de Titulación de  
Ingeniería en Comunicaciones y Electrónica

COMITÉ DE TITULACIÓN  
INGENIERÍA EN COMUNICACIONES Y ELECTRÓNICA

c.c.p.- Dr. Víctor Manuel Pérez Abreu Carrión- Director de Titulación  
c.c.p.- Dr. Rubén Sánchez Gómez.-Asesor  
c.c.p.- Mtro. José Benjamín Mercado Sánchez.-Asesor  
PMS/arrf.





CUCEI

FORMA CDT-4

# División de Electrónica y Computación

Guadalajara, Jal., a 24 de Junio de 2011

**COMITÉ DE TITULACIÓN DE  
ING. EN COMUNICACIONES Y ELECTRÓNICA  
P R E S E N T E .**

Por este conducto se hace de su conocimiento que el trabajo de titulación denominado:  
Análisis de la Eficiencia Espectral Ergódica Asintótica de Sistemas MIMO con Correlación de Kronecker

que fue aprobado por ustedes mediante la solicitud No. 5735/10 con la opción específica:  
Artículo 13 Tesis, Tesina e Informes. Opción I Tesis.

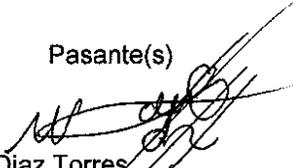
ha sido concluido, bajo la dirección de: Dr. Víctor Manuel Pérez Abreu Carrión y  
desarrollado con los siguientes puntos:

En esta tesis se expuso una breve introducción a la teoría de la información, así como la relación que tiene con la codificación y la teoría ergódica. Se expuso el estado del arte en el análisis de la eficiencia espectral ergódica asintótica de sistemas MIMO con correlación de Kronecker, utilizando resultados de la teoría de matrices aleatorias. Además, se desarrollo e implemento un método para aproximar numéricamente la eficiencia espectral ergódica asintótica de sistemas MIMO con correlación de Kronecker en presencia de una entrada isotrópica. Así mismo, se expuso el análisis numérico de los efectos de la correlación en el desempeño del sistema.

Se anexa copia del trabajo para que sea revisado por las personas que integrarán el jurado y si lo consideran conforme al reglamento , se autorice su impresión.

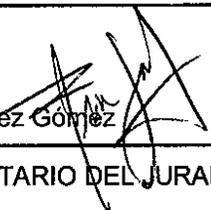
**ATENTAMENTE**

Pasante(s)

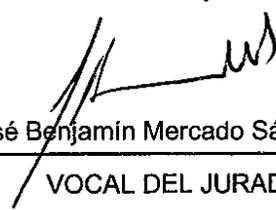
  
Mario Alberto Díaz Torres

Vo. Bo.

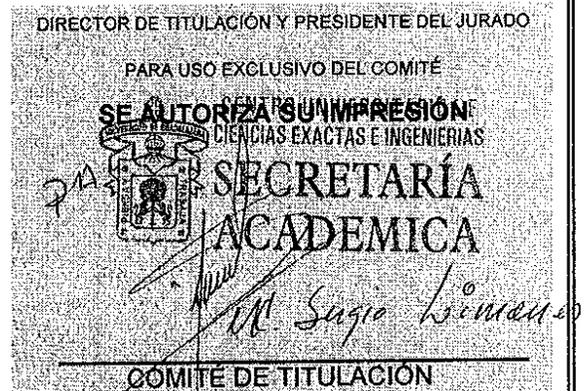
  
Dr. Víctor Manuel Pérez Abreu Carrión

  
Dr. Rubén Sánchez Gómez

SECRETARIO DEL JURADO

  
Mtro. José Benjamín Mercado Sánchez

VOCAL DEL JURADO



UTILICE ESTA FORMA HASTA QUE EL TRABAJO SEA CONCLUIDO Y REVISADO POR EL DIRECTOR DE TITULACIÓN



A mis padres, Angelina y Mario,  
a mi hermana Ana, y a todos mis  
seres queridos quienes le dan sen-  
tido a mi vida.



# Agradecimientos

Quiero agradecer a mis padres por todo su apoyo, comprensión y enseñanzas, además de ser un ejemplo para mí. En particular, les agradezco el haberme enseñado día a día que trabajando arduamente siempre se obtienen buenos resultados.

También agradezco a mi director de tesis, el Dr. Víctor Manuel Pérez Abreu Carrión, por todo el tiempo, paciencia y apoyo que me ha brindado. En especial, por todas sus enseñanzas, las cuales sobrepasan las puramente matemáticas.

Así mismo, doy las gracias a mis sinodales, el Mtro. José Benjamín Mercado Sánchez y el Dr. Rubén Sánchez Gómez, por todas sus sugerencias y comentarios.

Finalmente, agradezco a la Universidad de Guadalajara por el todo apoyo que me brindó a lo largo de mi carrera, al Centro de Investigación en Matemáticas por la beca proporcionada para realizar mi tesis de licenciatura y al Sistema Nacional de Investigadores por la beca de ayudante de investigador nacional SNI4337.

Es difícil mencionar a todas las personas e instituciones que contribuyeron de alguna manera a la realización de esta tesis, aprovecho estas líneas para brindarles mi más sincero agradecimiento.



# Índice general

<b>Introducción</b>	<b>xv</b>
<b>Notación</b>	<b>xxi</b>
<b>1. Teoría de la Información</b>	<b>1</b>
1.1. Fuente de Información . . . . .	1
1.2. Cantidad de Información . . . . .	4
1.3. Canal de Comunicación . . . . .	9
1.3.1. Canal de Ruido Aditivo Gaussiano Complejo . . . . .	13
1.4. Capacidad Ergódica . . . . .	14
1.5. Capacidad Ergódica Asintótica . . . . .	17
1.6. Notas Históricas . . . . .	19
<b>2. Modelo Canónico de los Sistemas MIMO</b>	<b>21</b>
2.1. Introducción a los Sistemas MIMO . . . . .	21
2.2. Canal MIMO Constante . . . . .	23
2.3. Modelo Canónico . . . . .	25
2.3.1. Capacidad Ergódica . . . . .	26
2.3.2. Capacidad Ergódica Asintótica . . . . .	29
2.4. Análisis Numérico de la Capacidad Ergódica . . . . .	31
2.4.1. Canal Constante . . . . .	31
2.4.2. Modelo Canónico . . . . .	32
2.4.3. Modelo Canónico Asintótico . . . . .	33
2.5. Algunas Aplicaciones de los Sistemas MIMO . . . . .	34

2.6.	Cálculo Numérico de la Capacidad Ergódica . . . . .	35
2.6.1.	Canal Constante . . . . .	35
2.6.2.	Modelo Canónico . . . . .	38
2.6.3.	Modelo Canónico Asintótico . . . . .	39
2.7.	Notas Históricas y de Frontera . . . . .	39
<b>3.</b>	<b>Sistemas MIMO con Correlación de Kronecker</b>	<b>41</b>
3.1.	Correlación de Kronecker . . . . .	42
3.2.	Correlación Unilateral . . . . .	45
3.2.1.	Receptor No Correlacionado . . . . .	45
3.2.2.	Transmisor No Correlacionado . . . . .	50
3.3.	Correlación Bilateral . . . . .	51
3.4.	Canales Relacionados . . . . .	53
3.4.1.	Canales IND . . . . .	53
3.4.2.	Canal de Rice No Correlacionado . . . . .	53
3.5.	Análisis Numérico de la Eficiencia Espectral Ergódica Isotrópica . . . . .	56
3.5.1.	Receptor No Correlacionado . . . . .	57
3.5.2.	Transmisor No Correlacionado . . . . .	60
3.5.3.	Correlación Bilateral . . . . .	63
3.6.	Cálculo Numérico de la Eficiencia Espectral Ergódica Isotrópica . . . . .	66
3.6.1.	Receptor No Correlacionado . . . . .	67
3.6.2.	Transmisor No Correlacionado . . . . .	71
3.6.3.	Correlación Bilateral . . . . .	73
3.6.4.	Canal de Rice . . . . .	76
<b>4.</b>	<b>Conclusiones</b>	<b>79</b>
<b>A.</b>	<b>Teoría Ergódica y Algunos Resultados de Procesos Estocásticos</b>	<b>81</b>
A.1.	Nociones Básicas de Probabilidad y Procesos Estocásticos . . . . .	81
A.2.	Breve Introducción a la Teoría Ergódica . . . . .	86
A.3.	Canales Ergódicos . . . . .	91

<b>B. Codificación</b>	<b>95</b>
B.1. Introducción . . . . .	95
B.2. Propiedad de Equipartición Asintótica . . . . .	96
B.3. AEP en Procesos Estocásticos . . . . .	102
B.4. Codificación en Canales Ruidosos . . . . .	105
B.5. Notas Históricas . . . . .	107
<b>C. Teoría de Matrices Aleatorias</b>	<b>109</b>
C.1. Matrices Aleatorias Gaussianas . . . . .	109
C.2. Transformada de Stieltjes . . . . .	110
C.3. Teoremas Tipo Marchenko-Pastur . . . . .	112
<b>Bibliografía</b>	<b>117</b>



# Introducción

Una innovación que ha revolucionado las comunicaciones electrónicas es la comunicación inalámbrica, en donde los sistemas MIMO<sup>1</sup> juegan un papel central en el desarrollo de nuevos sistemas. Esto requiere del uso de herramientas y modelos, clásicos y nuevos, para su comprensión y estudio. En particular, es de primordial importancia el análisis de la eficiencia espectral ergódica de dichos sistemas, lo cual nos permite estimar la cantidad de información que pasará a través un sistema MIMO previo a su implementación.

Esta tesis de licenciatura aborda el análisis de la eficiencia espectral ergódica asintótica de sistemas MIMO con correlación de Kronecker en presencia de una entrada isotrópica<sup>2</sup>, exponiendo de manera detallada los principales resultados y técnicas disponibles para ello. Dado que evaluar la eficiencia espectral ergódica no es fácil en general (véase por ejemplo el artículo de Telatar [Tel99]), en esta tesis se propone y analiza la bondad de la aproximación provista por la eficiencia espectral ergódica asintótica. Además, con el fin de calcular numéricamente la eficiencia espectral ergódica asintótica, se propone e implementa un método para aproximarla.

En 1948 el ingeniero eléctrico y matemático Claude E. Shannon publicó en el Bell System Technical Journal el artículo “A mathematical theory of communication” [Sha48], con el cual fundó la teoría de la información y en particular acuñó el concepto de capacidad de un canal<sup>3</sup>. La capacidad de un canal es una cantidad teórica que, como demostró Shannon,

---

<sup>1</sup>Un sistema MIMO (del inglés Multiple Input Multiple Output) es un sistema de comunicación inalámbrica en el cual tanto el transmisor como el receptor disponen de varias antenas para transmitir y recibir según sea el caso.

<sup>2</sup>En sistemas donde el transmisor no posee el suficiente equipo de cómputo para tener información del estado del canal, transmitir la información de manera isotrópica es en general la opción más natural.

<sup>3</sup>En el caso de canales ergódicos, para una distribución de la entrada fija, a la cantidad de información que puede pasar a través del sistema se le denomina eficiencia espectral ergódica. En este contexto, la capacidad ergódica se define como la máxima eficiencia espectral ergódica, maximizada sobre las posibles distribuciones de la entrada.

está íntimamente relacionada con la máxima cantidad de información que puede pasar a través del sistema. Desde entonces, la teoría de la información ha sido una herramienta muy importante en el desarrollo de las comunicaciones electrónicas, así como en el desarrollo de otras áreas de importancia actual como la codificación, economía, teoría de juegos, ciencias de la computación, entre otras.

Una de las características de los sistemas de comunicación inalámbrica previos a los sistemas MIMO es que las múltiples trayectorias que sigue la señal transmitida en su camino al receptor son perjudiciales para el desempeño del sistema, puesto que producen interferencia destructiva en el receptor. Sin embargo, explotando la existencia de las múltiples trayectorias, Gerard Foschini introdujo en 1996 los sistemas multiantena. Posteriormente Emre Telatar publicó en 1999 su artículo [Tel99] en el cual demostró que un sistema MIMO, bajo el llamado modelo canónico, posee una capacidad ergódica asintótica por antena receptora constante. Como consecuencia, es teóricamente posible transmitir tanta información a través de un sistema MIMO como se quiera sin ningún costo adicional de potencia, únicamente agregando más antenas transmisoras y receptoras, manteniendo una proporción constante entre ellas.

El supuesto acerca de la independencia de los coeficientes de propagación se puede garantizar bajo ciertas condiciones, tal y como lo supone el modelo canónico. Sin embargo, no siempre sucede esto, por lo cual es necesario introducir el concepto de sistema MIMO correlacionado. Es decir, un modelo para los sistemas MIMO donde hay correlación presente entre los coeficientes de propagación. De particular importancia y utilidad es el modelo de correlación separable o correlación de Kronecker, debido, entre otros factores, a su motivación teórica, validez experimental y utilidad analítica. Uno de los objetivos de esta tesis es estudiar este modelo, analizando por separado tres casos posibles: (a) el receptor no presenta correlación, (b) el transmisor no presenta correlación y (c) cuando ambos extremos presentan correlación.

Recientemente Couillet, Debbah y Silverstein en su artículo [Cou10], analizan expresiones para la capacidad asintótica de sistemas MIMO correlacionados multiusuario. Este análisis es particularmente importante puesto que contempla los efectos de la interferencia entre los usuarios del sistema, por lo que se ajusta totalmente a los canales multiaccesso MIMO. Sin embargo, en esta tesis se considera únicamente el caso en el que hay un único transmisor y un único receptor.

Actualmente existe una gran cantidad de temas relacionados con los sistemas MIMO. Esta

tesis presenta los conceptos básicos de dichos sistemas, ofreciendo una breve introducción al tema. Puesto que el desarrollo de la comunicación inalámbrica involucra varias teorías y conceptos, otro de los objetivos de esta tesis es hacer una recopilación de algunas de las teorías y conceptos más relevantes, comentando la importancia de los mismos en la comunicación inalámbrica. En particular, en los apéndices se exponen ciertos aspectos básicos de probabilidad, procesos estocásticos, teoría ergódica, codificación y teoría de matrices aleatorias.

Para calcular la eficiencia espectral ergódica asintótica de los canales MIMO es necesario utilizar resultados de la teoría de matrices aleatorias<sup>4</sup> y la transformada de Stieltjes de la distribución empírica espectral asintótica del canal. En el caso particular en que se tiene correlación unilateral, la misma que está compuesta de variables aleatorias independientes idénticamente distribuidas las cuales toman una cantidad finita de valores, la transformada de Stieltjes buscada está dada por las raíces de un polinomio. Por lo que, calcular numéricamente las raíces los polinomios respectivos nos permite obtener los valores de la transformada de Stieltjes. Mediante la implementación numérica de una fórmula de inversión es posible recuperar la densidad espectral asintótica del canal, con la cual se calcula la eficiencia espectral ergódica asintótica.

En el caso con correlación bilateral, aplicando un resultado de matrices aleatorias debido a Girko, se obtiene una ecuación de punto fijo a través de la cual se calcula la transformada de Stieltjes asintótica correspondiente. Es importante mencionar que dicho resultado es válido únicamente en el caso de distribuciones de correlación con soporte compacto. En esta tesis proponemos un método basado en discretizaciones de la función de distribución de la correlación el cual nos permite utilizar las expresiones polinomiales (en el caso con correlación unilateral) y el teorema de Girko (en el caso con correlación bilateral) para aproximar la transformada de Stieltjes correspondiente en cada caso. Con el fin de evaluar la eficacia de los programas implementados para calcular la eficiencia espectral ergódica asintótica de los canales correlacionados, se compara la eficiencia espectral ergódica asintótica cuando<sup>5</sup>  $\beta \rightarrow \infty$  con los valores calculados por el programa.

La tesis está organizada como sigue. En el Capítulo 1 se presentan los conceptos básicos de teoría de la información y el modelado de los sistemas de comunicación. Se pone especial énfasis

---

<sup>4</sup>En particular, se utilizan ciertas generalizaciones del teorema de Marchenko-Pastur (Teorema C.6).

<sup>5</sup>Se define  $\beta$  como el límite de  $n_T/n_R$  cuando  $n_R$  tiende a infinito.

en el concepto de eficiencia espectral ergódica y capacidad ergódica. Al final del capítulo, se motiva la introducción de la capacidad ergódica asintótica a través de una variación del canal de ruido aditivo Gaussiano complejo.

En el Capítulo 2 se presenta el modelo canónico de los sistemas MIMO introducido por Telatar. Se inicia el capítulo con el análisis del canal MIMO constante con el fin de ilustrar el papel que juegan los eigenvalores de la matriz de coeficientes de propagación en la eficiencia espectral ergódica del sistema. Así mismo, se muestra una comparativa entre la capacidad ergódica y la capacidad ergódica asintótica del canal MIMO bajo el modelo canónico, obteniendo como resultado particular (para un relación señal a ruido igual a cien) que la capacidad ergódica asintótica por antena receptora es una buena aproximación a la capacidad ergódica por antena receptora.

En el Capítulo 3 se presentan los sistemas MIMO correlacionados, en particular los sistemas con correlación de Kronecker, y se analiza la eficiencia espectral ergódica asintótica de dichos sistemas. Con el fin de ilustrar las técnicas propuestas para evaluar la eficiencia espectral ergódica asintótica, se presenta brevemente el análisis de otros dos canales no correlacionados: los canales independientes no idénticamente distribuidos y el canal de Rice. Por último, se presenta una comparación numérica entre la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica. Al final del capítulo se presenta la implementación del método de aproximación propuesto que se describió con anterioridad.

En el Apéndice A se proporciona un resumen de los resultados de probabilidad y procesos estocásticos que son necesarios en esta tesis. Además, se presenta una breve introducción a la teoría ergódica con el fin de motivar la suposición acerca de la ergodicidad de los canales de comunicación inalámbrica. En particular, justificamos intuitivamente que ciertos canales de comunicación inalámbrica (como procesos estocásticos) son mezclantes, lo cual implica la ergodicidad de dichos canales. Esto permite demostrar que la eficiencia espectral empírica converge casi seguramente a la eficiencia espectral ergódica, y por lo tanto es válido tomar la eficiencia espectral ergódica como medida de desempeño de un sistema de comunicación inalámbrica.

En el Apéndice B se presenta una breve introducción a la codificación. Primero se exponen los teoremas de compresión de fuentes por medio de la propiedad de equipartición asintótica (Asymptotic Equipartition Property), tanto en el caso de fuentes compuestas por variables

aleatorias independientes idénticamente distribuidas, como en el caso en el que la fuente es un proceso estocástico estacionario y ergódico. Por último, se enuncia el teorema de Shannon para canales ruidosos, el cual muestra la estrecha relación entre la capacidad ergódica teórica y la máxima cantidad de información que puede pasar a través del sistema de manera práctica.

En el Apéndice C se compilan los principales conceptos y resultados de teoría de matrices aleatorias utilizados en esta tesis, incluyendo algunos resultados sobre matrices aleatorias con entradas gaussianas y la transformada de Stieltjes de una distribución. Por último, se presenta el Teorema de Marchenko-Pastur y algunas generalizaciones acerca de la distribución empírica espectral de las matrices de correlación aleatorias bajo estudio en este trabajo.

Al final de los primeros dos capítulos y del Apéndice B se incluyen notas históricas, las cuales tienen como objetivo contextualizar desde el punto de vista histórico los resultados presentados en cada capítulo.



# Notación

Conjunto de los números naturales	$\mathbb{N} = \{1, 2, 3, \dots\}$
Conjunto de los números reales	$\mathbb{R}$
Conjunto de los números reales positivos	$\mathbb{R}^+$
Conjunto de los números complejos	$\mathbb{C}$
Conjunto de los números complejos con parte imaginaria positiva	$\mathbb{C}^+$
Parte real de un número complejo $z$	$\text{Re}(z)$
Parte imaginaria de un número complejo $z$	$\text{Im}(z)$
Cardinalidad de un conjunto $A$	$ A $
Variable aleatoria	v.a.
Variables aleatorias	v.a.s
Independientes idénticamente distribuidas	iid
Distribución gaussiana (media $\mu$ , varianza $\sigma^2$ )	$N(\mu, \sigma^2)$
Valor esperado de una v.a. $X$	$\mathbb{E}[X]$
Valor esperado de una v.a. $Y$ con respecto a una v.a. $X$	$\mathbb{E}_X[Y]$
Varianza de una v.a. $X$	$\text{var}(X)$
Covarianza de las v.a.s $X$ y $Y$ .	$\text{cov}(X, Y)$
Matriz identidad de $k \times k$	$\mathbf{I}_k$
Matriz transpuesta de una matriz $\mathbf{A}$	$\mathbf{A}^T$
Matriz transpuesta conjugada de una matriz $\mathbf{A}$	$\mathbf{A}^*$
Matriz diagonal compuesta por $a_1, \dots, a_n$	$\text{diag}(a_1, a_2, \dots, a_n)$
Traza de una matriz $\mathbf{A}$	$\text{tr}(\mathbf{A})$
Vector compuesto por la concatenación de las columnas de una matriz $\mathbf{A}$	$\text{vec}(\mathbf{A})$
Producto de Kronecker (producto tensorial)	$\otimes$
Propiedad de Equipartición Asintótica	AEP



# Capítulo 1

## Sistemas de Comunicación y Teoría de la Información

El objetivo de este capítulo es presentar el tipo de sistemas de comunicación que se analizan en los capítulos subsecuentes de esta tesis, así como introducir las propiedades de dichos sistemas en las que estamos interesados, en particular la eficiencia espectral ergódica y la capacidad ergódica. Para ello, se muestra la forma típica en la que se modelan estos sistemas, a la par que se exponen los principales conceptos y resultados de la teoría de la información<sup>1</sup>. Por último, se motiva a través de un ejemplo la definición de la capacidad ergódica asintótica.

### 1.1. Fuente de Información

La existencia de un sistema de comunicación está justificada por la necesidad de transmitir un mensaje, el cual contiene una cierta *información* o datos. Basados en este hecho podemos pensar, al menos de manera intuitiva, que una fuente de información es cualquier dispositivo u objeto que proporciona algún tipo de dato a través del tiempo.

**Ejemplo 1. Texto** Esta tesis se puede ver como una fuente de información que conforme se lee produce palabras, las cuales son la representación física de la información que queremos transmitir. Así, por ejemplo, si se lee una palabra cada segundo entonces podemos modelar

---

<sup>1</sup>Para una exposición más detallada de la Teoría de la Información remitimos al lector al libro de Cover y Thomas [Cov06].

la lectura de esta tesis como una función  $f : \mathbb{N} \rightarrow \{\text{Palabras (contenidas en la tesis)}\}$ , donde  $f(n)$  representa la palabra leída en el segundo  $n$ .

**Ejemplo 2. Voz** Podemos pensar en las variaciones de voltaje que produce la voz humana en un micrófono como una fuente de información, en la cual las variaciones de voltaje son la representación física de la información que se quiere transmitir. Una manera natural de modelar este proceso es representar las variaciones de voltaje como una función  $g : [0, \infty) \rightarrow \mathbb{R}$ , donde  $g(t)$  representa el valor del voltaje en el tiempo  $t$ .

**Ejemplo 3. Bits** Es posible pensar en los bits que trasmite un celular durante una llamada como una fuente de información en la que la información se representa mediante unos y ceros. Si se transmitieran bits solamente en los tiempos  $n\tau$  con  $n = 1, 2, 3, \dots$  ( $n \in \mathbb{N}$ ) y  $\tau$  una constante de tiempo fija, entonces la transmisión de dichos bits se podría representar como una función  $h : \mathbb{N} \rightarrow \{0, 1\}$ , donde  $h(n)$  representa el bit transmitido en el tiempo  $n\tau$ .

Si bien, es válido pensar que la fuente de información es la voz de la persona que habla por el celular, también es válido pensar que la fuente de información son los bits que se transmiten, los cuales se producen a partir de la voz. Considerar como nuestra fuente de información a la voz o a los bits dependerá meramente del aspecto particular del sistema que queramos analizar.

A priori no se puede saber exactamente qué dirá esta tesis, así como tampoco es posible saber de antemano que nos dirá una persona cuando nos llama por teléfono o que bits producirá su voz, es decir, a priori desconocemos  $f = \{f(n)|n \in \mathbb{N}\}$ ,  $g = \{g(t)|t \in [0, \infty)\}$  y  $h = \{h(n)|n \in \mathbb{N}\}$  en los tres ejemplos anteriores. Por lo tanto, es más conveniente pensar en  $f$ ,  $g$  y  $h$  como colecciones de variables aleatorias que como funciones (véanse las definiciones A.5 y A.7) ó en general como procesos estocásticos (véase la Definición A.18). En consecuencia, representaremos una fuente de información por un proceso estocástico  $X_t$  ( $t \in T$ ), cuyo espacio de estados lo denotaremos por  $E_X$ .

En los ejemplos anteriores, los espacios de estados  $E_X$  son el conjunto de palabras que conforman esta tesis, el rango de voltajes que puede tomar el micrófono y el conjunto  $\{0, 1\}$  respectivamente. A su vez, los conjuntos índices  $T$  son el conjunto  $\mathbb{N}$ , el intervalo  $[0, \infty)$  y  $\mathbb{N}$

respectivamente. Consideremos el siguiente ejemplo, el cual aparece en el artículo clásico de Shannon “A mathematical theory of communication” [Sha48].

Supóngase que se quiere modelar el idioma inglés (escrito). Denotemos por  $E$  al conjunto de todas las letras del inglés más el espacio, esto es

$$E = \{ \text{ }, a, b, c, \dots, z \}.$$

Bajo esta notación un texto en inglés es representado como un proceso estocástico  $X_t$  con espacio de estados  $E$  y conjunto índice  $\mathbb{N}$ , donde  $X_n$  representa<sup>2</sup> el caracter en la posición  $n$ . Un primer modelo consiste en tomar los caracteres de un texto como variables aleatorias independientes idénticamente distribuidas (iid), las cuales poseen una distribución uniforme en  $E$  (véase la Definición A.11). Un texto típico producido por este modelo es el siguiente: **XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD**. Observemos que este texto difícilmente puede ser considerado como una expresión en inglés, de hecho, difícilmente puede ser considerado como una expresión en cualquier idioma.

Un segundo modelo consiste en tomar los caracteres del texto como variables aleatorias iid, las cuales se distribuyen sobre el conjunto  $E$  de tal manera que la probabilidad  $\mathbb{P}(X_n = \alpha)$  con  $\alpha \in E$  es igual a la frecuencia relativa del caracter  $\alpha$  en el inglés. Un texto típico es el siguiente: **OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL**. Observemos que este texto se parece más a una expresión en inglés que el texto producido por el modelo anterior, ya que al menos las palabras tienen longitudes más similares a las reales. Sin embargo, aún dista mucho de parecer una expresión en inglés.

Dejando de lado la independencia entre los caracteres del texto, un tercer modelo consiste en modelar el texto como una cadena de Markov (véase la Definición A.19) de primer orden, en la que la probabilidad de transición  $\mathbb{P}(X_{n+1} = \beta | X_n = \alpha)$  iguala la frecuencia relativa con la que  $\beta$  aparece después de  $\alpha$  en el inglés. Un texto típico producido por esta cadena de Markov es el siguiente: **ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE**.

---

<sup>2</sup>Por comodidad, se utiliza  $X_n$  en lugar de  $X(n)$ .

Si seguimos el mismo procedimiento que en el modelo anterior, pero en lugar de tomar una cadena de Markov de primer orden tomamos una cadena de Markov de segundo orden, obtendremos textos como el siguiente: **IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.** Con este modelo se forman pequeñas palabras, por ejemplo *whhey* o *the*, pero no más allá.

Es claro que al incrementar el orden de la cadena obtendremos mejores aproximaciones. También es posible usar palabras en lugar de letras, es decir, tomar una cadena de Markov de un cierto orden donde el espacio de estados no sea las letras del inglés, sino las palabras. Claramente obtendremos mejores aproximaciones usando palabras en lugar de letras, sin embargo esto tendría un costo. Dado que en el inglés hay más palabras que letras, el orden máximo de la cadena<sup>3</sup> se reduciría drásticamente.

Esto muestra que, en general, entre más elaborado sea el proceso estocástico con el que modelamos nuestra fuente de información, mejor se ajustará éste a la fuente real. Sin embargo, entre más elaborado sea el proceso estocástico, más difícil será trabajar con él. Por lo tanto, es totalmente válido usar un modelo simplificado de la fuente, el cual sea más o menos aproximado a la fuente real en términos de la aplicación en cuestión.

## 1.2. Cantidad de Información

En la sección anterior presentamos el concepto de fuente de información, la cual intuitivamente es cualquier dispositivo que produzca algún tipo de dato. Si bien es cierto que todas las fuentes de información producen datos, no todas las fuentes de información producen la misma *cantidad de información*, tal y como lo muestra el siguiente ejemplo.

Supóngase que tenemos tres fuentes de información  $A : \mathbb{N} \rightarrow \{0, 1\}$ ,  $B : \mathbb{N} \rightarrow \{0, 1\}$  y  $C : \mathbb{N} \rightarrow \{0, 1\}$ . Supongamos además que las  $A_n$  son variables aleatorias iid con  $\mathbb{P}(A_n = 1) = 1$  y  $\mathbb{P}(A_n = 0) = 0$ , que las  $B_n$  son variables aleatorias iid con  $\mathbb{P}(B_n = 1) = 5/7$  y  $\mathbb{P}(B_n = 0) = 2/7$  y que las  $C_n$  son variables aleatorias iid con  $\mathbb{P}(C_n = 1) = 1/2$  y  $\mathbb{P}(C_n = 0) = 1/2$ . Las siguientes sucesiones son realizaciones típicas de  $A$ ,  $B$  y  $C$  respectivamente.

---

<sup>3</sup>Se puede ver que el número de probabilidades de transición es igual a  $N_e^{r+1}$  donde  $N_e$  es el número de estados y  $r$  es el orden de la cadena.

- 11111111111111111111...
- 110011110111110111...
- 1011110101000011010...

Intuitivamente la primer fuente no produce información, ya que es lo mismo observarla o no observarla, sabemos de antemano que siempre mostrará un 1. Por su parte,  $B$  y  $C$  no son predecibles, sin embargo podemos decir más de  $B$  que de  $C$ . Sabemos que es más probable ver un 1 que un 0 para cualquier  $B_n$  en particular, mientras que prácticamente no podemos decir nada acerca de las  $C_n$ , ya que es igualmente probable observar un valor que el otro.

Lo anterior nos hace pensar que la cantidad de información que produce una fuente está fuertemente relacionada con la incertidumbre o desconocimiento a priori que tenemos sobre dicha fuente. Una forma de cuantificar la cantidad de información que produce una variable aleatoria discreta es mediante la entropía de dicha variable aleatoria.

**Definición 1.1.** *Sea  $X$  una variable aleatoria discreta que toma valores  $s_i$  con probabilidades  $P(s_i) := \mathbb{P}(X = s_i)$  para  $i = \{1, 2, \dots, n\}$ . Se define la entropía de  $X$  como*

$$H(X) = - \sum_{i=1}^n P(s_i) \log P(s_i) = -\mathbb{E}[\log P(X)]. \quad (1.1)$$

En la definición anterior es usual tomar el logaritmo en base 2, en cuyo caso la unidad de medida de la entropía es el bit por símbolo (bit/símb). Si alternativamente se toma el logaritmo natural, entonces la unidad de medida correspondiente será el nat<sup>4</sup> por símbolo (nat/símb). En esta tesis utilizaremos el bit/símb a menos que se especifique lo contrario.

Podemos convencernos de la efectividad de tomar a la entropía como función que cuantifica la cantidad de información que produce una variable aleatoria de al menos dos formas. La primera, la cual se puede encontrar en el libro de Guiasu “Information Theory with Applications” [Gui77], consiste en establecer ciertos axiomas que deben ser satisfechos por cualquier función que mida la cantidad de información que produce una variable aleatoria discreta. Después, basados en estos axiomas, demostrar que la única<sup>5</sup> función que los satisface es la entropía.

---

<sup>4</sup>Observemos que  $1 \text{ nat/símb} = 1.4426 \text{ bits/símb}$ .

<sup>5</sup>Salvo por la base del logaritmo.

La otra forma, es convencerse uno mismo de la veracidad de tal afirmación al observar el papel fundamental que la entropía juega en la codificación. Siendo más específicos, hay que observar que la entropía de una variable aleatoria discreta está íntimamente relacionada con la longitud mínima a la cual se puede codificar dicha variable aleatoria (véanse los teoremas B.3 y B.5 del Apéndice B).

Para el caso de las fuentes  $A$ ,  $B$  y  $C$ , tenemos que las entropías correspondientes a las variables aleatorias  $A_n$ ,  $B_n$  y  $C_n$  ( $n \in \mathbb{N}$ ) son iguales a

$$H(A_n) = 0 \text{ bits/símb}, \quad (1.2)$$

$$H(B_n) = 0.8631 \text{ bits/símb}, \quad (1.3)$$

$$H(C_n) = 1 \text{ bit/símb}. \quad (1.4)$$

Estos valores son totalmente acordes con nuestra intuición, es decir,  $A_n$  no produce información alguna y  $B_n$  produce más información que  $A_n$  pero menos que  $C_n$ . Es posible extender la definición de entropía a variables aleatorias absolutamente continuas (véase la Definición A.8), en cuyo caso se le denomina entropía diferencial.

**Definición 1.2.** *Sea  $X$  una variable aleatoria absolutamente continua con función de densidad  $f_X$  y soporte  $S_X$ . Se define la entropía diferencial de  $X$  como*

$$H(X) = - \int_{S_X} f_X(x) \log_2 f_X(x) dx = -\mathbb{E}[\log_2 f_X(X)]. \quad (1.5)$$

Si bien, las definiciones de entropía para variables aleatorias discretas y para variables aleatorias absolutamente continuas son muy similares, hay que resaltar que la entropía de una variable aleatoria absolutamente continua en general no es el límite de entropías de variables aleatorias discretas que convergen en distribución (véase la Definición A.22) a la variable aleatoria absolutamente continua en cuestión. Por ejemplo, sea  $W_n$  una variable aleatoria discreta que toma valores  $i/n$  para  $i = \{1, \dots, n\}$  con probabilidad  $1/n$  para cada valor. En este caso  $W_n$  posee una entropía igual a

$$H(W_n) = -\frac{1}{n} \sum \log_2 \frac{1}{n} = -\log_2 \frac{1}{n} = \log_2 n. \quad (1.6)$$

De la ecuación anterior es claro que cuando  $n \rightarrow \infty$  también  $H(W_n) \rightarrow \infty$ . Sin embargo, la entropía diferencial de una variable aleatoria absolutamente continua  $W$  la cual se distribuye uniformemente en el intervalo  $[0, 1]$  está dada por

$$H(W) = - \int_0^1 \log_2 1 \, dx = 0.$$

De esto observamos que  $H(W) \neq \lim_{n \rightarrow \infty} H(W_n)$  a pesar de que<sup>6</sup>  $W_n \xrightarrow{d} W$ , tal y como se había señalado. En lo sucesivo, nos referiremos a la entropía diferencial como entropía únicamente, siempre y cuando esto no produzca confusión.

El siguiente teorema afirma que de todas las v.a. absolutamente continuas, la v.a. con distribución gaussiana (véase la Definición A.13) tiene la máxima entropía para una varianza dada.

**Teorema 1.1.** [Sha48] *Supóngase que se tiene una variable absolutamente continua  $X$  con varianza fija  $\sigma^2$ . Entonces la entropía de  $X$  satisface*

$$H(X) \leq \frac{1}{2} \ln(2\pi e \sigma^2),$$

*con igualdad si y sólo si  $X$  tiene distribución gaussiana.*

Este resultado es particularmente importante ya que, como se había mencionado, tener la máxima entropía implica tener la mayor incertidumbre posible sobre la variable aleatoria en cuestión.

Es posible generalizar el teorema anterior para el caso multivariado complejo (véase la Definición A.14), en el cual los vectores Gaussianos complejos circulares simétricos (Definición C.2) juegan el papel que la distribución gaussiana tiene en el caso univariado real. Si se tiene un vector aleatorio complejo  $n$ -dimensional  $X$  el cual tiene función de densidad  $f_X$  y soporte  $S_X \subset \mathbb{C}^n$ , entonces para calcular la entropía de  $X$  basta aplicar directamente la definición de entropía diferencial (Definición 1.2), realizando los ajustes necesarios a la integral correspondiente.

**Teorema 1.2.** [Tel99] *Supóngase que se tiene un vector aleatorio complejo  $n$ -dimensional  $X$*

---

<sup>6</sup>Denotamos por  $X_n \xrightarrow{d} Y$  al hecho de que  $X_n$  converge en distribución a  $Y$ .

con función de densidad  $f_X$  y matriz de covarianza fija  $Q$ . Entonces la entropía de  $X$  satisface

$$H(X) \leq \ln \det(\pi e Q),$$

con igualdad si y sólo si  $X$  es un vector Gaussiano complejo circular simétrico.

Este teorema tiene como consecuencia inmediata el siguiente resultado, el cual utilizaremos al final de este capítulo para calcular la capacidad del canal de ruido aditivo Gaussiano.

**Corolario 1.1.** *Supóngase que  $X$  es una variable aleatoria compleja con función de densidad  $f_X$  y varianza fija  $\sigma^2$ . Entonces la entropía de  $X$  satisface*

$$H(X) \leq \ln(\pi e \sigma^2),$$

con igualdad si y sólo si  $X$  tiene parte real e imaginaria independientes, ambas con distribución gaussiana y varianza  $\sigma^2/2$ .

Observemos que hasta el momento únicamente hemos definido la entropía para variables aleatorias, por lo tanto tenemos que extender esta definición a fuentes de información, las cuales, propiamente dicho, son procesos estocásticos.

**Definición 1.3.** [Cov06] *Sea  $X_t$  un proceso estocástico indexado por el conjunto<sup>7</sup>  $T = \mathbb{N}$ . Se define la entropía de  $X_t$  como*

$$H(X_t) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n), \quad (1.7)$$

siempre y cuando el límite exista.

La entropía del lado derecho de la ecuación (1.7) es simplemente la entropía de la v.a. cuya función de densidad es la función de densidad conjunta de las variables aleatorias  $X_1, \dots, X_n$ .

Con el fin de evitar tecnicismos que nos desvíen de nuestra discusión acerca de los elementos principales de un sistema de comunicación, a lo largo de esta tesis supondremos que las fuentes con las que trabajamos están compuestas por v.a.s iid. Un tratamiento más completo sobre entropía y procesos estocásticos puede ser consultado en [Cov06]. Bajo estas suposiciones la entropía de una fuente iguala a la entropía de la variable aleatoria que la compone.

---

<sup>7</sup>Posteriormente mostraremos por qué tomamos  $T = \mathbb{N}$ .

**Teorema 1.3.** [Cov06] *Supóngase que se tiene un proceso estocástico  $X_t$  indexado por el conjunto  $T = \mathbb{N}$ . Si las  $X_n$  ( $n \in \mathbb{N}$ ) son variables aleatorias iid con distribución común  $X$ , entonces*

$$H(X_t) = H(X).$$

Bajo nuestras suposiciones la entropía de una fuente sólo depende de la variable aleatoria  $X$  que la compone, por lo tanto, algunas veces utilizaremos el término *la fuente  $X$*  en lugar del término preciso *la fuente  $X_t$  compuesta por la variable aleatoria  $X$* .

Uno de los objetivos de la codificación es cambiar unas fuentes por otras (véase la Sección B.1 del Apéndice B). En este sentido la codificación nos permite cambiar, bajo ciertas restricciones, las fuentes originales por otras que sean analíticamente más simples o más útiles. En otras palabras, la codificación nos permite controlar hasta cierto punto la entrada del sistema, esto implica que nuestra suposición acerca de que las fuentes con las que trabajamos están compuestas por v.a. iid es válida en ciertos escenarios.

### 1.3. Canal de Comunicación

Algunas veces la información no es observada directamente de la fuente, como sucede con una llamada telefónica por ejemplo. Por lo tanto, nuestro siguiente paso es modelar el recorrido que hace la información en su camino desde la fuente hasta el destinatario.

Llamaremos *canal* a cualquier medio físico por el cual se pueda transmitir información. En general, este medio imprimirá una cierta deformación aleatoria a la señal mediante la cual se transmite la información. Por lo tanto, conviene representar al canal por un proceso estocástico  $H_t$ , el cual tiene un cierto espacio de estados  $E_H$  y por simplicidad supondremos está indexado por el mismo conjunto con el que se indexa la fuente.

En algunos casos es posible suponer que la fuente está compuesta por variables aleatorias iid sin que esto repercuta fuertemente en el desempeño del sistema (véase la Sección B.3 del Apéndice B). Sin embargo, suponer que el canal es de una forma u otra, depende exclusivamente del tipo de canal que se quiera estudiar. En la siguiente sección se discutirá este punto con mayor detalle.

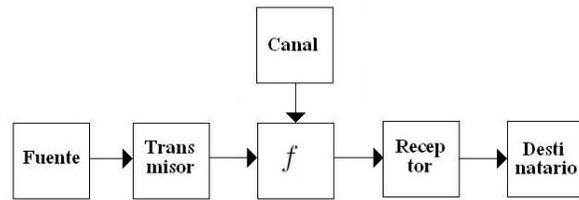
Dado que la entrada de nuestro sistema es aleatoria, así como el medio por el cual se

propaga la señal portadora de la información, en consecuencia la salida del sistema deberá ser aleatoria. Sin embargo, supondremos que para una realización de la entrada y una realización del canal dadas, la salida siempre será la misma. Por lo tanto, la salida  $Y_t$  de un sistema de comunicación deberá estar dada por

$$Y_t = f(X_t, H_t), \quad (1.8)$$

donde  $f : E_X \times E_H \rightarrow E_Y$  es una función determinista y  $E_Y$  es el conjunto de valores que puede tomar  $Y_t$ .

El proceso anterior se ilustra gráficamente en la siguiente figura.



Recordemos que las labores de codificación y decodificación se encuentran entre los bloques de la fuente y el transmisor, y entre los bloques del receptor y el destinatario respectivamente. Nuestro análisis estará enfocado únicamente en lo que pasa entre el bloque del transmisor y el bloque del receptor.

Hasta ahora hemos dicho poco sobre como son los conjuntos  $E_X$ ,  $E_H$  y  $E_Y$ , igualmente no hemos discutido sobre como es  $T$ , el conjunto índice de los procesos. Procedamos entonces a analizar estos conjuntos.

A lo largo de esta tesis se consideran exclusivamente sistemas en los cuales la información se transmite mediante señales electromagnéticas de la forma<sup>8</sup>

$$x(t) = a(t) \cos(w_c t) - b(t) \text{sen}(w_c t), \quad (1.9)$$

donde  $x : \mathbb{R} \rightarrow \mathbb{R}$  es la señal transmitida,  $w_c$  es la frecuencia angular de la señal y  $a, b : \mathbb{R} \rightarrow \mathbb{R}$  son las funciones denominadas componente en fase y componente en cuadratura respectivamente.

---

<sup>8</sup>Esta forma de señal es la forma típica utilizada en la mayoría de los sistemas de comunicación digital.

Además, supondremos que la comunicación se realiza en periodos de tiempo de longitud constante, y que durante cada periodo el transmisor mantiene constantes a  $a(t)$  y  $b(t)$ . También supondremos que el sistema es síncrono, es decir, tanto el receptor como el transmisor saben con exactitud cuando empieza y cuando termina cada periodo. Esto implica que el conjunto índice de los procesos estocásticos en cuestión será  $\mathbb{N}$ .

Si consideramos un periodo en particular y definimos  $a = a(t)$  y  $b = b(t)$ , entonces tendremos que

$$x(t) = a \cos(w_c t) - b \operatorname{sen}(w_c t) \quad (1.10)$$

$$= c \cos(w_c t + \phi), \quad (1.11)$$

donde  $c = \sqrt{a^2 + b^2}$  y  $\phi = \arctan(b/a)$ . Esto nos permite representar la señal transmitida en el periodo mediante el punto en coordenadas polares  $(c, \phi)$ , el cual está representado en el plano complejo por el punto  $a + ib$ . Por lo tanto, podemos modelar las señales transmitidas como números complejos  $X_n = a_n + ib_n$ .

Claramente la señal recibida será una versión modificada de la señal transmitida. Para modelar la señal recibida, y en consecuencia el efecto del canal en la señal transmitida, es necesario analizar primeramente el comportamiento de los canales inalámbricos.

La propiedad principal que caracteriza a los canales inalámbricos es la existencia de múltiples caminos a través de los cuales la señal transmitida llega al receptor. Comprender a fondo los mecanismos de propagación de las ondas electromagnéticas está fuera del propósito de esta tesis<sup>9</sup>. Sin embargo, podemos mencionar que los efectos dominantes en un canal de comunicación inalámbrica son tres: los desvanecimientos a gran escala (large-scale fading), los desvanecimientos a pequeña escala (small-scale fading) y el ruido aditivo.

El desvanecimiento a gran escala se refiere a las variaciones de la potencia promedio de la señal recibida en diferentes puntos geográficos, generalmente estas variaciones sólo son perceptibles a través de grandes distancias. Este fenómeno, también conocido como atenuación, se produce principalmente por las pérdidas de potencia debidas a la propagación de la señal a través del medio y por obstáculos de gran tamaño que obstruyen el paso de la señal.

---

<sup>9</sup>Para una explicación más detallada acerca de los mecanismos de propagación de las ondas electromagnéticas consúltese el libro de David Tse [Tse05].

El desvanecimiento a pequeña escala se refiere a las variaciones de la potencia recibida a través del tiempo. Estas variaciones son causadas por las múltiples copias de la señal transmitida que llegan al receptor (debido a los múltiples caminos entre el receptor y transmisor), por lo tanto, este fenómeno depende fuertemente de los objetos<sup>10</sup> que estén alrededor del transmisor y del receptor, así como el movimiento de los mismos. Estas variaciones son perceptibles en pequeños cambios geográficos, ya que dos puntos relativamente cercanos pueden tener caminos diferentes.

A lo largo de esta tesis pondremos especial énfasis en una clase de desvanecimiento conocido como desvanecimiento plano (flat fading). Este desvanecimiento es tal que la señal recibida es idéntica a la transmitida, excepto por un cambio en la amplitud y en la fase<sup>11</sup>.

Debido a que el entorno y movimiento del transmisor y receptor son en general aleatorios, el desvanecimiento a gran escala se modela estadísticamente. Como ya lo mencionamos, las señales utilizadas en los sistemas de comunicación inalámbrica son representadas como números complejos, y debido a que el desvanecimiento plano sólo modifica la amplitud y fase de la señal transmitida, entonces la señal recibida idealmente es de la forma  $\alpha x$  donde  $x$  es la señal transmitida y  $\alpha$  es el número complejo que representa la deformación producida por el canal. Es importante resaltar que el desvanecimiento a gran escala está contemplado en  $\alpha$  por medio del valor esperado de la potencia recibida  $\mathbb{E}[|\alpha x|^2]$ .

El ruido aditivo se produce generalmente por dos fenómenos, el ruido térmico (thermal noise) y el ruido de disparo (shot noise). El ruido térmico es producido por el movimiento de electrones libres en un resistor. Los electrones libres que crean este fenómeno, son electrones que se desprenden del resistor a causa de la temperatura y por lo tanto, este fenómeno aparece incluso en ausencia de corriente eléctrica a través del conductor. El ruido de disparo, por el contrario, se presenta cuando una corriente eléctrica circula a través de un dispositivo electrónico. Este ruido es generado por la variación entre la corriente promedio y la corriente instantánea, esto debido a que el número de electrones que pasan a través de un punto del circuito no es constante con respecto al tiempo.

Dado que el ruido térmico y el ruido de disparo se generan debido al efecto conjunto de una

---

<sup>10</sup>Del orden de una longitud de onda de la señal transmisora.

<sup>11</sup>El desvanecimiento plano se presenta en canales de banda angosta. En canales de banda ancha, existe un fenómeno conocido como desvanecimiento selectivo (frequency-selective fading). Para más detalles sobre las diferentes clases de desvanecimientos consúltese [Jaf05].

gran cantidad de electrones, en la mayoría de los casos estos ruidos se suponen Gaussianos. Cabe resaltar que el ruido aditivo no es único de las comunicaciones inalámbricas, dado que el ruido aditivo se produce prácticamente en cualquier proceso de conducción eléctrica, este fenómeno está presente en la mayoría de los dispositivos electrónicos.

Por lo tanto, modelaremos la señal recibida como un número complejo de la forma  $\alpha x + n$  donde  $\alpha$  es la variación producida por los desvanecimientos a gran escala y a pequeña escala, mientras que  $n$  es la variación producida por el ruido aditivo. Observemos que una realización del canal está compuesta por el par  $(\alpha, n)$  y en consecuencia los espacios de estados de los procesos estocásticos en cuestión serán  $E_X = E_Y = \mathbb{C}$  y  $E_H = \mathbb{C} \times \mathbb{C}$ .

Con el fin de hacer un uso más eficiente de la energía eléctrica, la señal transmitida deberá tener media cero<sup>12</sup>, es decir  $\mathbb{E}[X] = 0$ . Además, restringiremos la potencia promedio transmitida a algún valor  $P$ , esto es,  $\mathbb{E}[|X|^2] \leq P$ . El valor de  $P$  dependerá totalmente de la disponibilidad de energía eléctrica que tenga el sistema.

### 1.3.1. Canal de Ruido Aditivo Gaussiano Complejo

Un canal de ruido aditivo Gaussiano complejo es un canal el cual está descrito por la expresión

$$Y_n = X_n + \gamma N_n, \quad (1.12)$$

donde  $X_n \in \mathbb{C}$  es la señal transmitida en el periodo  $n$ ,  $Y_n \in \mathbb{C}$  es la señal recibida durante el mismo periodo,  $\gamma^2 \in \mathbb{R}$  es una constante denominada potencia de ruido y las  $N_n \in \mathbb{C}$  son variables aleatorias independientes gaussianas complejas de media cero y varianza unitaria, cada una con parte real y parte imaginaria iid con distribución  $N(0, 1/2)$ .

En este modelo el coeficiente que representa la modificación inducida en la señal transmitida debido al paso a través del canal es igual a 1, es decir no hay tal efecto. Por esta razón este modelo es bastante utilizado en comunicaciones alámbricas, donde la modificación debida al paso a través del canal es prácticamente constante salvo el efecto del ruido aditivo.

Es importante mencionar el porqué las variables aleatorias  $N_n$  se modelan como variables aleatorias gaussianas y no mediante otra distribución. En primer lugar, la motivación física

---

<sup>12</sup>Se puede demostrar que si  $X$  es una v.a. compleja de media cero y  $\mu$  un número complejo no nulo, entonces  $\mathbb{E}[|X + \mu|^2] > \mathbb{E}[|X|^2]$ . Por lo tanto, en términos de la potencia transmitida es óptimo que  $\mathbb{E}[X] = 0$ .

sugiere que el ruido es debido a la superposición de muchos efectos pequeños, lo cual hace natural pensar en la normalidad de la distribución estadística del ruido.

Además, recordemos que la entropía también se puede interpretar como la incertidumbre que se tiene acerca de un cierto fenómeno. Dado que la distribución gaussiana maximiza la entropía para una varianza dada, entonces esta distribución es la que nos produce la mayor incertidumbre acerca del ruido. En otras palabras, si usáramos cualquier otra distribución diferente a la gaussiana es porque se tiene evidencia que afirme la no normalidad del ruido.

## 1.4. Capacidad Ergódica

Hasta ahora se ha cuantificado la cantidad de información que produce una fuente mediante la entropía y se ha modelado la deformación que sufre la señal transmitida en su paso por el canal. El siguiente paso es investigar la cantidad de información que llega al receptor, ya que en general, ésta no será la misma que la transmitida.

Podemos pensar en la cantidad de información útil recibida como la cantidad de información recibida menos la cantidad de información o incertidumbre producida por el paso de la información a través del canal. Esto se puede pensar como si a la información recibida le quitáramos la información que generó el ruido. Por lo tanto, la información perdida se puede pensar como la incertidumbre que se tiene sobre la salida dada la entrada.

**Definición 1.4.** Sean  $X$  y  $Y$  dos variables aleatorias con densidades  $f_X$  y  $f_Y$  respectivamente. Se define la entropía condicional de  $Y$  dada  $X$  como

$$H(Y|X) = - \int_{S_X} \int_{S_Y} f(x, y) \log_2 f(y|x) dy dx, \quad (1.13)$$

donde  $S_X$  y  $S_Y$  son los soportes de las densidades  $f_X$  y  $f_Y$  respectivamente.

Por ejemplo, en el canal de ruido aditivo Gaussiano complejo se tiene que

$$H(Y_n|X_n) = H(\gamma N_n), \quad (1.14)$$

tal y como se esperaba. A continuación se define la información mutua de dos variables aleatorias, por lo expuesto en el párrafo anterior esta cantidad cuantifica la cantidad de

información que pasa a través de un sistema.

**Definición 1.5.** *Dadas dos variables aleatorias  $X$  y  $Y$ , se define la información mutua de ambas variables aleatorias como*

$$I(X; Y) = H(Y) - H(Y|X). \quad (1.15)$$

Observemos que la información mutua se puede expresar como

$$I(X; Y) = \int_{S_Y} \int_{S_X} f(x, y) \log_2 \left( \frac{f(x, y)}{f_X(x)f_Y(y)} \right) dx dy, \quad (1.16)$$

de lo cual se concluye rápidamente que  $I(X; Y) = I(Y; X)$ .

Si se tiene un sistema sin ruido, en el que  $Y = X$  entonces  $I(X; Y) = H(X) = H(Y)$  como se esperaría, ya que en este sistema no hay deformación de la información debido a la propagación a través del canal. Si por el contrario, se tuviera un canal en el que la señal recibida es independiente de la señal transmitida entonces  $I(X; Y) = 0$ , es decir, únicamente tendríamos la información del ruido en el receptor.

Es importante notar que durante un *uso* del canal la información mutua promedio por símbolo o periodo es igual a

$$\frac{1}{N} \sum_{i=1}^N I(X_i; Y_i) = \frac{1}{N} \sum_{i=1}^N I(X_i; f(X_i, H_i)), \quad (1.17)$$

donde  $N$  es el número de periodos que se utiliza el canal.

Si  $I(X_i; Y_i) = g(H_i)$  para alguna función  $g : E_H \rightarrow \mathbb{R}$ , es decir si la información mutua es una función<sup>13</sup> de  $H_i$ , entonces en el caso de canales ergódicos (véase la Sección A.3 del Apéndice A) se tiene que

$$\mathbb{P} \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N I(X_i; Y_i) = \mathbb{E}_H[g(H)] \right) = 1, \quad (1.18)$$

donde  $H$  es la variable aleatoria que tiene la misma distribución que las  $H_i$ .

---

<sup>13</sup>En el siguiente capítulo veremos que en el caso de los sistemas MIMO la información mutua del sistema es una función continua que sólo depende de  $H_i$ .

Dado que en la mayoría de sistemas convencionales  $N$  suele ser relativamente grande<sup>14</sup>, de la ecuación (1.18) se concluye que en el caso de canales ergódicos la información mutua empírica promedio  $\frac{1}{N} \sum_{i=1}^N I(X_i; Y_i)$  debe ser similar al valor esperado teórico  $\mathbb{E}_H[g(H)]$ . Con base en esta observación se define la eficiencia espectral ergódica de un sistema como la información mutua producida por la entrada y la salida dividida entre el ancho de banda utilizado. Por ejemplo, si  $X$  y  $Y$  son las variables aleatorias asociadas a la entrada y salida de un sistema respectivamente, y el ancho de banda utilizado es igual a  $B$  Hertz, entonces la eficiencia espectral ergódica<sup>15</sup> del sistema será igual a  $I(X; Y)/B$ .

Por simplicidad suponemos que el ancho de banda utilizado en nuestros sistemas es de 1Hz. Cabe resaltar que el término *eficiencia espectral* se refiere a la eficiencia con la que se utiliza el espectro electromagnético, y no tiene que ver con el espectro de una matriz.

Es importante notar que para un canal dado la eficiencia espectral ergódica depende únicamente de la distribución de la entrada, la cual es hasta cierto punto controlable mediante la codificación.

**Definición 1.6.** *Se define la capacidad ergódica de un canal como la máxima eficiencia espectral ergódica, maximizada sobre las posibles distribuciones de la entrada. En otras palabras, la capacidad ergódica  $C$  de un sistema se define como*

$$C = \max_{f_X} I(X; Y). \quad (1.19)$$

Ésta es la cantidad sobre la cual gira todo el análisis de esta tesis, sin embargo, no es el único aspecto de interés en este tipo de sistemas<sup>16</sup>. La capacidad ergódica es muy importante debido a que nos proporciona un máximo teórico para la cantidad de información que puede pasar a través del sistema (véase la Sección B.4 del Apéndice B), lo cual nos ayuda a evaluar la utilidad de un sistema antes de implementarlo. Esto es crucial en las comunicaciones, ya que una mala inversión es inaceptable en sistemas de la magnitud de una red de telefonía celular, por ejemplo.

<sup>14</sup>Por ejemplo, un sistema que transmite 1Kbps con 10 bits por símbolo utiliza 100 periodos por segundo.

<sup>15</sup>La unidad de medida de la eficiencia espectral ergódica es el bit/símb/Hz.

<sup>16</sup>En sistemas donde el canal permanece prácticamente constante a lo largo de todo un uso, la capacidad ergódica deja de ser significativa. En este tipo de canales se suele utilizar como medida de desempeño la *Capacidad de Apagón* (Outage Capacity), la cual a grandes rasgos, es una función  $C_{\text{out}} : [0, 1] \rightarrow \mathbb{R}$  donde el valor  $C_{\text{out}}(\rho)$  representa la máxima cantidad de información  $C$  con la que el  $100(1 - \rho)$  por ciento de las realizaciones del canal produce una información mutua mayor o igual a  $C$ .

Además, la capacidad ergódica nos ayuda a evaluar el desempeño de los sistemas ya implementados. Imaginemos que tenemos un sistema que alcanza una eficiencia del 99% con respecto a la capacidad ergódica (máximo teórico). En esta situación difícilmente podremos hacer mejoras que afecten significativamente el desempeño del sistema.

En resumen, calcular la capacidad ergódica es fundamental en la planeación e implementación de cualquier sistema de comunicación inalámbrica. A continuación se desarrolla el cálculo de la capacidad del canal de ruido aditivo Gaussiano complejo. Posteriormente se muestra una variación del canal de ruido aditivo Gaussiano complejo muy interesante, la cual nos permite introducir la noción de capacidad ergódica asintótica.

## 1.5. Capacidad Ergódica Asintótica

A fin de motivar la introducción de la capacidad ergódica asintótica, primero analizamos el canal de ruido aditivo Gaussiano complejo seguido del análisis de una variación de dicho sistema, la cual refleja claramente el comportamiento asintótico de la capacidad ergódica.

Supóngase que se tiene un canal de ruido aditivo Gaussiano complejo<sup>17</sup>  $Y = X + \gamma N$  tal que  $\mathbb{E}[|X|^2] = P\gamma^2$ . En este caso la capacidad del sistema está dada por

$$C = \max_{f_X} \{H(X + \gamma N) - H(X + \gamma N|X)\} \quad (1.20)$$

$$= \max_{f_X} \{H(X + \gamma N) - H(\gamma N)\}. \quad (1.21)$$

Despejando el valor de la entropía de una variable aleatoria gaussiana compleja (Corolario 1.1) obtenemos que

$$C = \max_{f_X} \{H(X + \gamma N)\} - \ln(\pi e \gamma^2). \quad (1.22)$$

Dado que la varianza de  $X + \gamma N$  es fija ( $P\gamma^2 + \gamma^2$ ), tenemos que el máximo de  $H(X + \gamma N)$  se alcanza cuando  $X + \gamma N$  es gaussiana. Dado que  $N$  es gaussiana, entonces la capacidad se

---

<sup>17</sup>Al igual que lo hicimos con las fuentes, omitimos el índice temporal del canal así como el de la salida del sistema, ya que las cantidades de interés solo dependen de las distribuciones comunes  $X$ ,  $H$  y  $Y$ .

alcanza cuando  $X$  es gaussiana. Por lo tanto la capacidad iguala a la conocida expresión

$$C = \ln(\pi e(P\gamma^2 + \gamma^2)) - \ln(\pi e\gamma^2) \quad (1.23)$$

$$= \ln(1 + P). \quad (1.24)$$

Supóngase que en lugar de transmitir toda la potencia por un mismo canal, ésta se divide en  $n$  canales idénticos al descrito en el párrafo anterior, es decir con potencia de ruido  $\gamma^2$  para cada canal. En este caso la potencia para cada canal individual será  $P\gamma^2/n$ , y la capacidad de cada uno de ellos será igual a

$$C = \ln\left(1 + \frac{P}{n}\right). \quad (1.25)$$

Dado que la capacidad total del sistema es la suma de las capacidades de los canales individuales, tendremos que la capacidad total del sistema está dada por

$$C_n = n \ln\left(1 + \frac{P}{n}\right). \quad (1.26)$$

Haciendo los cálculos correspondientes se puede mostrar que si  $n > m$  entonces  $C_n > C_m$ . Esto implica que, con la misma potencia, entre más canales individuales se tengan mayor capacidad total se tendrá. Sin embargo, el límite de la sucesión  $C_n$  cuando  $n \rightarrow \infty$  es igual a

$$C_\infty = \lim_{n \rightarrow \infty} \ln\left(1 + \frac{P}{n}\right)^n = P, \quad (1.27)$$

lo que implica que, si bien cada vez tenemos mayor capacidad, ésta nunca será mayor a  $P$ . La cantidad  $C_\infty$  es denominada capacidad ergódica asintótica y por lo general representa un máximo teórico, el cual está íntimamente relacionado con la escalabilidad del sistema.

Si bien es cierto que en general no es posible alcanzar una eficiencia espectral igual a la capacidad ergódica asintótica, el conocimiento de ésta nos permite tener una idea general del comportamiento y limitaciones del sistema. De hecho, en muchos casos, para una  $n$  razonablemente grande la capacidad  $C_n$  se comporta muy similar a  $C_\infty$ .

Por lo tanto, es deseable analizar la capacidad ergódica asintótica de cualquier sistema de comunicación que se pretenda implementar.

## 1.6. Notas Históricas

El nacimiento de la teoría de la información se dio con la publicación del artículo clásico de Claude E. Shannon “A mathematical theory of communication” publicado en el Bell System Technical Journal en 1948. Existieron algunos trabajos previos en los laboratorios Bell relacionados con el análisis de canales de comunicación, pero hasta el artículo de Shannon fue que se introdujo por primera vez el concepto de entropía<sup>18</sup> y su unidad de medición el bit. Además, en este artículo se dio por primera vez una definición formal del término *capacidad* de un canal de comunicación. Más allá de la definición teórica de capacidad, en dicho artículo se mostró la relevancia de la capacidad teórica en términos de sus implicaciones prácticas (véase el Teorema B.7 del Apéndice B).

---

<sup>18</sup>Fue la primera vez que se utilizó en el contexto de la medición de la cantidad de información, ya que la entropía ya existía en el contexto termodinámico.



## Capítulo 2

# Modelo Canónico de los Sistemas MIMO

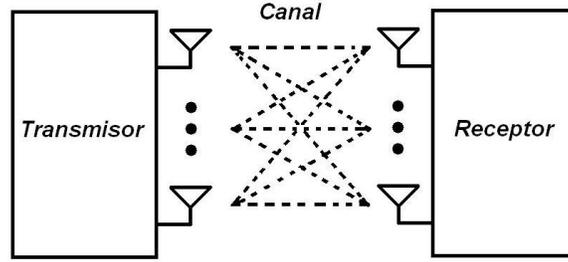
En este capítulo se presentan los sistemas MIMO, así como el modelo canónico de dichos sistemas. Con el objeto de mostrar el papel que juegan los eigenvalores de la matriz de coeficientes de propagación del canal, previo a la presentación del modelo canónico analizaremos el caso en el que la matriz de coeficientes de propagación es constante.

Una vez introducido el modelo canónico presentamos el análisis de la capacidad ergódica asintótica de los canales MIMO bajo este modelo. Por último, presentamos una comparación numérica entre la capacidad ergódica y la capacidad ergódica asintótica, seguido de la presentación de algunas aplicaciones que tienen dichos sistemas.

### 2.1. Introducción a los Sistemas MIMO

Un sistema MIMO (del inglés Multiple Input Multiple Output) es un sistema de comunicación inalámbrica en el que tanto el transmisor como el receptor cuentan con varias antenas para transmitir y recibir según sea el caso. Denotemos por  $n_R$  y  $n_T$  al número de antenas receptoras y transmisoras respectivamente.

En este tipo de sistemas todas las antenas transmisoras transmiten al mismo tiempo, y de igual manera todas las antenas receptoras reciben al mismo tiempo. Esta situación es ilustrada gráficamente en la siguiente figura.



La señal transmitida por una antena transmisora  $t \in \{1, \dots, n_T\}$ , será recibida en la antena receptora  $r \in \{1, \dots, n_R\}$  como una versión modificada de la señal transmitida. Si  $x_t$  es la señal transmitida por la antena  $t$ , entonces en la antena receptora  $r$  será recibida la señal  $H_{rt}x_t$  (véase la Sección 1.3). Se considera que tanto  $x_t$  como  $H_{rt}$  son números complejos. A los  $H_{rt}$  se les denomina coeficientes de propagación.

Cabe señalar que si  $r_1, r_2 \in \{1, \dots, n_R\}$  y  $t_1, t_2 \in \{1, \dots, n_T\}$  con  $(r_1, t_1) \neq (r_2, t_2)$  entonces el camino que sigue la señal transmitida por la antena  $t_1$  hacia la antena  $r_1$  es en general distinto al camino que sigue la señal transmitida por la antena  $t_2$  hacia la antena  $r_2$ . Por lo tanto, los coeficientes de propagación  $H_{rt}$  son en general diferentes.

Por lo tanto, para cada antena receptora  $r$  la señal recibida  $y_r$  será una combinación lineal de las señales transmitidas  $x_t$  más un ruido aditivo Gaussiano, es decir

$$y_r = \sum_{t=1}^{n_T} H_{rt}x_t + n_r. \quad (2.1)$$

Es importante recordar (véase la Sección 1.3) que el ruido presente en la ecuación anterior se debe, entre otros factores, al ruido de disparo y al ruido térmico, los cuales se generan en los dispositivos electrónicos presentes en el sistema. Debido a que los ruidos  $n_r$  se originan en diferentes componentes electrónicos (antenas, amplificadores, etc.) se supone que son v.a.s independientes<sup>1</sup>, las cuales se normalizan de tal forma que  $\text{var}(n_r) = 1$  para toda  $r \in \{1, \dots, n_R\}$ .

Si escribimos la ecuación (2.1) en notación matricial, obtenemos la siguiente expresión

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (2.2)$$

---

<sup>1</sup>Puede haber fuentes de ruido que afecten a más de una antena, como por ejemplo el ruido eléctrico que produce un rayo. Sin embargo, en general esta clase de fuentes es menos significativa.

donde  $\mathbf{x} = [x_1 \dots x_{n_T}]^T$  representa el vector de entrada del sistema,  $\mathbf{y} = [y_1 \dots y_{n_R}]^T$  es el vector de salida,  $\mathbf{H} = \{H_{rt}\}$  es la matriz de coeficientes de propagación del canal y  $\mathbf{n} = [n_1 \dots n_{n_R}]^T$  es el vector de ruido del sistema, el cual tiene matriz de covarianza  $\mathbb{E}[\mathbf{nn}^*] = \mathbf{I}_{n_R}$ .

En lo sucesivo analizaremos la capacidad ergódica de sistemas MIMO los cuales estén descritos por la ecuación (2.2). Es claro que la capacidad ergódica dependerá fuertemente de la matriz  $\mathbf{H}$ , de hecho, en la siguiente sección mostraremos que la capacidad ergódica está íntimamente relacionada con los valores singulares de la matriz de coeficientes de propagación<sup>2</sup>.

## 2.2. Canal MIMO Constante

Para ilustrar el papel que tienen los eigenvalores de la matriz  $\mathbf{HH}^*$  en el análisis de la capacidad ergódica de un sistema MIMO, estudiaremos el siguiente esquema de transmisión-recepción en el caso simplificado donde el canal permanece constante a través del tiempo.

Supongamos que se tiene un sistema MIMO, el cual tiene una matriz de coeficientes de propagación  $\mathbf{H}$  constante, es decir,  $\mathbf{H}$  es una matriz determinista. Supongamos además que tanto el transmisor como el receptor conocen la matriz  $\mathbf{H}$ . Es importante notar que en este caso la aleatoriedad en el sistema será producida únicamente por el vector de ruido Gaussiano  $\mathbf{n}$ .

La transmisión se realiza de la siguiente manera. Se codifica la información a transmitir en un vector  $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_{n_T}]^T$ , el cual debe tener una potencia menor o igual a  $P$ , esto es

$$\sum_{t=1}^{n_T} \text{var}(\tilde{x}_t) \leq P. \quad (2.3)$$

Dado que  $\mathbf{H}$  acepta una descomposición en valores singulares (SVD), podemos escribirla como

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^*, \quad (2.4)$$

donde  $\mathbf{U}$  es una matriz unitaria de  $n_R \times n_R$ ,  $\mathbf{V}$  es una matriz unitaria de  $n_T \times n_T$  y  $\mathbf{D}$  es una matriz de  $n_R \times n_T$  con  $\mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_{\min\{n_R, n_T\}})$ , donde los valores  $\sigma_i$  son los valores

---

<sup>2</sup>Recordemos que el cuadrado de los valores singulares no nulos de  $\mathbf{H}$  igualan a los eigenvalores no nulos de la matriz  $\mathbf{HH}^*$

singulares de  $\mathbf{H}$ .

Dado que el transmisor y el receptor conocen la matriz  $\mathbf{H}$ , y ésta es constante, entonces también conocen las matrices  $\mathbf{U}$  y  $\mathbf{V}$ . Por lo tanto, previo a la transmisión del vector  $\tilde{\mathbf{x}}$  éste se multiplica por  $\mathbf{V}$ , para dar paso al vector a transmitir

$$\mathbf{x} := \mathbf{V}\tilde{\mathbf{x}}. \quad (2.5)$$

Observemos que  $\text{tr}(\mathbf{x}\mathbf{x}^*) = \text{tr}(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*)$  y por lo tanto  $\sum_{t=1}^{n_T} \text{var}(x_t) \leq P$ , es decir,  $\mathbf{x}$  irradia una potencia menor o igual a  $P$ . Despejando las ecuaciones (2.4) y (2.5) en la ecuación (2.2) obtenemos que el vector recibido  $\mathbf{y}$  es igual a

$$\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{V}^*\mathbf{V}\tilde{\mathbf{x}} + \mathbf{n} = \mathbf{U}\mathbf{D}\mathbf{x} + \mathbf{n}. \quad (2.6)$$

Aplicando un procesamiento similar, el receptor multiplica el vector recibido  $\mathbf{y}$  por  $\mathbf{U}^*$ , para dar paso al nuevo vector  $\tilde{\mathbf{y}} := \mathbf{U}^*\mathbf{y}$ . Despejando en la ecuación anterior obtenemos que

$$\tilde{\mathbf{y}} = \mathbf{U}^*(\mathbf{U}\mathbf{D}\mathbf{x} + \mathbf{n}) = \mathbf{D}\mathbf{x} + \mathbf{U}^*\mathbf{n} = \mathbf{D}\mathbf{x} + \tilde{\mathbf{n}}, \quad (2.7)$$

donde  $\tilde{\mathbf{n}} := \mathbf{U}^*\mathbf{n}$ . Dado que  $\mathbf{U}$  es una matriz unitaria y  $\tilde{\mathbf{n}}$  es un vector de ruido Gaussiano entonces  $\mathbf{U}^*\mathbf{n}$  tiene la misma distribución que  $\mathbf{n}$ .

Dado que la matriz  $\mathbf{D}$  es una matriz diagonal, entonces la ecuación (2.7) es equivalente al sistema de ecuaciones<sup>3</sup>

$$\begin{aligned} \tilde{y}_1 &= \sigma_1 \tilde{x}_1 + \tilde{n}_1 \\ &\vdots \\ \tilde{y}_{n_R} &= \sigma_{n_R} \tilde{x}_{n_R} + \tilde{n}_{n_R}. \end{aligned} \quad (2.8)$$

Cada una de las ecuaciones anteriores representa un canal de ruido aditivo Gaussiano complejo, por lo tanto, si las  $\tilde{x}_r$  son v.a.s independientes entonces la capacidad ergódica de cada subcanal  $\tilde{y}_r = \sigma_r \tilde{x}_r + \tilde{n}_r$  es igual a

$$C_r = \log_2(1 + \sigma_r^2 P_r), \quad (2.9)$$

---

<sup>3</sup>Si  $n_R > n_T$  el sistema de ecuaciones tendrá únicamente  $n_T$  ecuaciones.

donde  $P_r = \mathbb{E}[|x_r|^2]$  (recordemos que  $\mathbb{E}[|\tilde{n}_r|^2] = 1$ ). En esta situación la eficiencia espectral ergódica del sistema es

$$I(\mathbf{x}; \mathbf{y}) = \sum_{r=1}^{n_R} C_r. \quad (2.10)$$

Dado que la capacidad para cada subcanal se alcanza cuando  $x_r$  es gaussiana circular simétrica, entonces la capacidad del sistema se alcanza para un vector Gaussiano complejo circular simétrico (véase el Corolario C.1). Por último, para obtener la capacidad ergódica hay que optimizar sobre los valores de las  $P_r$ , es decir

$$C = \max_{P_r: \sum_r P_r \leq P} \sum_{r=1}^{n_R} \log_2(1 + \sigma_r^2 P_r). \quad (2.11)$$

Esta optimización puede realizarse computacionalmente para cada matriz  $\mathbf{H}$  que se proporcione (véase la Sección 2.6.1).

Algo que es importante resaltar es que la capacidad de este canal MIMO constante depende totalmente de los eigenvalores de la matriz  $\mathbf{H}\mathbf{H}^*$ . De hecho, el sistema de ecuaciones (2.8) el cual relaciona una antena transmisora con una antena receptora nos sugiere que un sistema MIMO está compuesto por una especie de *cableado oculto*, el cual comunica a cada antena transmisora con una antena receptora (siempre y cuando haya suficientes antenas receptoras). En esta analogía, los eigenvalores de la matriz  $\mathbf{H}\mathbf{H}^*$  representan las características individuales de los *cables ocultos* que componen al sistema.

## 2.3. Modelo Canónico

Si bien en la sección anterior encontramos la capacidad de un sistema MIMO con matriz de coeficientes de propagación constante, cabe resaltar que esta situación es poco común en sistemas convencionales (a menos que el medio prácticamente no cambie con el paso del tiempo). Por lo tanto, es necesario introducir un modelo en el que la matriz de coeficientes de propagación sea aleatoria. En esta sección introduciremos un modelo, el cual contempla la aleatoriedad de la matriz de coeficientes de propagación.

El modelo canónico de los sistemas MIMO es un modelo básico relativamente simple para analizar la capacidad ergódica de canales MIMO en los que la matriz de coeficientes de

propagación es aleatoria. Por lo tanto, empezaremos nuestro estudio de los sistemas MIMO aleatorios con este modelo.

El modelo canónico de los sistemas MIMO supone que los coeficientes de propagación son iid con distribución gaussiana compleja circular simétrica de media cero y varianza unitaria. A este tipo de atenuación se le conoce como atenuación de Rayleigh. Es importante señalar que bajo este modelo la fase de la señal recibida tiene una distribución uniforme en el intervalo  $[0, 2\pi)$ , lo cual refleja la suposición implícita acerca de la total aleatoriedad de la posición del transmisor y del receptor, así como la aleatoriedad del entorno cercano a ellos.

En el siguiente capítulo analizaremos los sistemas MIMO bajo un modelo, el cual no supone que los coeficientes de propagación son independientes. Sin embargo, en ciertos escenarios si es posible obtener coeficientes de propagación prácticamente independientes<sup>4</sup>.

El modelo canónico supone además que el receptor conoce la matriz  $\mathbf{H}$  y que el transmisor conoce la distribución de  $\mathbf{H}$ . A esta configuración se le suele llamar CSIR/CDIT (del inglés *Channel State Information at the Receiver* y *Channel Distribution Information at the Transmitter*). Esta configuración es común en sistemas similares a la telefonía celular en el enlace de subida, donde el transmisor es un equipo móvil con un hardware limitado, y por lo tanto no puede conocer el estado del canal, mientras que el receptor es una estación base, la cual tiene el hardware suficiente para conocer el estado del canal en todo momento.

### 2.3.1. Capacidad Ergódica

Como se mencionó en la Sección 1.4, una de las cantidades fundamentales que se debe conocer de cualquier sistema de comunicación es la capacidad ergódica. El siguiente teorema proporciona una expresión para la capacidad ergódica de un sistema MIMO bajo el modelo canónico.

**Teorema 2.1.** *La capacidad del canal MIMO con matriz de coeficientes de propagación  $\mathbf{H}$  de  $n_R \times n_T$  se alcanza cuando  $\mathbf{x}$  es un vector aleatorio Gaussiano complejo circular simétrico con media cero y matriz de covarianza  $\frac{P}{n_T} \mathbf{I}_{n_T}$ . Además, en esta situación la capacidad está dada por*

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I}_{n_R} + \frac{P}{n_T} \mathbf{H} \mathbf{H}^* \right) \right]. \quad (2.12)$$

---

<sup>4</sup>En situaciones donde se tiene un ambiente rico en múltiples trayectorias, es suficiente separar las antenas media longitud de onda para obtener coeficientes estadísticamente independientes [Big07].

Con la intención de comparar el desarrollo realizado para obtener la capacidad ergódica del canal de ruido aditivo Gaussiano complejo y el desarrollo necesario para obtener la capacidad ergódica del sistema MIMO bajo el modelo canónico, bosquejamos la demostración del teorema anterior.

**Demostración. (Bosquejo)** Recordemos que la capacidad de un canal está dada por

$$C(n_R, n_T) = \max_{f_{\mathbf{x}}} I(\mathbf{x}; \mathbf{y}) = \max_{f_{\mathbf{x}}} \{H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})\}. \quad (2.13)$$

Uno de los supuestos del modelo es que el receptor conoce la matriz  $\mathbf{H}$  en todo momento, por lo tanto, la ecuación anterior puede reescribirse como

$$C(n_R, n_T) = \max_{f_{\mathbf{x}}} \{H(\mathbf{y}) - H(\mathbf{H}\mathbf{x} + \mathbf{n}|\mathbf{x})\} \quad (2.14)$$

$$= \max_{f_{\mathbf{x}}} \mathbb{E}_{\mathbf{H}} [H(\mathbf{y}) - H(\mathbf{H}\mathbf{x} + \mathbf{n}|\mathbf{x})|\mathbf{H}] \quad (2.15)$$

$$= \max_{f_{\mathbf{x}}} \mathbb{E}_{\mathbf{H}} [H(\mathbf{y})|\mathbf{H}] - H(\mathbf{n}) \quad (2.16)$$

$$= \max_{f_{\mathbf{x}}} \mathbb{E}_{\mathbf{H}} [H(\mathbf{y})|\mathbf{H}] - \log_2 \det(\pi e \mathbf{I}). \quad (2.17)$$

Observemos que para una matriz de covarianza dada  $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^*]$  y una matriz  $\mathbf{H}$  dada, el vector  $\mathbf{y}$  tiene una matriz de covarianza constante igual a

$$\mathbb{E}[\mathbf{y}\mathbf{y}^*] = \mathbb{E}[(\mathbf{H}\mathbf{x} + \mathbf{n})(\mathbf{H}\mathbf{x} + \mathbf{n})^*] \quad (2.18)$$

$$= \mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*. \quad (2.19)$$

Por el Teorema 1.2 la máxima entropía se obtiene cuando  $\mathbf{y}$  es un vector Gaussiano complejo circular simétrico, y dado que  $\mathbf{n}$  es Gaussiano complejo circular simétrico entonces  $\mathbf{x}$  también tiene que serlo. De esto se obtiene que la capacidad ergódica está dada por la expresión<sup>5</sup>

$$C(n_R, n_T) = \max_{\mathbf{Q}} \mathbb{E}_{\mathbf{H}} [\log_2 \det[\pi e(\mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*)] - \log_2 \det(\pi e \mathbf{I})] \quad (2.20)$$

$$= \max_{\mathbf{Q}} \mathbb{E}_{\mathbf{H}} [\log_2 \det(\mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*)]. \quad (2.21)$$

---

<sup>5</sup>Obsérvese que la eficiencia espectral para una matriz de covarianza fija  $\mathbf{Q}$  está dada por la expresión  $I(X; Y) = \mathbb{E}_{\mathbf{H}}[\log_2 \det(\mathbf{I}_{n_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*)]$ , la cual solamente depende de  $\mathbf{H}$ . Por lo dicho en la Sección 1.4 (ecuación 1.18) la cantidad de información transmitida durante un uso del canal será similar a la capacidad teórica en cuestión.

Usando la invarianza unitaria de la matriz  $\mathbf{H}$  (véase el Teorema C.2), se concluye que la última optimización se realiza sobre las matrices  $\mathbf{Q}$  diagonales. Utilizando la concavidad de la función  $\log_2 \det$  se llega a que

$$\log_2 \det \left( \mathbf{I} + \mathbf{H} \left( \frac{1}{n_T!} \sum_{\pi \in S_{n_T}} \pi \mathbf{Q} \pi^* \right) \mathbf{H}^* \right) \geq \frac{1}{n_T!} \sum_{\pi \in S_{n_T}} \log_2 \det (\mathbf{I} + \mathbf{H} \pi \mathbf{Q} \pi^* \mathbf{H}^*), \quad (2.22)$$

donde  $S_{n_T}$  es el conjunto de matrices de permutación de orden  $n_T$ . Dado que  $\pi$  es una matriz unitaria, la invarianza unitaria de  $\mathbf{H}$  implica que

$$\log_2 \det (\mathbf{I} + \mathbf{H} \pi \mathbf{Q} \pi^* \mathbf{H}^*) = \log_2 \det (\mathbf{I} + \mathbf{H} \mathbf{Q} \mathbf{H}^*). \quad (2.23)$$

Despejando la ecuación anterior en la ecuación (2.22) obtenemos que

$$\log_2 \det \left( \mathbf{I} + \mathbf{H} \left( \frac{1}{n_T!} \sum_{\pi \in S_{n_T}} \pi \mathbf{Q} \pi^* \right) \mathbf{H}^* \right) \geq \log_2 \det (\mathbf{I} + \mathbf{H} \mathbf{Q} \mathbf{H}^*). \quad (2.24)$$

Y puesto que  $\frac{1}{n_T!} \sum_{\pi \in S_{n_T}} \pi \mathbf{Q} \pi^* = \frac{P}{n_T} \mathbf{I}$ , el resultado se sigue. La demostración detallada se puede encontrar en el artículo de Emre Telatar “Capacity of Multi-Antenna Gaussian Channels” [Tel99].

El teorema anterior nos asegura que la capacidad se alcanza con  $\mathbf{Q} = \frac{P}{n_T} \mathbf{I}_{n_T}$ , es decir tomando independientes las señales transmitidas por las antenas del transmisor. La intuición nos indica que al no tener información del estado del canal en el transmisor, lo mejor que se puede hacer es irradiar la potencia disponible uniformemente en todas las antenas transmisoras. En este sentido, el resultado anterior corrobora dicha intuición.

Calcular la capacidad por medio del teorema anterior no es fácil en general, ya que implica utilizar los polinomios de Laguerre (véase [Tel99], Teorema 2). En la siguiente sección se presenta un método el cual nos permite aproximar la capacidad ergódica mediante la capacidad ergódica asintótica.

### 2.3.2. Capacidad Ergódica Asintótica

En la Sección 1.5 se calculó la capacidad ergódica asintótica del canal Gaussiano dividido, y nos percatamos que este canal es poco escalable<sup>6</sup>. Esta sección está dedicada a estudiar qué tan escalable es un sistema MIMO bajo el modelo canónico.

Como se mencionó en la sección anterior, calcular la capacidad de un sistema MIMO para valores dados de  $n_R$  y  $n_T$  no es fácil en general. En esta sección, además de estudiar la escalabilidad de los sistemas MIMO, encontraremos un método alternativo para aproximar la capacidad ergódica de un sistema MIMO con  $n_R$  y  $n_T$  fijos.

Dado que el determinante de toda matriz compleja iguala al producto de sus eigenvalores (Teorema C.1), entonces la ecuación (2.12) puede ser escrita como

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \prod_{r=1}^{n_R} \lambda_r \left( \mathbf{I}_{n_R} + \frac{P}{n_T} \mathbf{H}\mathbf{H}^* \right) \right], \quad (2.25)$$

donde  $\lambda_1(\mathbf{I}_{n_R} + (P/n_T)\mathbf{H}\mathbf{H}^*) \leq \dots \leq \lambda_{n_R}(\mathbf{I}_{n_R} + (P/n_T)\mathbf{H}\mathbf{H}^*)$  representan los eigenvalores ordenados<sup>7</sup> de la matriz  $\mathbf{A}$ . Observemos que si  $\mathbf{x}$  es un eigenvector de  $\mathbf{H}\mathbf{H}^*$  y  $\lambda$  el eigenvalor correspondiente entonces

$$\left( \mathbf{I}_{n_R} + \frac{P}{n_T} \mathbf{H}\mathbf{H}^* \right) \mathbf{x} = \mathbf{x} + \lambda \frac{P}{n_T} \mathbf{x} = \left( 1 + \lambda \frac{P}{n_T} \right) \mathbf{x}, \quad (2.26)$$

por lo tanto  $1 + \lambda(P/n_T)$  es un eigenvalor de  $\mathbf{I}_{n_R} + (P/n_T)\mathbf{H}\mathbf{H}^*$  y  $\mathbf{x}$  el eigenvector correspondiente, lo cual pone en relación uno a uno los eigenvalores de  $\mathbf{H}\mathbf{H}^*$  y los eigenvalores de  $\mathbf{I}_{n_R} + (P/n_T)\mathbf{H}\mathbf{H}^*$ . Utilizando esta relación en la ecuación (2.25) obtenemos que

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \prod_{r=1}^{n_R} \left( 1 + \frac{P}{n_T} \lambda_r(\mathbf{H}\mathbf{H}^*) \right) \right]. \quad (2.27)$$

Dado que el logaritmo del producto es la suma de los logaritmos, tenemos que

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \sum_{r=1}^{n_R} \log_2 \left( 1 + \frac{P}{n_T} \lambda_r(\mathbf{H}\mathbf{H}^*) \right) \right]. \quad (2.28)$$

<sup>6</sup>Por *poco escalable* nos referimos a que la cantidad de información que puede pasar a través del sistema está acotada.

<sup>7</sup>Si  $\mathbf{I}_{n_R} + (P/n_T)\mathbf{H}\mathbf{H}^*$  es una matriz Hermitiana no negativa definida, entonces los eigenvalores de  $\mathbf{A}$  son reales no negativos.

Si escribimos la última suma en términos de la distribución empírica espectral (véase la Definición C.7) tenemos que

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{P}{n_T} x \right) n_R dF^{\mathbf{H}\mathbf{H}^*}(x) \right]. \quad (2.29)$$

Dado que

$$F^{\mathbf{H}\mathbf{H}^*}(n_R x) = \frac{1}{n_R} \sum_{r=1}^{n_R} \mathbb{1}_{\{\lambda_r(\mathbf{H}\mathbf{H}^*) \leq n_R x\}} \quad (2.30)$$

$$= \frac{1}{n_R} \sum_{r=1}^{n_R} \mathbb{1}_{\{\lambda_r(\mathbf{H}\mathbf{H}^*)/n_R \leq x\}} \quad (2.31)$$

$$= \frac{1}{n_R} \sum_{r=1}^{n_R} \mathbb{1}_{\{\lambda_r(\mathbf{H}\mathbf{H}^*/n_R) \leq x\}} = F^{\mathbf{H}\mathbf{H}^*/n_R}(x), \quad (2.32)$$

la ecuación (2.29) se puede reescribir como

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{P}{n_T} n_R x \right) n_R dF^{\mathbf{H}\mathbf{H}^*}(n_R x) \right] \quad (2.33)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{P}{n_T} n_R x \right) n_R dF^{\mathbf{H}\mathbf{H}^*/n_R}(x) \right]. \quad (2.34)$$

Normalizando la capacidad por el número de antenas receptoras tenemos que

$$\frac{C(n_R, n_T)}{n_R} = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + P \frac{n_R}{n_T} x \right) dF^{\mathbf{H}\mathbf{H}^*/n_R}(x) \right]. \quad (2.35)$$

Si  $n_R \rightarrow \infty$  y  $n_T/n_R \rightarrow \beta$  aplicando el teorema de Marchenko-Pastur (véase el Teorema C.6) tenemos que

$$\frac{C(n_R, n_T)}{n_R} \xrightarrow{c.s.} \int_l^r \log_2 \left( 1 + \frac{P}{\beta} x \right) dF_\beta(x), \quad (2.36)$$

donde  $l = (1 - \sqrt{\beta})^2$ ,  $r = (1 + \sqrt{\beta})^2$  y  $F_\beta(x)$  es la distribución de Marchenko-Pastur (Definición C.8).

Observemos que el lado derecho de la ecuación (2.36) es una constante que únicamente depende de  $\beta$ . Si denotamos por  $\alpha$  a dicha constante, entonces  $C(n_R, n_T) \approx n_R \alpha$  para  $n_R$  suficientemente grande. Esto significa que el sistema es totalmente escalable, en el sentido que basta agregar más antenas transmisoras y receptoras (manteniendo la proporción  $\beta$ ) para

obtener más capacidad sin necesidad de invertir más potencia.

Esta propiedad de los sistemas MIMO es muy importante, ya que prácticamente un sistema MIMO nos da tanta capacidad como queramos sin ningún costo extra de potencia. Algo que es importante remarcar es que esta ganancia no es totalmente gratis, entre más antenas se tienen, la cantidad de procesamiento computacional necesario aumenta, y por lo tanto el costo de los equipos transmisores y receptores también aumenta.

Por otro lado, la ecuación (2.36) nos permite aproximar la capacidad de un sistema MIMO con  $n_R$  y  $n_T$  fijos mediante la aproximación

$$C(n_R, n_T) \approx n_R \int_l^r \log_2 \left( 1 + \frac{P}{\beta} x \right) dF_\beta(x), \quad (2.37)$$

donde  $\beta = n_T/n_R$ . En la siguiente sección analizaremos más a fondo la aproximación que proporciona la ecuación anterior.

## 2.4. Análisis Numérico de la Capacidad Ergódica

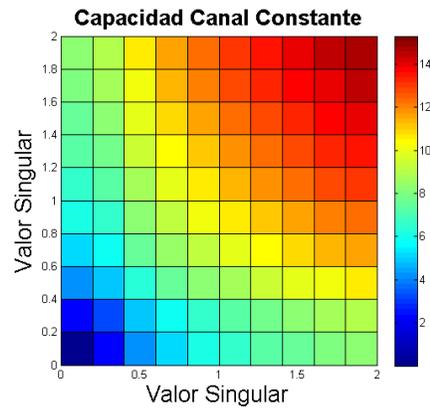
En esta sección se analizan numéricamente algunos casos particulares de las situaciones expuestas en las secciones anteriores. En particular, se enfatiza el análisis de la bondad de la aproximación de la capacidad ergódica mediante la capacidad ergódica asintótica. Para facilitar las comparaciones, todas las simulaciones de este capítulo y el siguiente se realizan para una potencia  $P = 100$ .

### 2.4.1. Canal Constante

Dado que la capacidad ergódica únicamente depende de los valores singulares de la matriz  $\mathbf{H}$  y no de su realización particular, la siguiente gráfica<sup>8</sup> muestra la capacidad para  $n_R, n_T = 2$  y valores singulares  $\sigma_1, \sigma_2 \in [0, 2]$ .

---

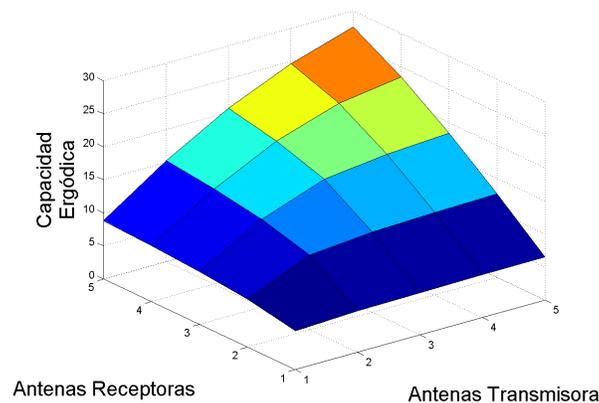
<sup>8</sup>Los valores de la gráfica se obtuvieron directamente de la ecuación (2.11) mediante la implementación de un algoritmo genético. Los detalles acerca del algoritmo así como su implementación se presentan en la Sección 2.6.1.



Esta gráfica es totalmente simétrica, tal y como se esperaba, puesto que el transmisor no tiene ningún interés particular hacia ninguna de las antenas con las que transmite. Si bien es cierto que en el caso del canal MIMO constante hay que optimizar la potencia a transmitir para la realización particular de  $\mathbf{H}$ , esta optimización se realiza únicamente una vez. Por lo tanto, una vez implementado el sistema, éste permanecerá en las mismas condiciones de operación hasta que se tenga evidencia de un cambio significativo del canal.

### 2.4.2. Modelo Canónico

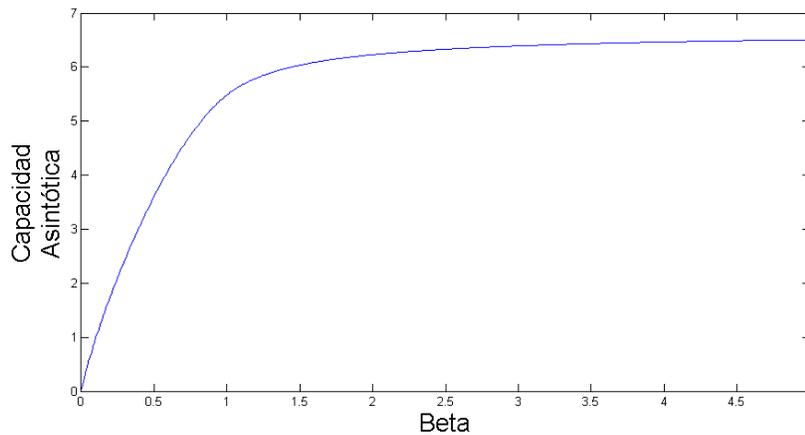
La siguiente gráfica ilustra la capacidad ergódica no asintótica para sistemas MIMO los cuales poseen entre una y cinco antenas por cada extremo (transmisor y receptor). Cada punto de la gráfica se obtuvo promediando 10,000 simulaciones, los detalles se pueden consultar en la Sección 2.6.2.



Observemos que la gráfica no es simétrica, por ejemplo, es mejor tener una antena transmisora y cinco receptoras que cinco antenas transmisoras y una receptora. Esto implica que en general los enlaces de subida y bajada de un sistema MIMO son asimétricos.

### 2.4.3. Modelo Canónico Asintótico

A continuación se presenta la gráfica de la capacidad ergódica asintótica con respecto a  $\beta$ . Dicha gráfica se obtuvo mediante la implementación directa de la ecuación (2.36) (véase la Sección 2.6.3 para más detalles acerca de la implementación).



De esta gráfica se observa que cuando  $n_R = 1$  y  $n_T \rightarrow \infty$ , la capacidad ergódica del sistema tiende a  $\log_2(1 + P)$ . Esto puede ser deducido del hecho que si  $n_R = 1$ , entonces  $(P/n_T)\mathbf{H}\mathbf{H}^* = (P/n_T)\sum_{j=1}^{n_T}|h_{1j}|^2$ . Por la ley débil de los grandes números se tiene que  $(P/n_T)\mathbf{H}\mathbf{H}^* \xrightarrow{p} P$  cuando  $n_T \rightarrow \infty$ , y aplicando el Teorema 2.1 el resultado deseado se sigue.

En la sección anterior se comentó que es posible aproximar la capacidad ergódica utilizando la capacidad ergódica asintótica vía la ecuación (2.37). La siguiente tabla muestra una comparativa de la capacidad ergódica y la capacidad ergódica asintótica, en cada casilla se encuentra la capacidad ergódica por antena receptora (obtenida por el método de la sección anterior) y la capacidad ergódica asintótica por antena receptora separadas por una diagonal.

$n_R \backslash n_T$	1	2	3	4	5
1	5.88/5.48	6.28/6.23	6.42/6.39	6.47/6.47	6.51/6.51
2	3.63/3.61	5.64/5.48	6.07/6.04	6.25/6.23	6.34/6.33
3	2.66/2.65	4.43/4.40	5.57/5.48	5.95/5.92	6.13/6.12
4	2.11/2.11	3.61/3.61	4.76/4.74	5.53/5.48	5.87/5.85
5	1.76/1.76	3.06/3.06	4.11/4.10	4.94/4.93	5.52/5.48

En la tabla anterior la aproximación menos precisa se da cuando  $n_R = n_T = 1$ , donde la capacidad ergódica presenta una diferencia porcentual del siete por ciento por arriba de la capacidad ergódica asintótica. Además, observemos que la capacidad ergódica asintótica siempre está por arriba de la capacidad ergódica. Por tanto, la tabla anterior sugiere que la capacidad ergódica puede ser aproximada satisfactoriamente por la capacidad ergódica asintótica.

## 2.5. Algunas Aplicaciones de los Sistemas MIMO

Actualmente son muchas las aplicaciones de los sistemas MIMO, entre las más conocidas se encuentran las siguientes [Sha06].

**WiFi 802.11n** La familia de estándares 802.11 de la IEEE definen un conjunto de protocolos para proveer acceso a internet mediante las llamadas WLAN (del inglés Wireless Local Area Network). Entre estos estándares se encuentra el 802.11n el cual incluye tecnología MIMO en la capa física. Este estándar es capaz de proporcionar velocidades de hasta 600Mbps.

**WiMAX 802.16e** La familia de estándares 802.16 de la IEEE, oficialmente llamada WirelessMAN y comercializada bajo el nombre de WiMAX, proporciona una serie de protocolos para proveer acceso a internet, entre los cuales el 802.16e integra el uso de tecnología MIMO en la capa física. Debido al gran alcance geográfico que tiene esta tecnología, actualmente se encuentra en desarrollo un proyecto en el Estado de Jalisco (México) el cual pretende proveer servicio de internet a comunidades de difícil acceso. Mediante el internet se prestarán servicios médicos y educativos, los cuales mejorarán la calidad de vida en esas comunidades [PWJ08].

**4G.** Entre los estándares candidatos para la cuarta generación (4G) de telefonía celular se encuentran entre otros WiFi, WiMAX y LTE Advanced. LTE Advance, al igual que WiFi y WiMAX, emplea tecnología MIMO para aumentar la cantidad de información total que pasa a través del sistema [Qua10].

Si bien la ecuación (2.36) promete un buen desempeño por parte de los sistemas MIMO, uno de los supuestos fue que los coeficientes de propagación son independientes idénticamente distribuidos. Como se verá en el siguiente capítulo, estos supuestos no siempre se tienen, pero sí es posible tener modelos en esas circunstancias.

## 2.6. Cálculo Numérico de la Capacidad Ergódica

En esta sección se recopila el software producido para realizar los cálculos numéricos presentados en este capítulo. El código fue realizado en Matlab 2007b, por lo que recomendamos verificar la compatibilidad de las instrucciones en caso de usarse una versión diferente.

### 2.6.1. Canal Constante

El cálculo de la capacidad ergódica de un canal MIMO constante se realizó mediante un algoritmo genético el cual optimiza la distribución de la potencia en la ecuación (2.11). En la Figura 2.1 se muestra el diagrama de flujo de este algoritmo.

La siguiente función evalúa la eficiencia espectral ergódica producida por un conjunto de eigenvalores  $\lambda$  y una distribución de potencia **pot**. Ésta es la función a optimizar, la cual sirve como función de aptitud (véase [Coe95]).

```

1 function varret = f(lambda,pot)
2 % Esta funcion calcula la eficiencia espectral ergodica de un canal
3 % constante con eigenvalores "lambda" y distribucion de potencia "pot".
4 varret = sum(log2(1+lambda.*pot));

```

El algoritmo de selección consiste en ordenar los individuos en orden descendiente de aptitud y tomar los **tseleccion** mejores. En este algoritmo de selección se procura no repetir individuos, lo cual normalmente no es utilizado en un algoritmo genético, pero debido a nuestros algoritmos de cruce y mutación esto es más eficiente.

Aprovechando la concavidad de la función a optimizar en el conjunto de matrices definidas positivas (véase [Tel99]) la función de cruce no se realizará aleatoriamente. El algoritmo de

cruza selecciona todas las posibles parejas (permitiendo cruzar alguien con sí mismo), sumarlas y normalizar el individuo resultante, de tal manera que la suma de sus componentes sea  $P$ . Este método de cruce asegura que los mejores individuos de la generación anterior pasen a la siguiente, es decir es un algoritmo elitista.

La mutación se realiza con una probabilidad  $p_{muta} = 0.05$ . El algoritmo de mutación consiste en tomar aleatoriamente **nsemillas** individuos alrededor del individuo a mutar<sup>9</sup> y agregarlos a la población. Para hacer más eficiente el algoritmo de mutación en etapas avanzadas del algoritmo, el radio en el cual se dispersan las semillas es proporcional a la distancia entre el mejor y el peor elemento de la generación anterior.

La población inicial siempre es muy importante, y bajo este algoritmo de cruce particular lo es aún más. Siendo coherentes con esta importancia, la parte más demandante del algoritmo (computacionalmente hablando) es elegir la población inicial. El proceso de elección de la población inicial es el siguiente. Se crea una población inicial del orden de **tinicial**, la cual se distribuye uniformemente en  $[0, P]^{n_R}$  para después ser proyectada al hiperplano  $P_1 + \dots + P_{n_R} = P$ . Esto garantiza que la población inicial está distribuida de manera prácticamente uniforme en el conjunto solución, y por lo tanto la respuesta obtenida será aproximada a la capacidad ergódica real.

La función `CapacidadCanalConstante` toma como entradas a la matriz  $\mathbf{H}$  y la potencia  $P$ , y como valores de retorno devuelve la capacidad ergódica del canal y la distribución de la potencia que alcanza dicha capacidad.

```

1 function [capacidad, distribucion_potencia] = CapacidadCanalConstante(H,P)
2 % Esta funcion aproxima la capacidad de un canal MIMO con matriz H constante
3 % y potencia disponible P.
4
5 [n_R, n_T] = size(H); % #ok< NASGU > % Obtencion de las dimensiones de H.
6

```

<sup>9</sup>La variable **nsemillas** debe su nombre a que este algoritmo de mutación es similar a la manera en que los árboles dispersan sus semillas alrededor de ellos.

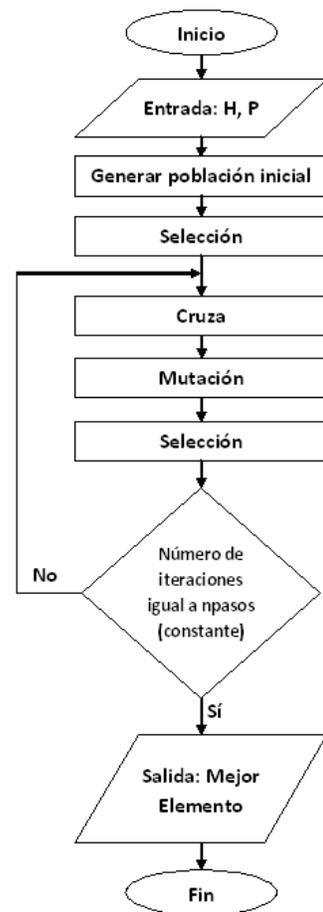


Figura 2.1: Diagrama de flujo del algoritmo genético utilizado.

```

7  tinicial = 10000; %Tama~no de la poblacion inicial.
8  tseleccion = 100; %Tama~no de la poblacion despues de la seleccion.
9  npasos = 20; %Numero de iteraciones del algoritmo.
10 pmuta = 0.05; % Probabilidad de mutacion.
11 nsemillas = 10; % Cantidad de descendientes por cada mutacion.
12
13 tpoblacion = 0; %#ok< NASGU > %Tama~no de la poblacion.
14 lambda = eig(H*H')'; % Obtencion de los eigenvalores de HH*.
15
16 % Generar la poblacion inicial. Se distribuyen puntos uniformemente en
17 % [0,P]^n_R y despues se proyectan al hiperplano P1+...+Pn_R = P.
18 G = []; % Variable en cuyas filas se almacena la poblacion.
19 base = ceil(tinicial^(1/n_R))+1; % Notese que la poblacion inicial real es mas
20 for n = 0:base^n_R-1 % grande que la variable tinicial.
21     temp = [];
22     k = n;
23     for it = 1:n_R
24         temp = [temp mod(k,base)]; % #ok< AGROW >
25         k = (k-mod(k,base))/base;
26     end
27     temp = P/(base-1)*temp;
28     if sum(temp)>0
29         temp = P*temp/sum(temp);
30     end
31     G = [G; temp f(lambda,temp)]; % #ok< AGROW >
32 end
33 tpoblacion = length(G(:,1));
34
35 % Seleccion. Se toman los mejores.
36 [X,I] = sort(G(:,n_R+1), 'descend');
37 G = G(I,:);
38 % Evitar repetidos.
39 nI = 1;
40 for it = 2:tpoblacion
41     if G(it,1:n_R) ~= G(it-1,1:n_R)
42         nI = [nI it]; % #ok< AGROW >
43     end
44 end
45 G = G(nI(1:min(length(nI),tseleccion)),:);
46 tpoblacion = length(G(:,1));
47
48 for n = 1:npasos
49     delta = norm(G(1,1:n_R)-G(tpoblacion,1:n_R)); % Radio de mutacion adaptivo.
50
51     % Cruza. Se cruzan todos con todos mediante promedio, esto para
52     % aprovechar la convexidad de la funcion a optimizar.
53     for it1 = 1:tpoblacion
54         for it2 = 1:tpoblacion
55             temp = G(it1,1:n_R)+G(it2,1:n_R);
56             if sum(temp)>0
57                 temp = P*temp/sum(temp);
58             end
59             G = [G; temp f(lambda,temp)]; % #ok< AGROW >
60         end
61     end
62
63 % Mutacion. Se generan puntos alrededor del elemento mutado.
64 for it = 1:tpoblacion
65     desicion = random('unif',0,1);
66     if (desicion < pmuta)

```

```

67         for w = 1:nsemillas
68             temp = G(it,1:n_R)+delta*random('unif',-1,1,1,n_R);
69             if sum(temp)>0
70                 temp = P*temp/sum(temp);
71             end
72             if (sum(temp>=0) == n_R)
73                 G = [G; temp f(lambda,temp)]; % #ok< AGROW >
74             end
75         end
76     end
77 end
78 tpoblacion = length(G(:,1));
79
80 % Seleccion. Se toman los mejores.
81 [X,I] = sort(G(:,n_R+1),'descend');
82 G = G(I,:);
83 % Evitar repetidos.
84 nI = 1;
85 for it = 2:tpoblacion
86     if G(it,1:n_R) ~ G(it-1,1:n_R)
87         nI = [nI it]; % #ok< AGROW >
88     end
89 end
90 G = G(nI(1:min(length(nI),tseleccion)),:);
91 tpoblacion = length(G(:,1));
92 end
93
94 % Generacion del resultado.
95 capacidad = G(1,n_R+1);
96 distribucion_potencia = G(1,1:n_R);

```

## 2.6.2. Modelo Canónico

Para calcular la capacidad ergódica del canal MIMO bajo el modelo canónico, se utilizó el siguiente procedimiento. Se simulan matrices de coeficientes de propagación, se evalúa la eficiencia espectral ergódica de cada matriz y se promedian dichas eficiencias espectrales. Por la ley débil de los grandes números (Teorema A.1) dicho promedio empírico converge (cuando el número de simulaciones va a infinito) al valor esperado teórico deseado (ecuación 2.12). Esto es lo que se conoce como método de Monte Carlo.

La función CapacidadErgodica toma como parámetros  $n_R$ ,  $n_T$ ,  $P$  y el número de simulaciones que se realizarán para estimar la capacidad ergódica.

```

1 function capacidad = CapacidadErgodica(n_R,n_T,P,nsimulaciones)
2 % Esta funcion aproxima la capacidad ergodica de un canal MIMO de n_Rxn_T
3 % bajo el modelo canonico mediante el promedio de varias simulaciones
4 % (nsimulaciones).
5
6 capacidad = 0;
7
8 for w = 1:nsimulaciones
9     % La matriz de coeficientes de propagacion.

```

```

10 H=random('norm',0,1/sqrt(2),n_R,n_T)+i*random('norm',0,1/sqrt(2),n_R,n_T);
11
12 % Obtener los eigenvalores de HH*.
13 lambda=eig(H*H');
14
15 % La eficiencia espectral ergodica obtenida para esta realizacion
16 % de H en particular.
17 capacidad_H=sum(log2(1+P/n_T*lambda));
18
19 % Agregar la eficiencia espectral ergodica obtenida para la rea-
20 % lizacion en cuestion.
21 capacidad=capacidad+capacidad_H/nsimulaciones;
22 end

```

### 2.6.3. Modelo Canónico Asintótico

Para calcular la capacidad ergódica asintótica dada por la ecuación (2.36) implementamos el método de integración numérica por rectángulos. Si bien este método es de los menos precisos, es muy eficiente y debido al grado de discretización con el que se utiliza, éste se desempeña de manera aceptable. La función `CapacidadAsintotica` toma como parámetros a  $\beta$  y a la potencia total  $P$ . La función calcula los límites  $l$  y  $r$ , y utilizando un ancho de  $\Delta = 10^{-6}$  para cada rectángulo, se aproxima la integral mediante la suma de las áreas de los rectángulos correspondientes.

```

1 function varret = CapacidadAsintotica(beta,P)
2 % Esta funcion calcula la capacidad ergodica asintotica de un canal MIMO con
3 % relacion beta y potencia P mediante una aproximacion de la integral que
4 % describe esta cantidad.
5
6 delta = 0.000001; %Tamaño del intervalo.
7
8 l = (1-sqrt(beta))^2; % Limite izquierdo.
9 r = (1+sqrt(beta))^2; % Limite derecho.
10
11 X = [1+delta:delta:r]; % #ok< NBRAK > % Evitar el cero.
12 fX = log2(1+P/beta*X).*sqrt((r-X).*(X-l))./(2*pi*X); % Aproximacion de la inte-
13 varret = delta*sum(fX); % gral.

```

## 2.7. Notas Históricas y de Frontera

En 1996 Gerard Foschini publicó su artículo “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas” [Fos96]. En él presentó una arquitectura con la cual, en el caso  $n = n_R = n_T$ , se obtienen crecimientos lineales en la capacidad ergódica respecto a  $n$ . El caso general para el modelo canónico (Teorema 2.1)

fue estudiado por Emre Telatar en su artículo “Capacity of multi-antenna Gaussian channels” [Tel99] en 1999. En este artículo se estudian tanto el caso asintótico como el no asintótico; sin embargo, es de especial interés el caso asintótico, del que se deriva la ecuación (2.37).

Desde entonces y hasta la fecha, los sistemas MIMO han cobrado importancia tanto en la industria como en la academia, prueba de ello es la gran cantidad de aplicaciones en las que se encuentran y la gran cantidad de artículos sobre el tema en revistas como la IEEE Transactions on Information Theory. Este campo sigue siendo un área activa de investigación en la que destacan los siguientes temas: comunicaciones MIMO multipunto-punto<sup>10</sup> [Cou10], comunicaciones MIMO punto-multipunto [Liu10], correlación en sistemas MIMO<sup>11</sup>[Tul05, Cou10], canal MIMO con relevo [Jin10] entre otros.

---

<sup>10</sup>En esta tesis solamente se analiza el caso punto-punto.

<sup>11</sup>Este tema es tratado en el siguiente capítulo para el caso punto-punto, sin embargo el tema sigue siendo estudiado en el caso más general.

## Capítulo 3

# Sistemas MIMO con Correlación de Kronecker

En este capítulo se presenta el análisis de la eficiencia espectral ergódica asintótica de ciertos canales MIMO cuyos coeficientes de propagación no son independientes. Primeramente se introduce el modelo de correlación de Kronecker, el cual utilizamos para calcular la eficiencia espectral ergódica asintótica de canales MIMO con correlación unilateral y bilateral.

Para encontrar la eficiencia espectral ergódica asintótica de tales sistemas MIMO correlacionados se propone un método para aproximar la transformada de Stieltjes asintótica de la distribución empírica espectral correspondiente. Con el fin de ilustrar la versatilidad del método, después de analizar la eficiencia espectral ergódica asintótica de los canales con correlación bilateral, analizamos una clase de canales independientes no idénticamente distribuidos así como el canal de Rice.

Posteriormente se compara numéricamente la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica de algunos casos particulares de los canales expuestos. Por último, se exponen los métodos desarrollados para calcular la eficiencia espectral ergódica asintótica así como su implementación en Matlab.

### 3.1. Correlación de Kronecker

En la Sección 2.3.2 del capítulo anterior encontramos que los sistemas MIMO (bajo el modelo canónico) tienen una capacidad ergódica asintótica proporcional al número de antenas receptoras, siempre y cuando se mantenga una relación constante entre el número de antenas transmisoras y receptoras. Sin embargo, uno de los supuestos principales fue que los coeficientes de propagación eran independientes e idénticamente distribuidos, lo cual no necesariamente es cierto en varias situaciones reales de interés.

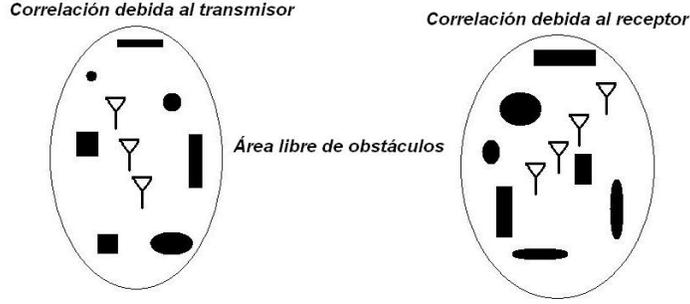
Por ejemplo, pensemos que se ponen cien antenas en un teléfono celular convencional. Debido a las dimensiones físicas de un teléfono celular, las antenas prácticamente estarían en el mismo lugar y en esta situación difícilmente podemos pensar que los coeficientes de propagación son independientes. En consecuencia, es necesario modelar la correlación entre los coeficientes de propagación<sup>1</sup> de un sistemas MIMO a fin de encontrar expresiones más precisas para la capacidad ergódica del sistema.

Denotaremos por  $\mathbf{R}_h = \mathbb{E}[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^*]$  a la matriz de covarianza de los coeficientes de propagación del canal. Por lo tanto, para modelar la correlación en su forma más general sería necesario tomar en cuenta  $(n_R n_T)^2$  coeficientes de correlación ( $\text{vec}(\mathbf{H})$  es un vector de dimensión  $n_R n_T$ ). Si bien utilizar todos estos coeficientes de correlación es bastante preciso, es al mismo tiempo poco manejable. Con el fin de introducir un modelos más simple, pero al mismo tiempo útil, consideremos la siguiente situación:

Supóngase que se entabla comunicación entre un celular y una radio base (ambos multiantena). La trayectoria de la señal portadora de la información, la cual supondremos va del celular hacia la radio base, puede ser dividida en tres etapas. En la primer etapa, justo después de su transmisión, la señal debe salir del entorno cercano al celular, el cual podría estar compuesto, por ejemplo, por el mueble e inmueble de un edificio. Después de salir del entorno cercano al transmisor, pasaría a una etapa de movimiento prácticamente libre. Por último, la señal volvería a interactuar con ciertos objetos, pero esta vez en el entorno cercano al receptor. Esta situación se ilustra gráficamente en la siguiente figura.

---

<sup>1</sup>Por correlación entre dos coeficientes de propagación nos referimos a la covarianza de ambos coeficientes vistos como variables aleatorias. Si bien es cierto que la correlación y la covarianza de dos variables aleatorias son diferentes (véanse las definiciones A.15 y A.16), aquí nos referimos por correlación al hecho de que los coeficientes no sean independientes.



Por lo tanto, podemos pensar que la correlación entre dos coeficientes de propagación depende de las antenas con las que se transmite y las antenas con las que se recibe de manera independiente. Es decir, si se transmite con las antenas  $j$  y  $l$ , y se recibe con las antenas  $i$  y  $k$ , entonces el coeficiente de propagación correspondientes será igual a  $C_r(i, k)C_t(j, l)$ , donde  $C_r$  y  $C_t$  son las funciones que modelan la *correlación* entre las antenas receptoras y transmisoras correspondientes. A este modelo se le llama modelo de correlación de Kronecker o también modelo de correlación separable<sup>2</sup>. Las suposiciones hechas por el modelo de Kronecker implican que la matriz  $\mathbf{R}_h$  tiene la forma

$$\mathbf{R}_h = \mathbf{R}_t^T \otimes \mathbf{R}_r, \quad (3.1)$$

donde  $\mathbf{R}_t$  y  $\mathbf{R}_r$  son matrices de covarianza de dimensión  $n_T$  y  $n_R$  respectivamente, las cuales modelan la correlación entre las antenas transmisoras y receptoras respectivamente. El símbolo  $\otimes$  denota el producto de Kronecker o producto tensorial (véase la Definición A.17).

Denotemos por  $\mathbf{H}_w$  a la matriz con coeficientes Gaussianos complejos circulares simétricos iid con media cero y varianza unitaria. Si suponemos que los coeficientes de la matriz que modela el canal son v.a.s gaussianas, entonces la matriz  $\mathbf{H}$  definida por  $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}$  tiene matriz de covarianza  $\mathbf{R}_h = \mathbf{R}_t^T \otimes \mathbf{R}_r$  y coeficientes con distribución gaussiana, y en consecuencia ajusta a los supuestos hechos sobre el canal.

Supondremos que las matrices de correlación  $\mathbf{R}_r$  y  $\mathbf{R}_t$  son aleatorias, es decir, cambian con el paso del tiempo. Esta suposición se basa en que, al menos intuitivamente, ciertas disposiciones geográficas del sistemas pueden generar ciertas matrices de correlación  $\mathbf{R}_r$  y  $\mathbf{R}_t$ , pero una vez que estas condiciones geográficas cambian, las matrices de correlación también lo hacen.

---

<sup>2</sup>Debido a que separa la correlación del sistema en correlación debida al transmisor y correlación debida al receptor.

A través de este capítulo normalizaremos la matriz  $\mathbf{H}$  de tal forma que  $\mathbb{E}_{\mathbf{H}}[\text{tr}(\mathbf{H}\mathbf{H}^*)] = n_R n_T$ . Dado que las matrices  $\mathbf{R}_r$  y  $\mathbf{R}_t$  son matrices Hermitianas semidefinidas positivas, podemos escribirlas como

$$\mathbf{R}_r = \mathbf{U}\mathbf{D}_r\mathbf{U}^*, \quad (3.2)$$

$$\mathbf{R}_t = \mathbf{V}\mathbf{D}_t\mathbf{V}^*, \quad (3.3)$$

donde  $\mathbf{U}$  y  $\mathbf{V}$  son matrices unitarias aleatorias de dimensiones  $n_R \times n_R$  y  $n_T \times n_T$  respectivamente,  $\mathbf{D}_r = \text{diag}(d_{r1}, \dots, d_{rn_R})$  es la matriz diagonal real cuyos coeficientes  $d_{ri}$  son los eigenvalores ordenados de la matriz  $\mathbf{R}_r$  y  $\mathbf{D}_t = \text{diag}(d_{t1}, \dots, d_{tn_T})$  es la matriz diagonal real cuyos coeficientes  $d_{ti}$  son los eigenvalores ordenados de la matriz  $\mathbf{R}_t$ .

Después de ciertas manipulaciones algebraicas a la expresión  $\mathbb{E}_{\mathbf{H}}[\text{tr}(\mathbf{H}\mathbf{H}^*)] = n_R n_T$ , se llega a la condición equivalente

$$\mathbb{E}_{\mathbf{D}_r, \mathbf{D}_t} \left[ \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} d_{ri} d_{tj} \right] = n_R n_T. \quad (3.4)$$

Esta última ecuación es una condición general sobre la distribución de los coeficientes de las matrices  $\mathbf{R}_r$  y  $\mathbf{R}_t$  a fin de que se satisfaga la normalización  $\mathbb{E}_{\mathbf{H}}[\text{tr}(\mathbf{H}\mathbf{H}^*)] = n_R n_T$ . Sin embargo, si suponemos que  $\mathbf{R}_r$  y  $\mathbf{R}_t$  son independientes entonces la condición anterior se simplifica a la condición equivalente

$$\mathbb{E}_{\mathbf{D}_r} \left[ \sum_{i=1}^{n_R} d_{ri} \right] \mathbb{E}_{\mathbf{D}_t} \left[ \sum_{j=1}^{n_T} d_{tj} \right] = n_R n_T. \quad (3.5)$$

La correlación de Kronecker es el modelo más utilizado para estudiar la correlación en sistemas MIMO, tanto por la motivación teórica como por su validez práctica [Bli04, Ker02, Yu01]. Como veremos más adelante, este modelo posee propiedades analíticas útiles en términos del análisis de la capacidad ergódica de sistemas MIMO correlacionados. En las siguientes secciones analizaremos la capacidad ergódica asintótica de canales con correlación de Kronecker.

En general estaremos en presencia de sistemas con una cantidad finita de antenas, y por lo tanto debemos adaptar la correlación presente en el sistema finito para extenderla al caso asintótico. En particular, trabajaremos con dos formas de extender la correlación. La primera

será suponer que los eigenvalores de las matrices de correlación en cuestión son v.a.s iid cuya distribución común tiene como imagen ciertos valores observados<sup>3</sup>. La otra forma en que modelaremos la correlación es tomando los eigenvalores como v.a.s iid con una cierta distribución continua<sup>4</sup>. En secciones subsecuentes denominaremos como caso discreto y caso continuo a dichas formas de extender la correlación respectivamente.

## 3.2. Correlación Unilateral

Como ya se había mencionado en la introducción de la Sección 2.3, en ambientes ricos en múltiples trayectorias se puede asegurar la independencia entre los coeficientes de propagación mediante el espaciado adecuado de las antenas. Por lo tanto cabe pensar que en sistemas donde exista la posibilidad de brindar de ese espacio al receptor o al transmisor, la correlación en el sistema será originada únicamente en el extremo opuesto del que dispone de dicho espacio. Debido a la naturaleza de esta situación, a este tipo de correlación se le suele denominar *correlación unilateral*. Esta correlación se presenta en dos formas diferentes, cuando el receptor no presenta correlación y cuando el transmisor no presenta correlación. Si bien ambos casos son muy similares, en general no son idénticos. A continuación se analiza el caso en el que el receptor no presenta correlación, seguido del caso análogo para el transmisor.

### 3.2.1. Receptor No Correlacionado

En esta sección supondremos que el receptor no está correlacionado, es decir  $\mathbf{R}_r = \mathbf{I}$ . Además, supondremos que se tiene una entrada  $\mathbf{x}$  gaussiana compleja circular simétrica con matriz de covarianza  $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = \frac{P}{n_T}\mathbf{I}$ , es decir, supondremos que se tiene una entrada isotrópica<sup>5</sup> con potencia disponible  $P$ . El siguiente teorema, el cual puede derivarse de la demostración del Teorema 2.1, nos permite calcular la eficiencia espectral ergódica de un sistema MIMO en general cuando se tiene una entrada gaussiana compleja circular simétrica.

---

<sup>3</sup>Los valores observados a los que hacemos referencia son los eigenvalores de las matrices de correlación empíricas, las cuales se estiman directamente de las mediciones del canal. Debido a la invarianza unitaria de la matriz  $\mathbf{H}_w$ , únicamente estamos interesados en la distribución de los eigenvalores de las matrices de correlación y no en la distribución de las matrices por sí mismas. Por lo tanto, supondremos que las matrices de correlación son matrices diagonales.

<sup>4</sup>Claramente aquí se supone que se hizo inferencia sobre dicha distribución previamente.

<sup>5</sup>El análisis se centra en la eficiencia espectral ergódica para una entrada isotrópica ya que, como se mencionó anteriormente, en muchos casos los dispositivos móviles no disponen de información acerca del estado del canal, en cuyo caso la mejor opción es transmitir de manera isotrópica.

**Teorema 3.1.** *Supóngase que se tiene un canal MIMO con matriz de coeficientes de propagación  $\mathbf{H}$  y vector de entrada  $\mathbf{x}$ . Si  $\mathbf{x}$  es un vector Gaussiano complejo circular simétrico con media cero y matriz de covarianza  $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = \mathbf{\Phi}$ , entonces la eficiencia espectral ergódica del sistema es igual a*

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}}[\log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{\Phi}\mathbf{H}^*)]. \quad (3.6)$$

Por el teorema anterior, la eficiencia espectral ergódica en el caso en el que el receptor no está correlacionado es igual a

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{H}\mathbf{R}_t^{1/2} \left( \mathbf{H}\mathbf{R}_t^{1/2} \right)^* \right) \right] \quad (3.7)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{H}\mathbf{R}_t\mathbf{H}^* \right) \right]. \quad (3.8)$$

Aplicando el Teorema C.1 obtenemos que

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \lambda_i \left( \mathbf{I} + \frac{P}{n_T} \mathbf{H}\mathbf{R}_t\mathbf{H}^* \right) \right], \quad (3.9)$$

donde  $\lambda_1(\mathbf{I} + (P/n_T)\mathbf{H}\mathbf{R}_t\mathbf{H}^*) \leq \dots \leq \lambda_{n_R}(\mathbf{I} + (P/n_T)\mathbf{H}\mathbf{R}_t\mathbf{H}^*)$  son los eigenvalores ordenados de la matriz  $\mathbf{I} + (P/n_T)\mathbf{H}\mathbf{R}_t\mathbf{H}^*$ . Simplificando la expresión anterior y cambiando la suma por una integral respecto a la distribución empírica espectral de  $(1/n_R)\mathbf{H}\mathbf{R}_t\mathbf{H}^*$ , llegamos a

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \left( 1 + \frac{Pn_R}{n_T} \lambda_i \left( \frac{1}{n_R} \mathbf{H}\mathbf{R}_t\mathbf{H}^* \right) \right) \right] \quad (3.10)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{Pn_R}{n_T} x \right) n_R dF^{\mathbf{H}\mathbf{R}_t\mathbf{H}^*/n_R}(x) \right]. \quad (3.11)$$

Escribamos  $\mathbf{B}_{n_R} = \mathbf{H}\mathbf{R}_t\mathbf{H}^*/n_R$ , entonces

$$\frac{I(\mathbf{x}; \mathbf{y})}{n_R} = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{Pn_R}{n_T} x \right) dF^{\mathbf{B}_{n_R}}(x) \right]. \quad (3.12)$$

Por lo tanto, la eficiencia espectral ergódica para una entrada isotrópica solamente depende de la distribución empírica espectral de  $\mathbf{B}_{n_R}$ . Al igual que en el capítulo anterior, aproximaremos la eficiencia espectral ergódica mediante la eficiencia espectral ergódica asintótica, la cual obtendremos a partir de la distribución empírica espectral asintótica de  $\mathbf{B}_{n_R}$ .

Por la forma de extender los valores de correlación observados al caso asintótico, la función de distribución empírica espectral de  $\mathbf{R}_t$  converge con probabilidad uno a una función de distribución determinista  $R(x)$ .

Aplicando el Teorema C.7 concluimos que la transformada de Stieltjes de la función de distribución empírica espectral de  $\mathbf{B}_{n_R}$  converge cuando  $n_R \rightarrow \infty$  y  $n_T/n_R \rightarrow \beta$  a  $m_{\mathbf{B}}(z)$ , la cual está dada por la ecuación de punto fijo

$$m_{\mathbf{B}} = m_{\mathbf{0}} \left( z - \beta \int \frac{x}{1 + xm_{\mathbf{B}}(z)} R(dx) \right), \quad (3.13)$$

donde  $m_{\mathbf{0}}(z)$  es la transformada de Stieltjes de la función de distribución  $F(x) = \mathbb{1}_{\{x \geq 0\}}$ . Dado que  $m_{\mathbf{0}}(z) = 1/(-z)$ , se tiene que

$$m_{\mathbf{B}}(z) = \left( -z + \beta \int \frac{x}{1 + xm_{\mathbf{B}}(z)} R(dx) \right)^{-1}. \quad (3.14)$$

### Caso Discreto

En el caso en el que la distribución  $R(x)$  toma una cantidad finita de valores, digamos  $r_1, \dots, r_n$  con probabilidades  $p_1, \dots, p_n$ , entonces la ecuación (3.14) puede escribirse de las siguientes formas

$$m_{\mathbf{B}}(z) = \left( -z + \beta \sum_{i=1}^n \frac{r_i p_i}{1 + r_i m_{\mathbf{B}}(z)} \right)^{-1}, \quad (3.15)$$

$$m_{\mathbf{B}}(z) = \left( -z + \frac{\beta}{\prod_{i=1}^n (1 + r_i m_{\mathbf{B}}(z))} \sum_{i=1}^n r_i p_i \prod_{j \neq i} (1 + r_j m_{\mathbf{B}}(z)) \right)^{-1}, \quad (3.16)$$

$$\prod_{i=1}^n (1 + r_i m_{\mathbf{B}}(z)) - m_{\mathbf{B}}(z) \left( -z \prod_{i=1}^n (1 + r_i m_{\mathbf{B}}(z)) + \beta \sum_{i=1}^n r_i p_i \prod_{j \neq i} (1 + r_j m_{\mathbf{B}}(z)) \right) = 0. \quad (3.17)$$

Esta última fórmula que a primera vista parece un tanto complicada, para un valor dado de  $z \in \mathbb{C}^+$  es un polinomio en  $m_{\mathbf{B}}(z)$ . El Teorema C.7 implica que  $m_{\mathbf{B}}(z)$  es la única raíz del polinomio de la ecuación (3.17) con parte imaginaria positiva. Por lo tanto, para encontrar el valor de  $m_{\mathbf{B}}(z)$  basta encontrar numéricamente las soluciones al polinomio en cuestión y buscar la única solución en  $\mathbb{C}^+$ . Los detalles de la implementación se presentan en la Sección 3.6.

Por último, para evaluar la eficiencia espectral ergódica asintótica es necesario obtener la función de densidad de probabilidad que corresponde a la transformada de Stieltjes obtenida, para lo cual utilizaremos la fórmula de inversión (Teorema C.3).

### Caso Continuo

Dado que en el caso discreto la resolución numérica de la ecuación (3.14) es relativamente simple, en esta tesis proponemos aproximar la solución del caso continuo mediante las soluciones de ciertos casos discretos.

Para tal fin aproximaremos  $R(x)$  por una función  $R_n(x)$  la cual es la función de distribución de la v.a. uniforme en el conjunto  $\{R^{-1}(i/n) : 1 \leq i \leq n\}$ , donde  $R^{-1}(p) = \sup\{x \in \mathbb{R} : F(x) \leq p\}$ . Dado que  $R_n(x)$  es discreta, podemos encontrar la solución correspondiente mediante el método de la sección anterior. Por último, tomaremos la solución encontrada para  $R_n(x)$  como aproximación a la solución correspondiente a  $R(x)$ .

Para comprobar la validez de este método de aproximación demostramos el siguiente teorema, el cual afirma que cuando  $n \rightarrow \infty$  la solución aproximada que se propone converge a la solución exacta.

**Teorema 3.2.** *Supóngase que*

- a. *Para cada  $n_R = 1, 2, \dots$  sea  $\mathbf{H}_{n_R}$  una matriz de  $n_R \times n_T$  con coeficientes complejos idénticamente distribuidos para toda  $n, i, j$ .*
- b. *Los coeficientes  $\mathbf{H}_{n_R}(i, j)$  son independientes para cada  $n_R$  y  $\text{var}(\mathbf{H}_{n_R}(i, j)) = 1$ .*
- c.  *$n_T = n_T(n_R)$  con  $n_T/n_R \rightarrow \beta > 0$  cuando  $n_R \rightarrow \infty$ .*
- d.  *$\mathbf{R}_{n_T}^n = \text{diag}(r_{n_T}^n(1), \dots, r_{n_T}^n(n_T))$  con  $r_{n_T}^n(i) \in \mathbb{R}$  y la función de distribución empírica de  $\{r_{n_T}^n(1), \dots, r_{n_T}^n(n_T)\}$  converge débilmente, con probabilidad uno, a una función de distribución no aleatoria  $R_n(x)$  cuando  $n_R \rightarrow \infty$ . Además, supóngase que  $R_n(x) \xrightarrow{d} R(x)$  cuando  $n \rightarrow \infty$ , donde  $R(x)$  es una función de distribución de probabilidad.*
- e.  *$\mathbf{H}_{n_R}$  y  $\mathbf{R}_{n_T}$  son independientes.*

Sea  $\mathbf{B}_{n_R}^n = \mathbf{H}_{n_R} \mathbf{R}_{n_T}^n \mathbf{H}_{n_R}^* / n_R$ , entonces con probabilidad uno  $F^{\mathbf{B}_{n_R}^n} \xrightarrow{v} F^{\mathbf{B}}$  cuando  $n, n_R \rightarrow \infty$ , donde  $F^{\mathbf{B}}$  es una función de distribución no aleatoria cuya transformada de Stieltjes  $m_{\mathbf{B}}(z)$

satisface que

$$m_{\mathbf{B}}(z) = - \left( z - \beta \int \frac{x}{1 + xm_{\mathbf{B}}(z)} dR(x) \right)^{-1} \quad (3.18)$$

para todo  $z \in \mathbb{C}^+$ . Además,  $m_{\mathbf{B}}(z)$  es la única solución a la ecuación anterior con parte imaginaria positiva.

**Demostración.** Sea  $\mathbf{R}_{n_T}$  una matriz diagonal de  $n_T \times n_T$  con coeficientes iid con distribución común  $R(x)$ . Observemos que los coeficientes de la matriz  $\mathbf{B}_{n_R}^n$  son sumas de elementos de la forma  $gh_1h_2$  con  $g \in \mathbf{R}_{n_T}^n$  y  $h_1, h_2 \in \mathbf{H}_w$ . Debido a la independencia entre los coeficientes de  $\mathbf{R}_{n_T}^n$  y  $\mathbf{H}_w$ , la convergencia  $R_n(x) \xrightarrow{d} R(x)$  implica que los coeficientes de  $\mathbf{B}_{n_R}^n$  convergen en distribución a los coeficientes de la matriz  $\mathbf{B}_{n_R} := \mathbf{H}_{n_R} \mathbf{R}_{n_T} \mathbf{H}_{n_R}^* / n_R$  cuando  $n \rightarrow \infty$ .

Dado que los eigenvalores de una matriz son una función continua de sus coeficientes, los eigenvalores de  $\mathbf{B}_{n_R}^n$  convergen en distribución a los eigenvalores de  $\mathbf{B}_{n_R}$  (de manera conjunta) y por lo tanto  $F^{\mathbf{B}_{n_R}^n} \xrightarrow{d} F^{\mathbf{B}_{n_R}}$  cuando  $n \rightarrow \infty$ . Cuando  $n_R \rightarrow \infty$  se tiene que con probabilidad uno  $F^{\mathbf{B}_{n_R}}(x) \xrightarrow{v} F^{\mathbf{B}}(x)$ , donde  $F^{\mathbf{B}}$  es una función de distribución no aleatoria que satisface la ecuación de punto fijo del Teorema C.7. Dado que la convergencia en distribución es transitiva se tiene que con probabilidad uno  $F^{\mathbf{B}_{n_R}^n}(x) \xrightarrow{v} F^{\mathbf{B}}(x)$  cuando  $n, n_R \rightarrow \infty$ .

Por el Teorema C.5 se tiene que  $F^{\mathbf{B}_{n_R}^n}(x) \xrightarrow{v} F^{\mathbf{B}}(x)$  si y sólo si  $m_{\mathbf{B}_{n_R}^n}(z) \xrightarrow{n, n_R \rightarrow \infty} m_{\mathbf{B}}(z)$  para todo  $z \in \mathbb{C}$ . Si  $\mathbf{A}$  es la matriz nula de  $n_R \times n_R$ , aplicando el Teorema C.7 a  $\mathbf{B}_{n_R}$  se obtiene el resultado deseado.  $\square$

En términos de la demostración anterior,  $\mathbf{R}_{n_T}$  es la matriz de correlación cuyos coeficientes tienen como distribución común a  $R(x)$ , es decir, la capacidad en el caso continuo está dada por el espectro de la matriz  $\mathbf{B}_{n_R}$ . En consecuencia, la convergencia  $F^{\mathbf{B}_{n_R}^n}(x) \xrightarrow{v} F^{\mathbf{B}}(x)$  asegura que nuestro método es correcto.

En otras palabras, el teorema anterior demuestra que mediante una discretización de  $R(x)$  podemos aproximar la verdadera transformada de Stieltjes de tal forma que cuando  $n, n_R \rightarrow \infty$  dicha aproximación coincide con la transformada deseada.

La discretización de  $R(x)$  transforma este caso, en el que  $R(x)$  es continua, al caso de la sección anterior. Por lo tanto, utilizando el software de la sección anterior (con los  $r_i$  y  $p_i$  adecuados) podemos aproximar la transformada de Stieltjes en este caso más general, y en consecuencia dar una aproximación para la eficiencia espectral ergódica. Es importante

mencionar que el comportamiento de la convergencia depende totalmente de las  $r_i$  y las  $p_i$  tomadas, así como el valor de  $n$  elegido<sup>6</sup>.

### 3.2.2. Transmisor No Correlacionado

En este caso supondremos que el transmisor no presenta correlación, esto es  $\mathbf{R}_t = \mathbf{I}$ . El desarrollo que se realiza en esta sección para calcular la eficiencia espectral ergódica cuando el transmisor no presenta correlación es similar al realizado en la sección anterior, prácticamente la única diferencia es que en lugar de utilizar el Teorema C.7 se utiliza el Teorema C.8. La eficiencia espectral ergódica alcanzada por una matriz de covarianza  $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = \frac{P}{n_T}\mathbf{I}$  es igual a

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H} \left( \mathbf{R}_r^{1/2} \mathbf{H} \right)^* \right) \right] \quad (3.19)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \lambda_i \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{H}^* \mathbf{R}_r^{1/2} \right) \right] \quad (3.20)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \left( 1 + P \lambda_i \left( \frac{1}{n_T} \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{H}^* \mathbf{R}_r^{1/2} \right) \right) \right] \quad (3.21)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 (1 + Px) n_R dF^{\mathbf{R}_r^{1/2} \mathbf{H} \mathbf{H}^* \mathbf{R}_r^{1/2} / n_T}(x) \right]. \quad (3.22)$$

Si escribimos  $\mathbf{B}_{n_R} := \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{H}^* \mathbf{R}_r^{1/2} / n_T$ , entonces

$$\frac{I(\mathbf{x}; \mathbf{y})}{n_R} = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 (1 + Px) dF^{\mathbf{B}_{n_R}}(x) \right]. \quad (3.23)$$

Por lo tanto, la eficiencia espectral ergódica solamente depende de la distribución empírica espectral de  $\mathbf{B}_{n_R}$ . Al igual que en el capítulo anterior, aproximaremos la eficiencia espectral ergódica mediante la eficiencia espectral ergódica asintótica.

Por la forma de extender los valores de correlación observados al caso asintótico, es claro que en este caso la función de distribución empírica espectral de  $\mathbf{R}_t$  también converge con probabilidad uno a una función de distribución no aleatoria  $R(x)$  cuando  $n_R \rightarrow \infty$ .

Aplicando el Teorema C.8 se obtiene que la distribución empírica espectral de  $\mathbf{B}_{n_R}$  converge a una función de distribución no aleatoria  $F^{\mathbf{B}}$  cuando  $n_R \rightarrow \infty$  y  $n_T/n_R \rightarrow \beta$ . La

---

<sup>6</sup>El valor de  $n$  teóricamente debe ser tan grande como sea posible, sin embargo, conforme  $n$  crece los problemas originados por errores numéricos aumentan (véanse las Secciones 3.5 y 3.6).

transformada de Stieltjes  $m_{\mathbf{B}}(z)$  de dicha función de distribución está dada por la siguiente ecuación de punto fijo

$$m_{\mathbf{B}}(z) = \int \frac{1}{x(1-c-czm_{\mathbf{B}}(z))-z} dR(x), \quad (3.24)$$

con  $c = \beta^{-1}$ .

Si  $R(x)$  toma una cantidad finita de valores, digamos  $r_1, \dots, r_n$  con probabilidades  $p_1, \dots, p_n$ , entonces la transformada de Stieltjes buscada  $m_{\mathbf{B}}(z)$  satisface que

$$m_{\mathbf{B}}(z) = \sum_{i=1}^n \frac{p_i}{-r_i c z m_{\mathbf{B}}(z) + r_i - r_i c - z}, \quad (3.25)$$

$$m_{\mathbf{B}}(z) = \frac{\sum_{i=1}^n p_i \prod_{j \neq i} (-r_j c z m_{\mathbf{B}}(z) + r_j - r_j c - z)}{\prod_{i=1}^n (-r_i c z m_{\mathbf{B}}(z) + r_i - r_i c - z)}, \quad (3.26)$$

$$m_{\mathbf{B}}(z) \prod_{i=1}^n (-r_i c z m_{\mathbf{B}}(z) + r_i - r_i c - z) - \sum_{i=1}^n p_i \prod_{j \neq i} (-r_j c z m_{\mathbf{B}}(z) + r_j - r_j c - z) = 0. \quad (3.27)$$

Para  $z \in \mathbb{C}^+$  fija, la ecuación anterior es un polinomio en  $m_{\mathbf{B}}(z)$  y en consecuencia podemos proceder de la misma manera que en la sección anterior. En el caso continuo utilizamos el mismo método que el utilizado en el Teorema 3.2 de la sección anterior. La demostración de que este método es correcto en esta situación, es muy similar a la demostración del Teorema 3.2, por lo cual la omitimos. En la Sección 3.6 se discuten los detalles de implementación de este método.

### 3.3. Correlación Bilateral

En esta sección se analizan sistemas MIMO en los cuales la correlación está presente en ambos extremos del enlace, de aquí que a este tipo de correlación se le denomine *correlación bilateral*.

Siguiendo el mismo desarrollo que el utilizado en los casos con correlación unilateral, encontramos que la eficiencia espectral ergódica alcanzada por una matriz de covarianza

$\mathbb{E}[\mathbf{x}\mathbf{x}^*] = \frac{P}{n_T}\mathbf{I}$  es igual a

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t^{1/2} \left( \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t^{1/2} \right)^* \right) \right] \quad (3.28)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \lambda_i \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t \mathbf{H}^* \mathbf{R}_r^{1/2} \right) \right] \quad (3.29)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \left( 1 + \frac{P n_R}{n_T} \lambda_i \left( \frac{1}{n_R} \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t \mathbf{H}^* \mathbf{R}_r^{1/2} \right) \right) \right] \quad (3.30)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{P n_R}{n_T} x \right) n_R dF^{\mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t \mathbf{H}^* \mathbf{R}_r^{1/2} / n_R}(x) \right]. \quad (3.31)$$

Escribamos  $\mathbf{B}_{n_R} = \mathbf{R}_r^{1/2} \mathbf{H} \mathbf{R}_t \mathbf{H}^* \mathbf{R}_r^{1/2} / n_R$ , entonces

$$\frac{I(\mathbf{x}; \mathbf{y})}{n_R} = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 \left( 1 + \frac{P n_R}{n_T} x \right) dF^{\mathbf{B}_{n_R}}(x) \right]. \quad (3.32)$$

Por lo tanto, la eficiencia espectral ergódica únicamente depende de la distribución empírica espectral de  $\mathbf{B}_{n_R}$ .

Supóngase que los eigenvalores de la matriz  $\mathbf{R}_r$  son v.a.s iid con distribución común  $R(x)$  y que los eigenvalores de la matriz  $\mathbf{R}_t$  son v.a.s iid con distribución común  $T(x)$ . Supóngase que las distribuciones  $R(x)$  y  $T(x)$  se discretizan de la misma manera que en la sección anterior, esto es, proporcionando versiones discretas  $R_n(x)$  y  $T_n(x)$  las cuales toman únicamente  $n$  valores. Utilizando argumentos similares a los de la demostración del Teorema 3.2 se obtiene que en el caso asintótico la transformada de Stieltjes obtenida por la versión discretizada del sistema converge, cuando  $n \rightarrow \infty$ , a la transformada de Stieltjes teórica deseada.

Dado que las discretizaciones  $R_n(x)$  y  $T_n(x)$  tienen soporte compacto, entonces el Corolario C.3 es aplicable a esta situación. Debido a la complejidad intrínseca del sistema, no es posible aplicar técnicas basadas en polinomios para encontrar la transformada de Stieltjes de la distribución empírica espectral asintótica de  $\mathbf{B}_{n_R}$ . Por lo tanto, es necesario iterar la ecuación de punto fijo proporcionada por el Corolario C.3, los detalles de implementación pueden ser consultados en la Sección 3.6.

## 3.4. Canales Relacionados

En esta sección discutimos brevemente dos canales de comunicación inalámbrica conocidos como canales IND y canal de Rice no correlacionado. Si bien estos canales no son canales correlacionados, la eficiencia espectral ergódica de dichos canales se obtiene directamente de lo expuesto o se puede derivar utilizando el mismo método que se utilizó para los canales con correlación unilateral.

### 3.4.1. Canales IND

Supongamos que  $\mathbf{R}_r$  y  $\mathbf{R}_t$  son matrices diagonales semidefinidas positivas. En esta situación la matriz  $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w \mathbf{R}_r^{1/2}$  es una matriz cuyos coeficientes son v.a.s gaussianas independientes no idénticamente distribuidas. Este tipo de canales son conocidos como canales IND (por sus siglas del inglés Independent Non identically Distributed [Tul05]).

Estos canales surgen en sistemas donde la independencia estadística de los coeficientes de propagación puede ser lograda mediante ciertas técnicas las cuales comprometen la uniformidad de las distribuciones de los coeficientes de propagación. Por ejemplo, una técnica que propicia la obtención de coeficientes de propagación independientes es la polarización de las señales transmitidas, sin embargo, esto puede comprometer la uniformidad de las distribuciones de los coeficientes de propagación.

Claramente el tipo de canales IND presentados en este capítulo son un caso particular de los canales con correlación bilateral. De hecho, en la sección anterior se mostró (utilizando la invarianza unitaria de la matriz gaussiana) que en el caso de canales con correlación bilateral únicamente es relevante la distribución de los eigenvalores de las matrices de correlación y no su distribución particular, por lo cual se supuso, sin pérdida de generalidad, que dichas matrices de correlación eran matrices diagonales.

### 3.4.2. Canal de Rice No Correlacionado

El canal de Rice es un canal de comunicación inalámbrica similar al canal de Rayleigh. A diferencia del canal de Rayleigh, el canal de Rice tiene media no nula, la cual representa

físicamente el componente de línea de vista del sistema<sup>7</sup>. Denotemos por  $\mathbf{H}$  la matriz de coeficientes de propagación del canal y escribamos  $\mathbf{H}_m = \mathbb{E}[\mathbf{H}]$ . En esta situación el canal puede expresarse como

$$\mathbf{H} = \mathbf{H}_m + \mathbf{H}_w, \quad (3.33)$$

donde  $\mathbf{H}_w$  es la matriz gaussiana utilizada en el modelo canónico.

Una situación donde el análisis de este tipo de canales podría ser útil es en la siguiente: Supóngase que se disponen de varias ubicaciones geográficas para implementar un sistema de comunicaciones y que cada configuración tiene su propia media  $\mathbf{H}_m$ . Si extendemos  $\mathbf{H}_m$  al caso asintótico de la misma manera en que extendimos la correlación en la Sección 3.2, entonces la distribución empírica espectral de  $\frac{1}{n_T} \mathbf{H}_m \mathbf{H}_m^*$  converge a una distribución no aleatoria  $R(x)$  cuando  $n_R \rightarrow \infty$  para cada posible disposición geográfica.

En esta situación es de interés conocer la eficiencia espectral ergódica de cada posible ubicación del sistema, de esta forma es posible cuantificar el incremento en la eficiencia espectral ergódica (con respecto al canal de Rayleigh) causado por la componente de línea de vista de cada posible ubicación.

Realizando cálculos similares a los realizados en secciones anteriores, se tiene que la eficiencia espectral ergódica alcanzada por una entrada isotrópica con potencia disponible  $P$  es igual a

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{H} \mathbf{H}^* \right) \right] \quad (3.34)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \lambda_i \left( \mathbf{I} + \frac{P}{n_T} \mathbf{H} \mathbf{H}^* \right) \right] \quad (3.35)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \sum_{i=1}^{n_R} \log_2 \left( 1 + P \lambda_i \left( \frac{1}{n_T} \mathbf{H} \mathbf{H}^* \right) \right) \right] \quad (3.36)$$

$$= \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2 (1 + Px) n_R dF^{\mathbf{H} \mathbf{H}^* / n_T}(x) \right]. \quad (3.37)$$

Escribamos  $B_{n_R} = \mathbf{H} \mathbf{H}^* / n_T$ , entonces la eficiencia espectral ergódica por antena receptora es

---

<sup>7</sup>Teniendo en cuenta la existencia de múltiples trayectorias por las cuales una señal transmitida llega de la antena transmisora a una antena receptora, se le denomina componente de línea de vista a la trayectoria que va de la antena transmisora a la antena receptora directamente. Para que exista tal trayectoria es necesario que la antena receptora sea visible desde la antena transmisora, de aquí que a este componente se le llame componente de línea de vista.

igual a

$$\frac{I(\mathbf{x}; \mathbf{y})}{n_R} = \mathbb{E}_{\mathbf{H}} \left[ \int_0^\infty \log_2(1 + Px) \, dF^{\mathbf{B}_{n_R}}(x) \right]. \quad (3.38)$$

Aplicando el Teorema C.9 se tiene que la distribución empírica espectral de  $\mathbf{B}_{n_R}$  converge a una función de distribución no aleatoria  $F^{\mathbf{B}}$  cuya transformada de Stieltjes está dada por la siguiente ecuación de punto fijo

$$m_{\mathbf{B}}(z) = \int \frac{dR(x)}{\frac{x}{1+cm_{\mathbf{B}}(z)} - (1 + cm_{\mathbf{B}}(z))z + (1 - c)}, \quad (3.39)$$

con  $c = \beta^{-1}$ .

Si el conjunto de valores que toma  $R(x)$  es finito, digamos  $r_1, \dots, r_n$  con probabilidad  $p_1, \dots, p_n$ , entonces la transformada de Stieltjes buscada  $m_{\mathbf{B}}(z)$  satisface que

$$m_{\mathbf{B}}(z) = \sum_{i=1}^n \frac{p_i}{\frac{r_i}{1+cm_{\mathbf{B}}(z)} - (1 + cm_{\mathbf{B}}(z))z + (1 - c)}. \quad (3.40)$$

Simplificando la expresión anterior obtenemos la expresión equivalente

$$m_{\mathbf{B}}(z) = \sum_{i=1}^n \frac{p_i(1 + cm_{\mathbf{B}}(z))}{r_i - (1 + cm_{\mathbf{B}}(z))^2 z + (1 + cm_{\mathbf{B}}(z))(1 - c)}. \quad (3.41)$$

Multiplicando por el factor común  $\prod_{i=1}^n (r_i - (1 + cm_{\mathbf{B}}(z))^2 z + (1 + cm_{\mathbf{B}}(z))(1 - c))$  tenemos que

$$0 = m_{\mathbf{B}}(z) \prod_{i=1}^n (r_i - (1 + cm_{\mathbf{B}}(z))^2 z + (1 + cm_{\mathbf{B}}(z))(1 - c)) - \sum_{i=1}^n p_i(1 + cm_{\mathbf{B}}(z)) \prod_{j \neq i} (r_j - (1 + cm_{\mathbf{B}}(z))^2 z + (1 + cm_{\mathbf{B}}(z))(1 - c)). \quad (3.42)$$

Definamos  $A_i := -c^2 z$ ,  $B_i := c - c^2 - 2cz$  y  $C_i := r_i - z + 1 - c$  con  $i = 1, \dots, n$ , entonces

$$0 = m_{\mathbf{B}}(z) \prod_{i=1}^n (A_i m_{\mathbf{B}}(z)^2 + B_i m_{\mathbf{B}}(z) + C_i) - \sum_{i=1}^n p_i(1 + cm_{\mathbf{B}}(z)) \prod_{j \neq i} (A_j m_{\mathbf{B}}(z)^2 + B_j m_{\mathbf{B}}(z) + C_j). \quad (3.43)$$

Al igual que antes la expresión anterior es un polinomio en  $m_{\mathbf{B}}(z)$ , y por lo tanto encontrar la transformada de Stieltjes  $m_{\mathbf{B}}(z)$  es equivalente a encontrar las soluciones del polinomio en cuestión.

En el caso continuo podemos utilizar nuevamente discretizaciones similares a las del Teorema 3.2 para aproximar la transformada de Stieltjes buscada mediante otras transformada que son solución a ciertos casos discretos. En la Sección 3.6 se discuten los detalles de implementación del software que calcula la transformada de Stieltjes  $m_{\mathbf{B}}(z)$ .

### 3.5. Análisis Numérico de la Eficiencia Espectral Ergódica Isotrópica

En esta sección analizamos algunos casos particulares de los canales con correlación de Kronecker unilateral y bilateral<sup>8</sup>. Los resultados se exponen en tres secciones: receptor no correlacionado, transmisor no correlacionado y correlación bilateral. En cada una de las tres secciones se expone un análisis de la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica, así como una comparativa entre ellas (cuando se tienen hasta 5 antenas por extremo). Para ejemplificar los efectos de la correlación se utilizan cuatro distribuciones diferentes<sup>9</sup>:

*Bernoulli.* En este caso la correlación tiene una distribución uniforme en el conjunto  $\{1/2, 3/2\}$ . La distribución ya es discreta, por lo tanto no hay necesidad de discretizarla.

*Uniforme.* La correlación en este caso se distribuye uniformemente en el intervalo  $[0, 2]$ . Esta función de distribución se discretiza mediante una distribución uniforme en el conjunto  $\{1/20, \dots, 20/20\}$ . Por lo tanto, el grado de los polinomios utilizados en los casos con correlación unilateral es veintiuno.

*Exponencial.* Aquí la correlación tiene distribución exponencial de parámetro  $\lambda = 1$ . La discretización se realiza mediante una distribución uniforme en el conjunto  $\{F^{-1}(i/16) : i = 1, \dots, 15\} \cup \{F^{-1}(31/32)\}$ , de manera similar que en la Sección 3.2.1. Por lo tanto los polinomios en cuestión tienen grado diecisiete.

---

<sup>8</sup>Dado que el canal de Rice no es un sistema correlacionado, analizar su desempeño queda fuera del contexto y objetivo de esta tesis. Por esta razón no se incluyen simulaciones de este canal.

<sup>9</sup>En las secciones de correlación unilateral, la distribución de  $R(x)$  toma cada una de las cuatro distribuciones posibles. Para el caso con correlación bilateral, tanto  $R(x)$  como  $T(x)$  toman la misma distribución, es decir, cuando la correlación en el receptor tiene distribución Bernoulli también la correlación en el transmisor tiene distribución Bernoulli y así para las cuatro distribuciones. Sin embargo, el programa desarrollado también funciona en el caso en el que las distribuciones de  $R(x)$  y  $T(x)$  son diferentes.

*Chi Cuadrada.* En este caso la correlación tiene distribución chi cuadrada con un grado de libertad. Al igual que con la distribución exponencial la discretización se realizó mediante una distribución uniforme en el conjunto  $\{F^{-1}(i/16) : i = 1, \dots, 15\} \cup \{F^{-1}(31/32)\}$ .

Obsérvese que los conjuntos utilizados no pueden ser de la forma  $\{F^{-1}(i/16) : i = 1, \dots, 16\}$  puesto que  $F^{-1}(1) = \infty$  lo cual produce errores numéricos.

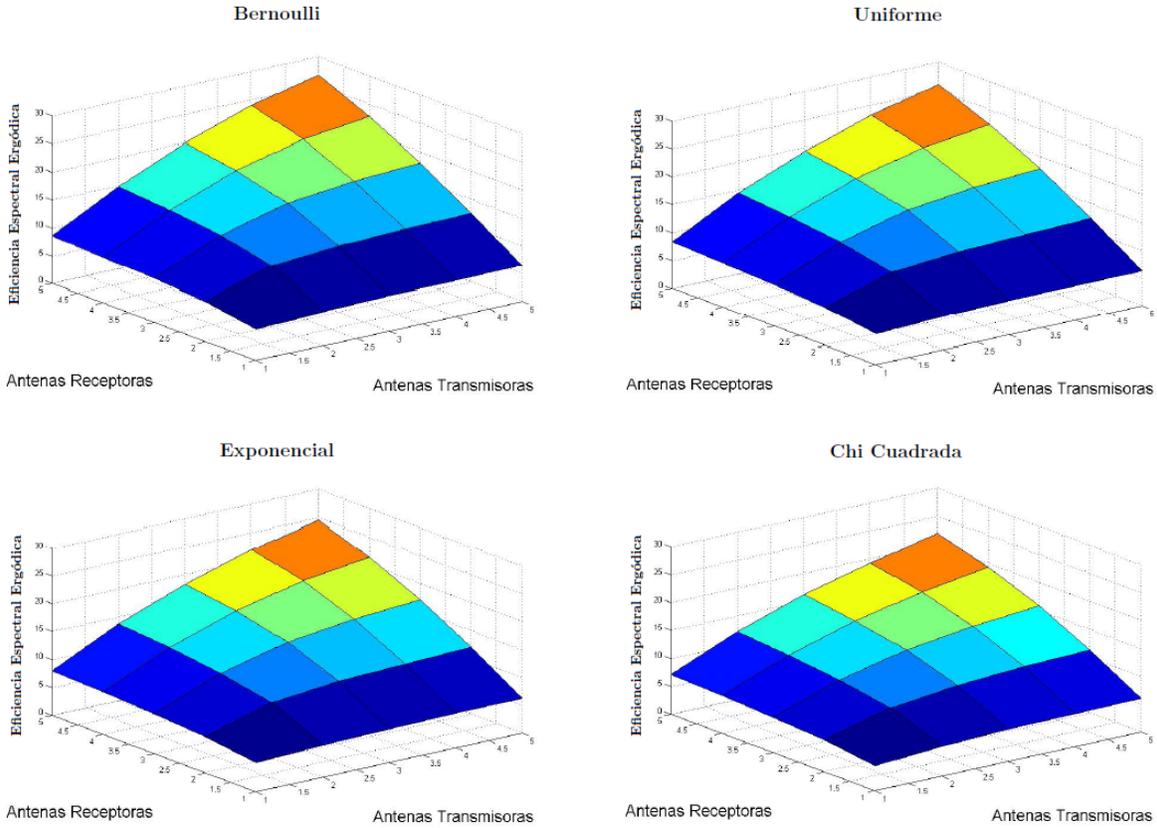
Para facilitar la comparación de las simulaciones de este capítulo con las del capítulo anterior, todas las simulaciones satisfacen la condición de la ecuación (3.5) (la media de las variables aleatorias utilizadas para modelar la correlación en los cuatro casos es uno) y se realizan para una potencia  $P = 100$ .

### 3.5.1. Receptor No Correlacionado

Con el fin de ilustrar el efecto que tiene la correlación en la eficiencia espectral ergódica cuando el receptor no está correlacionado, las siguientes gráficas<sup>10</sup> muestran la eficiencia espectral ergódica alcanzada cuando la correlación en el transmisor tiene distribución Bernoulli, uniforme, exponencial y chi cuadrada respectivamente. Obsérvese que la forma de las cuatro gráficas es la misma, la única variación es la escala. Se puede observar que en todos los casos la correlación tiene un efecto negativo en la eficiencia espectral ergódica respecto a la alcanzada por el modelo canónico.

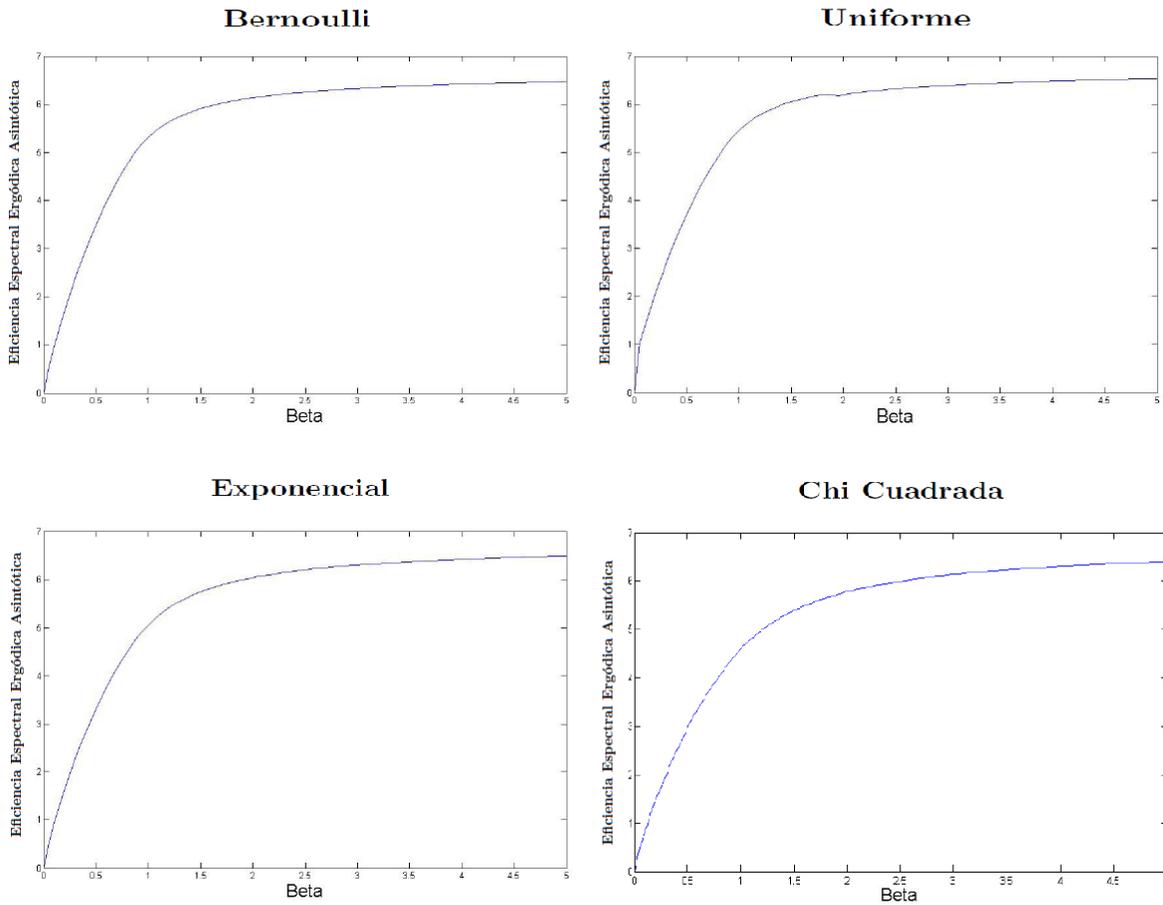
---

<sup>10</sup>Los programas utilizados para realizar las gráficas de esta sección se encuentran en la siguiente sección.



Si bien el método de aproximación funciona teóricamente, es lógico esperar que se presenten errores numéricos al aproximar las raíces de los polinomios utilizados o al aplicar la fórmula de inversión. Teniendo esto en cuenta, es necesario tener un parámetro que nos permita decidir si la simulación sigue dentro de las condiciones de operación correctas o no, para este fin utilizaremos la eficiencia espectral ergódica asintótica cuando  $\beta \rightarrow \infty$ .

Observemos que si  $n_R = 1$  entonces  $1 + \frac{P}{n_T} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w^* = 1 + \frac{P}{n_T} \sum_{j=1}^{n_T} d_{t_j} h_{1j} h_{1j}^*$ , donde  $\mathbf{R}_t = \text{diag}(d_{t_1}, \dots, d_{t_{n_T}})$  y  $\mathbf{H}_w = \{h_{ij}\}$ . Debido a la independencia entre  $\mathbf{R}_t$  y  $\mathbf{H}_w$  se tiene que  $\mathbb{E}[d_{t_j} h_{1j} h_{1j}^*] = 1$  para toda  $j$ . Aplicando la ley débil de los grandes números se tiene que  $1 + \frac{P}{n_T} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w^* \xrightarrow{P} 1 + P$ . Por lo tanto, independientemente de la distribución particular de los coeficientes de  $\mathbf{R}_t$ , la eficiencia espectral ergódica asintótica cuando  $\beta \rightarrow \infty$  debe ser igual a  $\log_2(1 + P) \approx 6.65$ .



Las gráficas sugieren que las simulaciones están dentro de las condiciones de operación aceptables. Observemos que la distribución particular de los coeficientes de  $\mathbf{R}_t$  influyen en la velocidad con la que la eficiencia espectral ergódica asintótica converge al valor calculado cuando  $\beta \rightarrow \infty$ .

Es importante notar que, si bien la función implementada retorna 6.56 cuando se tiene distribución Bernoulli y  $\beta = 10$ , para  $\beta = 100$  retorna un valor menor a  $3 \cdot 10^{-9}$ . Por lo tanto, estas funciones deben ser utilizadas con precaución, verificando que el programa funcione dentro de los condiciones de operación deseadas.

Las siguientes tablas muestran una comparación entre la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica.

Bernoulli						Uniforme					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	5.68/5.31	6.16/6.13	6.31/6.33	6.40/6.42	6.45/6.47	1	5.48/5.45	6.05/6.19	6.28/6.39	6.35/6.48	6.42/6.53
2	3.54/3.51	5.47/5.31	5.91/5.91	6.11/6.13	6.24/6.25	2	3.42/3.72	5.29/5.45	5.82/6.05	6.04/6.19	6.19/6.32
3	2.60/2.59	4.29/4.27	5.38/5.31	5.79/5.78	6.01/6.00	3	2.53/2.86	4.16/4.45	5.24/5.45	5.69/5.92	5.93/6.14
4	2.06/2.06	3.51/3.51	4.62/5.00	5.36/5.31	5.71/5.70	4	2.01/2.37	3.41/3.72	4.48/4.76	5.20/5.45	5.58/5.84
5	1.72/1.72	2.98/2.97	3.99/3.98	4.79/4.77	5.34/5.31	5	1.67/2.07	2.90/3.23	3.87/4.18	4.64/4.92	5.20/5.45

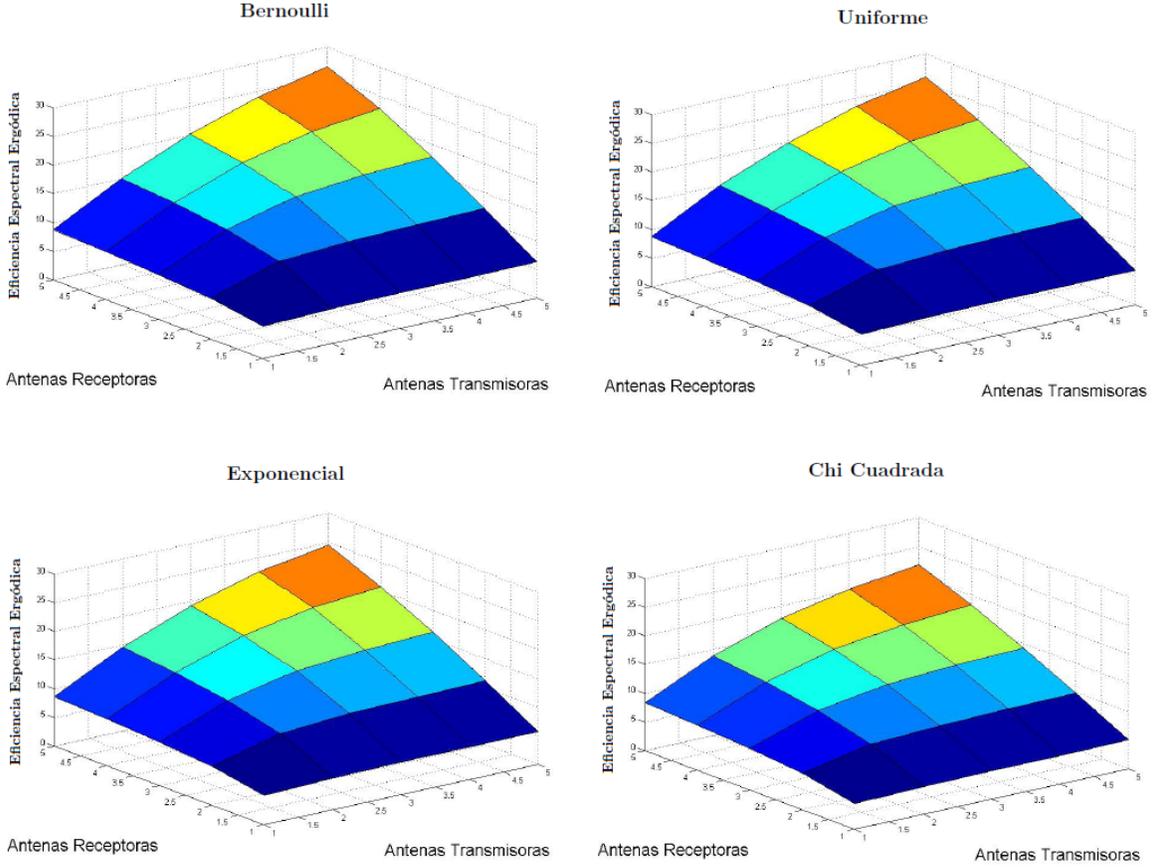
Exponencial						Chi Cuadrada					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	5.20/5.06	5.82/6.04	6.06/6.30	6.21/6.42	6.29/6.49	1	4.54/4.59	5.34/5.78	5.78/6.14	5.95/6.31	6.05/6.40
2	3.24/3.33	5.00/5.06	5.58/5.74	5.83/6.04	6.01/6.20	2	2.87/2.99	4.39/4.59	5.09/5.39	5.46/5.78	5.68/6.00
3	2.40/2.46	3.93/4.05	4.92/5.06	5.40/5.59	5.68/5.87	3	2.11/2.22	3.47/3.63	4.39/4.59	4.90/5.19	5.28/5.55
4	1.92/1.97	3.24/3.33	4.22/4.35	4.91/5.06	5.32/5.48	4	1.70/1.78	2.85/2.99	3.73/3.91	4.37/4.59	4.81/5.07
5	1.60/1.65	2.75/2.83	3.66/3.78	4.38/4.52	4.90/5.06	5	1.43/1.50	2.43/2.54	3.24/3.39	3.85/4.06	4.38/4.59

De las tablas anteriores se observa que la aproximación es buena en general. Un detalle importante que hay que resaltar es que la eficiencia espectral ergódica asintótica es menor que la eficiencia espectral ergódica en el caso de la distribución Bernoulli, mientras que en el resto es mayor.

Un detalle que conviene resaltar es que en el caso de la distribución Bernoulli la eficiencia espectral ergódica converge de manera casi exacta a la eficiencia espectral ergódica asintótica, esto se debe a que la distribución Bernoulli ya es una distribución discreta. En los demás casos se puede ver que cuando  $n_R = n_T$  la eficiencia espectral ergódica tiende a converger a un cierto valor, sin embargo, no converge a la eficiencia espectral ergódica propuesta. Esto claramente es efecto de la discretización, ya que la distribución Bernoulli es discreta mientras que las demás son continuas.

### 3.5.2. Transmisor No Correlacionado

A continuación mostramos una gráfica comparativa de la eficiencia espectral ergódica para las cuatro distribuciones en cuestión. Al igual que antes, la forma de las gráficas es esencialmente la misma para todas las distribuciones salvo variaciones en la escala. En este caso la eficiencia espectral ergódica también es menor que la del canal canónico, lo cual muestra nuevamente el efecto nocivo que tiene la correlación en la eficiencia espectral ergódica.

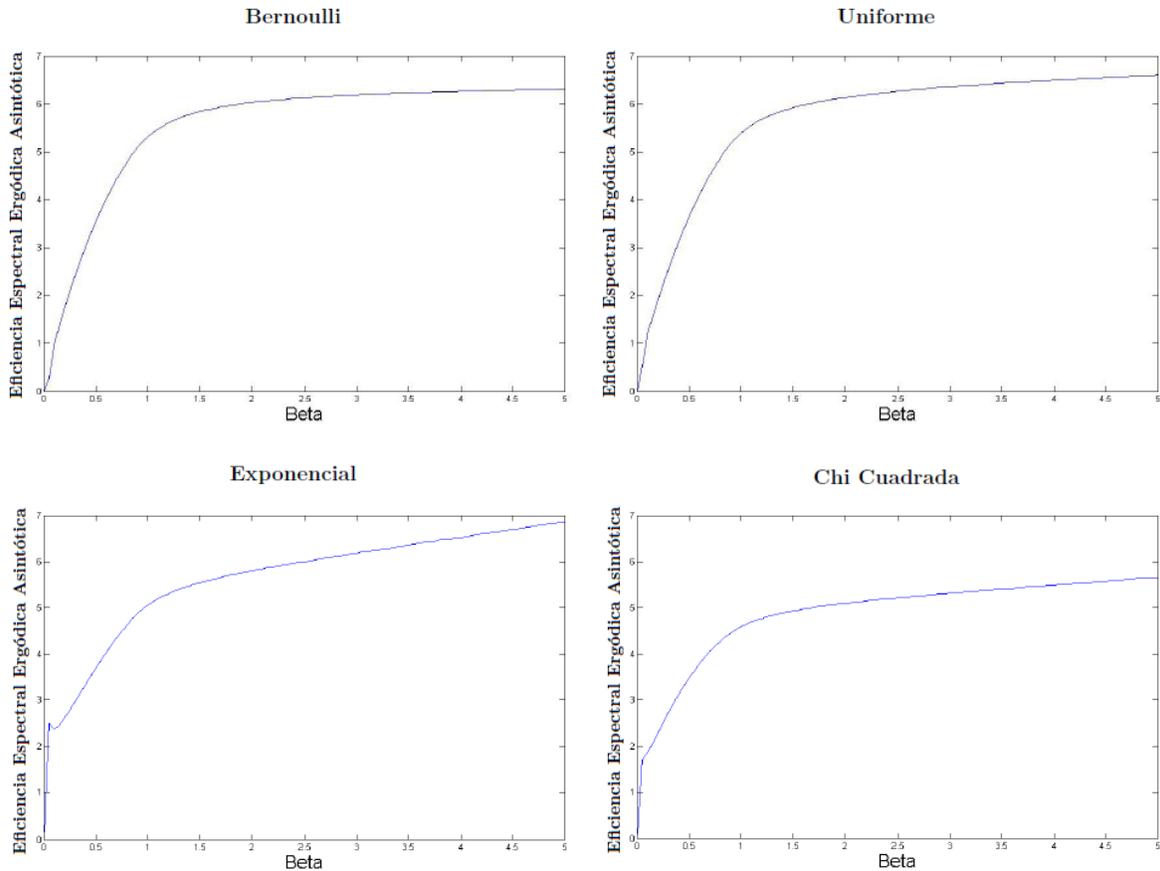


Nuevamente es necesario tener un parámetro que nos permita decidir si el programa que calcula la eficiencia espectral ergódica asintótica está operando de manera aceptable. Al igual que antes, utilizaremos la eficiencia espectral ergódica asintótica cuando  $\beta \rightarrow \infty$ . Si  $n_R = 1$  entonces  $1 + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{H}_w^* \mathbf{R}_r^{1/2} = 1 + \frac{P}{n_T} \mathbf{R}_r \sum_{j=1}^{n_T} h_{1j} h_{1j}^*$ , donde  $\mathbf{R}_r = \text{diag}(d_{r1}, \dots, d_{rn_R})$  y  $\mathbf{H}_w = \{h_{ij}\}$ . Puesto que  $\mathbb{E}[h_{1j} h_{1j}^*] = 1$  para toda  $j$ , aplicando la ley débil de los grandes números se tiene que  $1 + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{H}_w^* \mathbf{R}_r^{1/2} \xrightarrow{p} 1 + P \mathbf{R}_r$ . Por lo tanto, la eficiencia espectral ergódica asintótica cuando  $\beta \rightarrow \infty$  es igual a

$$\mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{H}_w^* \mathbf{R}_r^{1/2} \right) \right] \rightarrow \int \log_2(1 + Px) dR(x), \quad (3.44)$$

donde  $R(x)$  es la distribución de la correlación y  $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w$ . Por lo tanto la eficiencia espectral ergódica asintótica cuando  $\beta \rightarrow \infty$  es aproximadamente 6.46, 6.25, 5.88, y 5.16 en cada caso respectivamente. El hecho de que la eficiencia espectral ergódica asintótica dependa de la distribución de la correlación nos hace pensar, al menos intuitivamente, que en general

la correlación en el receptor es más notoria que la correlación en el transmisor. Las siguientes gráficas muestran la evolución de la eficiencia espectral ergódica asintótica respecto a  $\beta$ .



Es claro que en el caso con correlación exponencial el programa está fuera de las condiciones de operación aceptables<sup>11</sup>. Si bien en el caso de la distribución chi cuadrada el ajuste no es perfecto, aún está dentro de los patrones de forma y escala.

Las siguientes tablas muestran una comparación entre la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica. Si bien el desempeño del programa que calcula la eficiencia espectral ergódica asintótica no fue el óptimo, al menos en el caso de las distribuciones

<sup>11</sup>La correctitud de los programas fue comprobada mediante la comparación de la función de densidad generada por la fórmula de inversión correspondiente y la distribución empírica espectral de matrices de dimensión grande (del orden de  $1000 \times 1000$ ). Por lo tanto, que la gráfica ajuste o no nos dice únicamente si los parámetros con los cuales opera el programa son correctos. En particular, nos dice si el grado del polinomio utilizado es correcto y si el  $\epsilon$  de la fórmula de inversión es adecuado (véase la siguiente sección para más detalles). Entre mayor es el grado del polinomio utilizado es mejor la aproximación, sin embargo, pasado cierto umbral en el grado del polinomio la función de densidad obtenida se aleja demasiado de la función de densidad real, por lo tanto es necesario monitorear el desempeño del programa.

Bernoulli y uniforme, nos da una pauta para comparar la correlación en el transmisor contra la correlación en el receptor.

Bernoulli						Uniforme					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	5.68/5.31	6.08/6.03	6.23/6.19	6.29/6.26	6.32/6.31	1	5.49/5.39	5.90/6.14	6.04/6.36	6.07/6.50	6.09/6.61
2	3.57/3.56	5.45/5.31	5.87/5.84	6.04/6.03	6.13/6.13	2	3.52/3.66	5.29/5.39	5.69/5.92	5.84/6.14	5.94/6.27
3	2.62/2.63	4.34/4.32	5.38/5.31	5.75/5.73	5.94/5.92	3	2.61/2.75	4.26/4.40	5.23/5.39	5.57/5.81	5.74/6.01
4	2.10/2.10	3.55/3.56	4.64/4.63	5.36/5.31	5.68/5.66	4	2.09/2.24	3.51/3.66	4.56/4.72	5.21/5.39	5.50/5.73
5	1.75/1.76	3.01/3.02	4.04/4.03	4.80/4.80	5.34/5.31	5	1.75/1.91	3.00/3.14	3.98/4.12	4.71/4.88	5.21/5.39

Exponencial						Chi Cuadrada					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	5.15/5.06	5.55/5.80	5.67/6.19	5.72/6.52	5.75/6.87	1	4.51/4.59	4.81/5.10	4.95/5.32	5.00/5.49	5.05/5.65
2	3.39/3.70	4.97/5.06	5.35/5.55	5.50/5.80	5.59/6.00	2	3.17/3.48	4.42/4.59	4.69/4.93	4.85/5.10	4.92/5.22
3	2.54/3.09	4.06/4.26	4.93/5.06	5.25/5.43	5.40/5.64	3	2.43/2.86	3.75/3.98	4.39/4.59	4.63/4.85	4.75/5.00
4	2.05/2.79	3.40/3.70	4.35/4.50	4.91/5.06	5.18/5.36	4	1.97/2.52	3.20/3.48	3.98/4.18	4.37/4.59	4.58/4.80
5	1.72/2.62	2.92/3.34	3.83/4.05	4.49/4.64	4.90/5.06	5	1.67/2.29	2.78/3.13	3.57/3.80	4.07/4.29	4.38/4.59

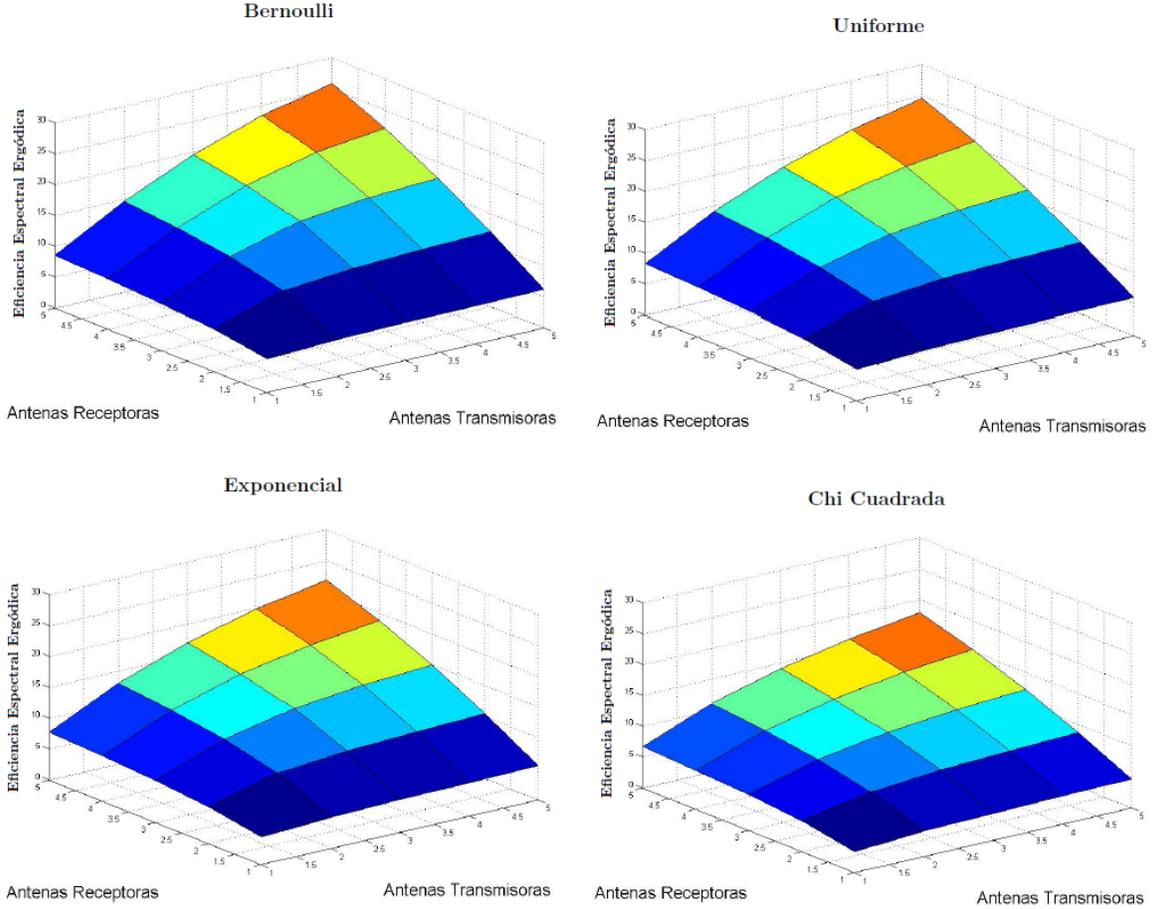
El ajuste mostrado en las tablas anteriores es parcialmente aceptable, al menos en los casos con distribuciones Bernoulli y uniforme. Incluso con estas distribuciones donde se da un mejor ajuste, éste es menos preciso que en el caso del receptor no correlacionado. Esto corrobora hasta cierto punto la intuición obtenida con el análisis de la eficiencia espectral ergódica cuando  $\beta \rightarrow \infty$ , donde en el caso del transmisor no correlacionado la eficiencia espectral ergódica dependía de la distribución de la correlación, lo cual desde un punto de vista intuitivo implica que la correlación agrega más aleatoriedad en este caso que en el anterior.

Algo que se puede comparar directamente de las eficiencias espectrales ergódicas es que en general, si  $n_T > n_R$  entonces es más conveniente no tener correlación presente en el receptor, mientras que si  $n_T < n_R$  es mejor tener un transmisor no correlacionado. Esto es acorde a la intuición, la cual nos dice que el extremo con menos antenas debe ser el extremo con menor correlación, de tal forma que la diversidad debida a las múltiples trayectorias no se vea afectada.

### 3.5.3. Correlación Bilateral

Las siguientes gráficas muestran la eficiencia espectral ergódica de los cuatro casos analizados cuando se tiene correlación bilateral presente en el sistema. Es importante observar

que de los tres casos (correlación en el transmisor, correlación en el receptor y correlación bilateral), éste es el caso donde la eficiencia espectral ergódica es menor, lo cual nos da una idea del efecto de la correlación bilateral. Esto es totalmente natural puesto que la correlación bilateral compromete la existencia de las múltiples trayectorias en ambos extremos.

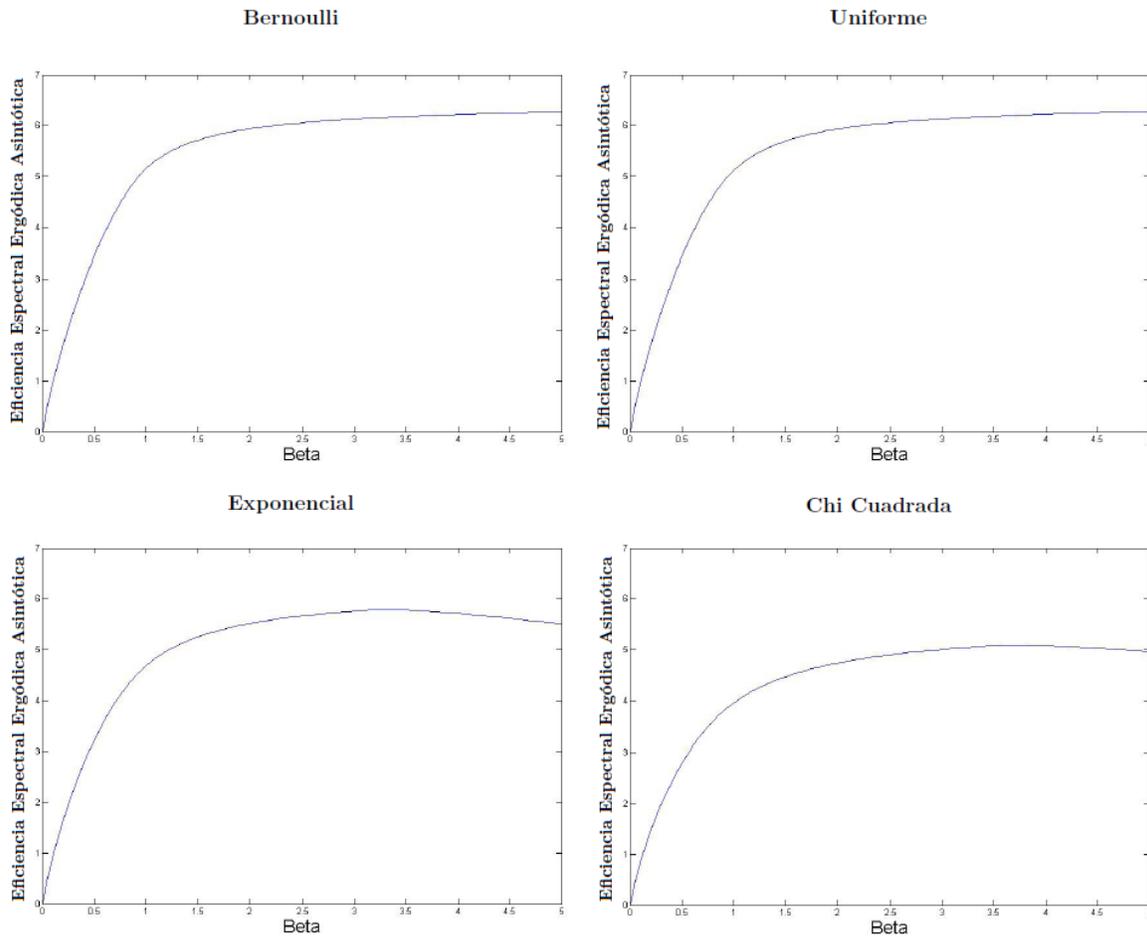


Al igual que en las secciones anteriores, utilizamos la eficiencia espectral ergódica cuando  $\beta \rightarrow \infty$  como variable de control del programa. Si  $n_R = 1$  entonces  $1 + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w^* \mathbf{R}_r^{1/2} = 1 + \frac{P_r}{n_T} \mathbf{R}_r \sum_{j=1}^{n_T} d_{tj} h_{1j} h_{1j}^*$ . Puesto que  $\mathbb{E}[d_{tj} h_{1j} h_{1j}^*] = 1$  para toda  $j$ , la ley débil de los grandes números implica la convergencia de la eficiencia espectral ergódica

$$\mathbb{E}_{\mathbf{H}} \left[ \log_2 \det \left( \mathbf{I} + \frac{P}{n_T} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w^* \mathbf{R}_r^{1/2} \right) \right] \rightarrow \int \log_2(1 + Px) dR(x), \quad (3.45)$$

donde  $R(x)$  es la distribución de la correlación en el receptor y  $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}$ .

Por lo tanto, la eficiencia espectral ergódica cuando  $\beta \rightarrow \infty$  es igual cuando se tiene correlación bilateral que cuando se tiene correlación en el receptor.



Al igual que en el caso con correlación presente en el receptor, aquí hay ciertos problemas numéricos, en especial cuando  $\beta$  crece. Sin embargo, el comportamiento es mejor que en el caso del transmisor no correlacionado, ya que tanto la forma como la escala son más similares a la esperada<sup>12</sup>. A continuación se muestran las tablas comparativas entre la eficiencia espectral ergódica y la eficiencia espectral ergódica asintótica.

<sup>12</sup>Es importante observar que los máximos de las gráficas de las distribuciones exponencial y chi cuadrada están a menos de 0.1 del valor calculado cuando  $\beta \rightarrow \infty$ . Esto da indicios de posibles problemas con el  $\epsilon$  (véase la siguiente sección para consultar los detalles de implementación), el cual al crecer  $\beta$  se vuelve inapropiado para obtener la solución con la suficiente precisión. Claramente hay errores por la discretización de la distribución de la correlación, pero este efecto es mínimo comparado con el efecto relacionado a la selección de  $\epsilon$ .

Bernoulli						Uniforme					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	5.51/5.15	5.97/5.94	6.12/6.13	6.20/6.22	6.25/6.27	1	5.14/5.12	5.69/5.93	5.86/6.13	5.99/6.22	6.01/6.28
2	3.46/3.46	5.27/5.15	5.72/5.71	5.93/5.94	6.06/6.06	2	3.31/3.45	4.95/5.12	5.46/5.70	5.66/5.93	5.80/6.05
3	2.55/2.56	4.20/4.19	5.22/5.15	5.61/5.59	5.80/5.81	3	2.47/2.56	4.00/4.17	4.90/5.12	5.33/5.57	5.55/5.80
4	2.04/2.05	3.45/3.46	4.50/4.49	5.20/5.15	5.51/5.51	4	1.99/2.05	3.32/3.45	4.30/4.47	4.90/5.12	5.27/5.48
5	1.71/1.71	2.94/2.94	3.91/3.91	4.66/4.65	5.16/5.15	5	1.66/1.72	2.84/2.94	3.74/3.90	4.42/4.63	4.91/5.12

Exponencial						Chi Cuadrada					
$n_R \backslash n_T$	1	2	3	4	5	$n_R \backslash n_T$	1	2	3	4	5
1	4.53/4.68	5.08/5.57	5.33/5.76	5.43/5.72	5.50/5.51	1	3.46/3.95	4.17/4.74	4.36/5.01	4.62/5.08	4.67/4.96
2	3.01/3.23	4.43/4.68	4.88/5.25	5.13/5.52	5.29/5.66	2	2.48/2.80	3.51/3.95	4.00/4.47	4.24/4.74	4.40/4.90
3	2.28/2.43	3.61/3.86	4.39/4.68	4.77/5.12	4.99/5.36	3	1.93/2.15	2.94/3.30	3.53/3.95	3.88/4.34	4.12/4.58
4	1.84/1.96	3.03/3.23	3.84/4.12	4.36/4.68	4.69/5.03	4	1.58/1.74	2.54/2.80	3.15/3.49	3.57/3.95	3.85/4.26
5	1.55/1.65	2.60/2.78	3.40/3.63	3.97/4.25	4.38/4.68	5	1.35/1.48	2.21/2.43	2.81/3.11	3.26/3.60	3.57/3.95

Tomando en cuenta que en esta situación la aleatoriedad proviene de  $\mathbf{H}_w$ ,  $\mathbf{R}_r$  y  $\mathbf{R}_t$  las aproximaciones obtenidas son buenas en general. En este caso la eficiencia espectral ergódica es aún menor que en el caso con transmisor no correlacionado, lo cual muestra que en términos de la eficiencia espectral ergódica éste es el peor escenario.

Dado que la correlación presente únicamente en el receptor es un caso particular de la correlación bilateral, los programas para la correlación bilateral pueden usarse para resolver el caso con el transmisor no correlacionado. La única desventaja, es que el método para resolver el caso con correlación bilateral es iterativo y no se tiene la certeza de que el algoritmo va a converger en algún momento. En otras palabras, a cambio de no tener la certeza teórica de que el algoritmo funciona tenemos un mejor comportamiento numérico, lo contrario de la Sección 3.5.2, donde se tenía la certeza teórica del funcionamiento del algoritmo pero no se tenía un buen comportamiento numérico.

### 3.6. Cálculo Numérico de la Eficiencia Espectral Ergódica Isotrópica

En esta sección presentamos la implementación de los teoremas y técnicas expuestas en este capítulo. La correlación unilateral se calcula utilizando dos programas diferentes, uno para el caso con correlación en el transmisor y otro para el caso con correlación en el receptor. Sin embargo, ambos programas son muy similares, ya que resuelven el polinomio correspondiente

para cada caso (véanse las ecuaciones (3.17) y (3.27)). De manera similar para el caso del canal de Rice se implementa la solución del polinomio equivalente en dicho caso (véase la ecuación (3.43)). Para el caso del canal con correlación bilateral se implementa la ecuación de punto fijo dada por el Corolario C.3.

Si bien el programa que calcula las transformadas de Stieltjes es el mismo para todas las distribuciones elegidas para la correlación, los programas que lo utilizan para calcular la eficiencia espectral ergódica son diferentes en cada caso, dependiendo de la distribución de la correlación. Por lo tanto, ejemplificamos el cálculo de las eficiencias espectrales ergódicas para la distribución chi cuadrada, aunque los programas respectivos para las demás distribuciones son prácticamente idénticos.

### 3.6.1. Receptor No Correlacionado

A continuación se presentan los programas con los cuales se calcularon las gráficas para el caso en el que no hay correlación presente en el receptor. Además de los programas se exponen los detalles de implementación más relevantes.

#### Eficiencia Espectral Ergódica

Este programa, el cual utiliza el método de Monte Carlo, simula **nsimulaciones** matrices de propagación con las matrices de correlación correspondientes. Después de obtener los eigenvalores de  $\mathbf{H}\mathbf{H}^*$  se agrega la eficiencia espectral ergódica obtenida para dicha realización al acumulado, el cual tras ser dividido regresa el promedio de las eficiencias espectrales ergódicas. Por la ley débil de los grandes números este procedimiento produce el resultado deseado cuando  $\mathbf{nsimulaciones} \rightarrow \infty$ .

```

1 function varret=EficienciaErgodicaRNC_Chi2(n_R,n_T,P,nsimulaciones)
2 % Esta funcion aproxima la eficiencia espectral ergodica de un
3 % canal MIMO de n_Rxn_T el cual tiene correlacion unilateral en
4 % el transmisor, dicha correlacion tiene una distribucion chi2.
5 varret=0;
6
7 for it=1:nsimulaciones
8     % La matriz de coeficientes de propagacion.
9     H=random('norm',0,1/sqrt(2),n_R,n_T)+i*random('norm',0,1/sqrt(2),n_R,n_T);
10    D=random('chi2',1,1,n_T);
11    Rt=diag(D);
12
13    % Obtener los eigenvalores de HH*.
14    E=eig(H*Rt*H'/n_R);

```

```

15
16     % Agregar la eficiencia espectral ergodica obtenida para la rea-
17     % lizacion en cuestion.
18     varret=varret+sum(log2(1+P*(n.R/n.T)*E));
19 end
20
21 varret=varret/nsimulaciones;

```

## Eficiencia Espectral Ergódica Asintótica

La estimación de la eficiencia espectral ergódica se realiza mediante la estimación de la función de densidad espectral correspondiente, para posteriormente encontrar el valor esperado que describe la eficiencia espectral ergódica por antena receptora. La estimación de la eficiencia espectral ergódica se realiza mediante los siguientes cuatro programas<sup>13</sup>.

```

1 function varret=EficienciaAsintoticaRNC_Chi2(beta,P)
2 % Esta funcion calcula la eficiencia espectral ergodica asintotica
3 % para una beta y P dadas cuando el receptor no presenta correlacion
4 % y el transmisor tiene correlacion con distribucion chi2.
5 varret=0;
6
7 % Se construye la discretizacion de la distribucion.
8 n=16;
9 r=[];
10 for it=1:(n-1)
11     r=[r chi2inv(it/n,1)]; %ok<AGROW>
12 end
13 r=[r chi2inv(1-1/(2*n),1)];
14 p=ones(1,n)/n; % Se asigna una distribucion uniforme.
15
16 % Se evalua la integral que proporciona la eficiencia espectral
17 % ergodica asintotica.
18 for x=0.001:0.001:30
19     varret=varret+InversionRNC(x,beta,r,p)*0.001*log2(1+(P/beta)*x);
20 end

```

La función de inversión evalúa  $m_{\mathbf{B}}(z)$  en un punto con  $\text{Im}(z)$  pequeño y retorna el valor de la parte imaginaria de  $m_{\mathbf{B}}(z)$ . Es importante notar que este límite se puede tomar de esta manera dado que el límite involucrado en la fórmula de inversión no implica la evaluación de expresiones complicadas numéricamente hablando<sup>14</sup>.

```

1 function den=InversionRNC(x,beta,r,p)
2 % Este programa implementa la formula de inversion.

```

<sup>13</sup>Las simulaciones se realizaron en un equipo de cómputo QuadCore a 2.33Ghz con 4Gb de memoria RAM. Con dicho equipo la obtención de la eficiencia espectral asintótica para una  $\beta$  y una potencia  $P$  dadas toma aproximadamente minuto y medio. Por lo tanto, cada gráfica que muestra el comportamiento de la eficiencia espectral ergódica asintótica con respecto a  $\beta$  requiere aproximadamente dos horas y media para su realización.

<sup>14</sup>Un ejemplo típico de un límite más difícil de evaluar numéricamente es el cálculo de una derivada. Por lo general, esto implica que se evalúen fracciones con numerador y denominador muy pequeños, lo cual es más propenso a tener errores numéricos.

```

3 |
4 | epsilon=0.000001;
5 |
6 | den=StieltjesRNC(x+epsilon*i , beta , r , p)/pi;

```

Es importante observar que esta implementación de la fórmula de inversión no toma la parte imaginaria de la transformada de Stieltjes porque el programa que calcula la transformada de Stieltjes lo hace.

La función `StieltjesRNC` almacena el polinomio dado por la ecuación (3.17) (el cual es creado por la función `PolinomioRNC`), encuentra sus raíces y de las que tienen parte imaginaria positiva selecciona la que tiene parte imaginaria mayor. El Teorema C.7 afirma que dicha raíz debe ser única, sin embargo por posibles errores numéricos puede haber ninguna o más de una. Para prevenir dichos errores numéricos se toma la raíz que tiene la parte imaginaria mayor.

Por último, la función `PolinomioRNC` crea el polinomio de la ecuación (3.17). El diagrama de flujo de la Figura 3.1 muestra el algoritmo para construir dicho polinomio. El polinomio se construye mediante la suma de otros tres polinomios más simples, los cuales denotaremos como

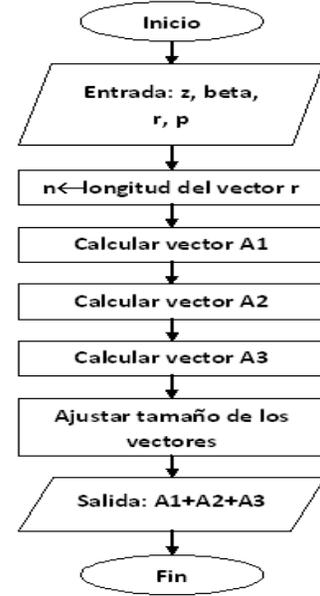


Figura 3.1: Diagrama de flujo de la función `PolinomioRNC`.

$$A1 = \prod_{i=1}^n (1 + r_i m_{\mathbf{B}}(z)), \quad (3.46)$$

$$A2 = -m_{\mathbf{B}}(z) \left( -z \prod_{i=1}^n (1 + r_i m_{\mathbf{B}}(z)) \right), \quad (3.47)$$

$$A3 = -m_{\mathbf{B}}(z) \left( \beta \sum_{i=1}^n r_i p_i \prod_{j \neq i} (1 + r_j m_{\mathbf{B}}(z)) \right). \quad (3.48)$$

Para construir los polinomios  $A1$ ,  $A2$  y  $A3$  se utiliza la función `poly` de Matlab, la cual crea un polinomio a partir de sus raíces. Es importante hacer que los polinomios tengan los coeficientes deseados, esto se puede lograr controlando el término constante (de manera predeterminada Matlab simplifica el polinomio). Esto se realiza en el momento en el que se

divide el polinomio en cuestión por su término constante (de la ecuación 3.17 se obtiene que el término constante debe ser uno).

Es importante observar que en los tres casos (receptor no correlacionado, transmisor no correlacionado y canal de Rice) el término constante no es cero puesto que es producto de números complejos no nulos, lo cual asegura el buen funcionamiento del programa.

```

1 function solucion=StieltjesRNC(z,beta,r,p)
2 % Esta funcion retorna la parte imaginaria de la transformada de
3 % Stieltjes en z para una beta dada cuando se tiene que la correlacion
4 % tiene distribucion en r con probabilidades p.
5
6 % Obtener las raices del polinomio.
7 raiz=roots(PolinomioRNC(z,beta,r,p));
8
9 solucion=0;
10
11 % Elegir la raiz correcta.
12 for it=1:length(raiz)
13     if imag(raiz(it))>solucion
14         solucion=imag(raiz(it));
15     end
16 end

```

```

1 function poli=PolinomioRNC(z,beta,r,p)
2 % Esta funcion crea el polinomio requerido para calcular la
3 % transformada de Stieltjes.
4
5 n=length(r);
6
7 if n==1
8     poli=[
9         r*z
10        r-beta*r+z
11        1
12        ];
13 else
14     % Calcular el producto
15     % \prod_{i=1}^n(1+r\cdot im_{\{bf B\}}(z))
16     raices=[];
17     for it=1:n
18         if (r(it)~=0)
19             raices=[raices -1/r(it)];
20         end
21     end
22     A1=poly(raices);
23     while length(A1)<n+1
24         A1=[0 A1];
25     end
26     A1=A1/A1(n+1);
27
28     % Calcular el termino
29     % z\cdot_{\{bf B\}}(z)\prod_{i=1}^n(1+r\cdot im_{\{bf B\}}(z))
30     A2=-z*A1;
31     A2=-[A2 0];
32
33     % Calcular el termino

```

```

34 %-\beta m_{\bf B}(z)\sum_{i=1}^n r_{ip-i}
35 %\prod_{j\neq i}(1+r_{im_{\bf B}}(z))
36 A3=zeros(1,n+1);
37
38 for ita=1:n
39     raices=[];
40     for itb=1:n
41         if itb~=ita && r(itb)~=0
42             raices=[raices -1/r(itb)];
43         end
44     end
45     An=poly(raices);
46     while length(An)<n+1
47         An=[0 An];
48     end
49     An=An/An(n+1);
50     An=beta*r(ita)*p(ita)*An;
51
52     A3=A3+An;
53 end
54
55 % Ajustar el grado de los polinomios.
56 A3=-[A3 0];
57 A1=[0 A1];
58
59 % Sumar los terminos del polinomio.
60 poli=A1+A2+A3;
61 end

```

### 3.6.2. Transmisor No Correlacionado

Dado que el código de esta sección es totalmente análogo al de la sección anterior (la única diferencia reside en el polinomio que se está resolviendo, ya que en lugar de la ecuación (3.17) se implementa la ecuación (3.27)), solamente se comentan los detalles particulares de este caso.

#### Eficiencia Espectral Ergódica

Este caso es totalmente análogo al caso donde el receptor no presenta correlación, con la diferencia que aquí se multiplica por  $\mathbf{R}_r$  en lugar de  $\mathbf{R}_t$ .

```

1 function varret=EficienciaErgodicaTNC_Chi2(n_R,n_T,P,nsimulaciones)
2
3 varret=0;
4
5 for it=1:nsimulaciones
6     H=random('norm',0,1/sqrt(2),n_R,n_T)+i*random('norm',0,1/sqrt(2),n_R,n_T);
7     D=random('chi2',1,1,n_R);
8     Rr=diag(D);
9     E=eig(Rr^(1/2)*(H*H')*Rr^(1/2)/n_T);
10

```

```

11     varret=varret+sum(log2(1+P*E));
12 end
13
14 varret=varret/nsimulaciones;

```

## Eficiencia Espectral Ergódica Asintótica

En este caso hay que tomar en cuenta la condición de unicidad proporcionada por el Teorema C.8 en lugar de la condición proporcionada por el Teorema C.7.

```

1 function varret=EficienciaAsintoticaTNC_Chi2(beta,P)
2
3 varret=0;
4
5 n=16;
6 r=[];
7 for it=1:(n-1)
8     r=[r chi2inv(it/n,1)];
9 end
10 r=[r chi2inv(1-1/(2*n),1)];
11 p=ones(1,n)/n;
12
13 for x=0.001:0.001:30
14     varret=varret+InversionTNC(x,beta,r,p)*0.001*log2(1+P*x);
15 end

```

```

1 function den=InversionTNC(x,beta,r,p)
2
3 epsilon=0.000001;
4
5 den=StieltjesTNC(x+epsilon*i,beta,r,p)/pi;

```

```

1 function solucion=StieltjesTNC(z,beta,r,p)
2
3 c=1/beta;
4
5 raiz=roots(PolinomioTNC(z,c,r,p));
6
7 solucion=0;
8
9 for it=1:length(raiz)
10     if imag(-(1-c)/z+c*raiz(it))>0 && imag(raiz(it))>solucion
11         solucion=imag(raiz(it));
12     end
13 end

```

Aquí la verificación del término constante se realiza mediante la variable escala puesto que en general este coeficiente no es uno.

```

1 function poli=PolinomioTNC(z,c,r,p)
2
3 n=length(r);
4
5 if n==1

```

```

6     poli=[
7         -r*c*z
8         r-c*r-z
9         1
10        ];
11    else
12        raices=[];
13        escala=1;
14        for it=1:n
15            escala=escala*(r(it)-r(it)*c-z);
16            if (r(it)~=0)
17                raices=[raices (r(it)-r(it)*c-z)/(r(it)*c*z)];
18            end
19        end
20        A1=poly(raices);
21        while length(A1)<n+1
22            A1=[0 A1];
23        end
24        A1=A1/A1(n+1)*escala;
25        A1=[A1 0];
26
27        A2=zeros(1,n+2);
28        for ita=1:n
29            raices=[];
30            escala=1;
31            for itb=1:n
32                if itb~=ita
33                    escala=escala*(r(itb)-r(itb)*c-z);
34                    if r(itb)~=0
35                        raices=[raices (r(itb)-r(itb)*c-z)/(r(itb)*c*z)];
36                    end
37                end
38            end
39            An=poly(raices);
40            while length(An)<n+2
41                An=[0 An];
42            end
43            An=An/An(n+2)*escala;
44            An=p(ita)*An;
45
46            A2=A2+An;
47        end
48
49        poli=A1-A2;
50    end

```

### 3.6.3. Correlación Bilateral

En este apartado se compilan los programas utilizados para el caso con correlación bilateral. En esta situación no se implementan métodos basados en polinomios, en su lugar se utiliza la ecuación de punto fijo proporcionada por el Corolario C.3.

## Eficiencia Espectral Ergódica

Para evaluar la eficiencia espectral ergódica se utiliza un código idéntico a los casos con correlación unilateral, salvo que aquí se tienen dos matrices de correlación en lugar de una.

```
1 function varret=EficienciaErgodicaCB_Chi2(n_R,n_T,P,nsimulaciones)
2
3 varret=0;
4
5 for it=1:nsimulaciones
6     H=random('norm',0,1/sqrt(2),n_R,n_T)+i*random('norm',0,1/sqrt(2),n_R,n_T);
7     D=random('chi2',1,1,n_R);
8     Rr=diag(D);
9     D=random('chi2',1,1,n_T);
10    Rt=diag(D);
11    E=eig(Rr^(1/2)*H*Rt*H'*Rr^(1/2)/n_R);
12
13    varret=varret+sum(log2(1+P*(n_R/n_T)*E));
14 end
15
16 varret=varret/nsimulaciones;
```

## Eficiencia Espectral Ergódica Asintótica

El cálculo de la correlación bilateral se realiza usando el Corolario C.3. Dado que una de las hipótesis de dicho corolario es que las distribuciones empíricas asintóticas de las matrices  $\mathbf{R}_r$  y  $\mathbf{R}_t$  tienen soporte compacto, tendremos que trabajar siempre con discretizaciones de las distribuciones reales.

```
1 function varret=EficienciaAsintoticaCB_Chi2(beta,P)
2
3 varret=0;
4
5 n=16;
6 r=[];
7 for it=1:(n-1)
8     r=[r chi2inv(it/n,1)];
9 end
10 r=[r chi2inv(1-1/(2*n),1)];
11 p=ones(1,n)/n;
12
13 for x=0.001:0.001:20
14     varret=varret+InversionCB(x,beta,r,p,r,p)*0.001*log2(1+(P/beta)*x);
15 end
```

El cálculo de la densidad de los eigenvalores se realiza mediante los siguientes dos programas. El primero únicamente implementa la fórmula de inversión para la transformada de Stieltjes y el segundo calcula la ecuación de punto fijo dada por el Corolario C.3.

```
1 function den=InversionCB(x,beta,rr,pr,rt,pt)
2
```

```

3 eps=0.000001;
4
5 den=imag(StieltjesCB(x+i*eps,beta,rr,pr,rt,pt))/pi;

```

La obtención de la transformada de Stieltjes en este caso es una implementación directa de la ecuación de punto fijo del Corolario C.3. Para el adecuado funcionamiento de la función StieltjesCB se debe crear una variable global  $\mathbf{B}$ , la cual tenga la misma cantidad de elementos que el grado de aproximación de  $\mathbf{R}_r$  (número de valores en la discretización). Puesto que en el programa EficienciaAsintoticaCB\_Chi2 se calculan las transformadas de Stieltjes para valores sucesivos de  $x$ , la variable global  $\mathbf{B}$  nos permite utilizar la solución anterior a la ecuación de punto fijo para calcular la siguiente.

```

1 function valor=StieltjesCB(z,beta,rr,pr,rt,pt)
2 % Esta funcion calcula el valor de la transformada de Stieltjes
3 % mediante la implementacion del Corolario C.3
4
5 global B;
6
7 valor=0;% Esta variable retorna el valor esperado final
8 % \mathbb{E}\left[\mathcal{B}(D,z)\right].
9 var=1;% Variacion de B entre iteraciones sucesivas.
10 epsilon=0.001;% Variacion umbral para detener iteraciones.
11
12 while var>epsilon
13     var=0;
14
15     for it=1:length(B)
16
17         A1=0;
18         for ita=1:length(rr)
19             A1=A1+rr(ita)*pr(ita)*B(ita);
20         end
21
22         A2=0;
23         for ita=1:length(rt)
24             A2=A2+rt(ita)*pt(ita)/(z+z*rt(ita)*A1);
25         end
26
27         temp=-z+z*beta*rr(it)*A2;
28         temp=1/temp;
29
30         var=var+norm(B(it)-temp);
31         B(it)=temp;
32     end
33 end
34
35 for it=1:length(rr)
36     valor=valor+pr(it)*B(it);
37 end

```

En este programa el  $\epsilon$  que se toma para detener las iteraciones es igual a 0.001. Claramente entre menor sea  $\epsilon$ , mejor será la aproximación y mayor será el tiempo de ejecución requerido

por el programa. Por lo tanto, es necesario ajustar el  $\epsilon$  teniendo en cuenta dicho intercambio precisión-tiempo de ejecución.

### 3.6.4. Canal de Rice

Como se mencionó con anterioridad, dado que el canal de Rice no es un canal correlacionado, analizar la eficiencia espectral ergódica de este canal queda fuera del objetivo de esta tesis. Sin embargo, por completez incluimos los programas necesarios para obtener la función de densidad del espectro asintótico de un canal de Rice no correlacionado.

#### Eficiencia Espectral Ergódica Asintótica

Esta implementación es muy similar a la implementación para el caso de la correlación unilateral, pero en este caso implementando la ecuación (3.43). La similitud de estas tres implementaciones (receptor no correlacionado, transmisor no correlacionado y canal de Rice) es muy fuerte debido a la relación tan estrecha entre los Teoremas C.7, C.8 y C.9.

```

1 function den=InversionR(x,beta,r,p)
2
3 epsilon=0.000001;
4 den=StieltjesR(x+epsilon*i,beta,r,p)/pi;

```

```

1 function solucion=StieltjesR(z,beta,r,p)
2
3 c=1/beta;
4
5 raiz=roots(PolinomioR(z,c,r,p));
6
7 solucion=0;
8
9 for it=1:length(raiz)
10     if imag(raiz(it)*z)>=0 && imag(raiz(it))>solucion
11         solucion=imag(raiz(it));
12     end
13 end

```

```

1 function poli=PolinomioR(z,c,r,p)
2
3 n=length(r);
4
5 if n==1
6     poli=[
7         -c^2*z
8         c-c^2-2*c*z
9         r-z+1-c-p*c
10        -p
11        ];
12 else

```

```

13 raices=[];
14 escala=1;
15 for it=1:n
16     Ai=-c^2*z;
17     Bi=c-c^2-2*c*z;
18     Ci=r(it)-z+1-c;
19     escala=escala*Ci;
20     raiz1=(-Bi+sqrt(Bi^2-4*Ai*Ci))/(2*Ai);
21     raiz2=(-Bi-sqrt(Bi^2-4*Ai*Ci))/(2*Ai);
22     if (raiz1~=0)
23         raices=[raices raiz1];
24     end
25     if (raiz2~=0)
26         raices=[raices raiz2];
27     end
28 end
29 A1=poly(raices);
30 while length(A1)<2*n+1
31     A1=[0 A1];
32 end
33 A1=A1/A1(2*n+1)*escala;
34 A1=[A1 0];
35
36 A2=zeros(1,2*n);
37 for ita=1:n
38     raices=[];
39     escala=1;
40     for itb=1:n
41         if itb~=ita
42             Ai=-c^2*z;
43             Bi=c-c^2-2*c*z;
44             Ci=r(itb)-z+1-c;
45             escala=escala*Ci;
46             raiz1=(-Bi+sqrt(Bi^2-4*Ai*Ci))/(2*Ai);
47             raiz2=(-Bi-sqrt(Bi^2-4*Ai*Ci))/(2*Ai);
48             if (raiz1~=0)
49                 raices=[raices raiz1];
50             end
51             if (raiz2~=0)
52                 raices=[raices raiz2];
53             end
54         end
55     end
56     raices=[raices -1/c];
57     An=poly(raices);
58     while length(An)<2*n
59         An=[0 An];
60     end
61     An=An/An(2*n)*escala;
62     An=p(ita)*An;
63
64     A2=A2+An;
65 end
66 A2=[0 0 A2];
67
68 poli=A1-A2;
69 end

```



## Capítulo 4

# Conclusiones

En esta tesis se analizaron sistemas MIMO con correlación de Kronecker en los que las matrices de correlación del modelo (ecuación 3.1) tienen eigenvalores independientes idénticamente distribuidos. En este contexto y motivados por el caso en el que la distribución de dichos eigenvalores toma una cantidad finita de valores, desarrollamos un método para calcular la eficiencia espectral ergódica asintótica por antena receptora para canales en los que los eigenvalores de las matrices de correlación tienen una distribución continua.

La prueba de la validez de este método, el cual está basado en la discretización de la función de distribución de los eigenvalores de las matrices de correlación, está dada por el Teorema 3.2. Para asegurarnos del correcto funcionamiento de la implementación de este método, la cual se realizó en Matlab, utilizamos la eficiencia espectral ergódica asintótica por antena receptora cuando  $\beta \rightarrow \infty$ . Esta cantidad, la cual se puede calcular analíticamente, se utilizó como parámetro de comparación para los valores retornados por la implementación, y así decidir si los parámetros utilizados al ejecutar el programa fueron los correctos.

Finalmente, se observó que dentro de los límites analizados (una relación señal a ruido igual a cien), la propuesta de estimar la eficiencia espectral ergódica por antena receptora mediante la eficiencia espectral ergódica asintótica por antena receptora es aceptable.



# Apéndice A

## Teoría Ergódica y Algunos

## Resultados de Procesos Estocásticos

En la primer sección de este apéndice se resumen algunos de los conceptos y resultados de probabilidad y procesos estocásticos utilizados en esta tesis. Para una exposición más detallada sobre probabilidad se recomienda al lector consultar el libro de Hoel, Port y Stone [Hoe71]. En las últimas dos secciones de este apéndice se da una breve introducción a la teoría ergódica y se discute la relevancia de esta teoría en el análisis de canales de comunicación.

### A.1. Nociones Básicas de Probabilidad y Procesos Estocásticos

**Definición A.1.** Una colección no vacía  $\mathbb{F}$  de subconjuntos de un conjunto  $\Omega$  se denomina  $\sigma$ -álgebra de subconjuntos de  $\Omega$  si las siguientes dos propiedades se satisfacen:

- Si  $A \in \mathbb{F}$  entonces  $A^c \in \mathbb{F}$ .
- Si  $A_1, A_2, \dots \in \mathbb{F}$ , entonces  $\bigcup_{n=1}^{\infty} A_n \in \mathbb{F}$ .

Una consecuencia directa de la definición de  $\sigma$ -álgebra es que si  $A_1, A_2, \dots \in \mathbb{F}$  entonces  $\bigcap_{n=1}^{\infty} A_n \in \mathbb{F}$ . Además, debido a que  $\mathbb{F}$  es no vacía y es cerrada bajo complementos, se tiene que  $\emptyset, \Omega \in \mathbb{F}$ .

**Definición A.2.** Una medida de probabilidad  $\mathbb{P}$  en una  $\sigma$ -álgebra  $\mathbb{F}$  de un conjunto  $\Omega$  es una función  $\mathbb{P} : \mathbb{F} \rightarrow \mathbb{R}^+ \cup \{0\}$  tal que

- $\mathbb{P}(\emptyset) = 0$  y  $\mathbb{P}(\Omega) = 1$ .
- Si  $A_1, A_2, \dots$  son mutuamente disjuntos, esto es,  $A_i \cap A_j = \emptyset$  si  $i \neq j$ , entonces

$$\mathbb{P} \left( \bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \quad (\text{A.1})$$

**Definición A.3.** Un espacio de probabilidad, denotado por  $(\Omega, \mathbb{F}, \mathbb{P})$ , es un conjunto  $\Omega$  dotado con una  $\sigma$ -álgebra  $\mathbb{F}$  y una medida de probabilidad  $\mathbb{P}$  definida en  $\mathbb{F}$ .

Con objeto de motivar las definiciones anteriores, considérese el siguiente ejemplo. Supóngase que se tiene un componente electrónico y estamos interesados en estudiar su tiempo de vida, claramente este tiempo será aleatorio. Suponiendo que el tiempo en el que se adquirió es el tiempo cero, entonces el tiempo de vida estará en el intervalo  $[0, \infty)$ , por lo que en este caso tomamos  $\Omega = [0, \infty)$ . Nosotros estamos interesados en hacer preguntas sobre la probabilidad de ciertos sucesos o eventos, por ejemplo, la probabilidad de que el componente dure entre  $x$  años y  $y$  años. De hecho, en este ejemplo particular podemos tomar como  $\mathbb{F}$  a la  $\sigma$ -álgebra generada<sup>1</sup> por los intervalos abiertos en  $\Omega$ . En general, la  $\sigma$ -álgebra  $\mathbb{F}$  se puede interpretar como un conjunto, el cual contiene las *preguntas* que son relevantes para el problema.

Dado que las preguntas se hacen sobre los conjuntos de  $\mathbb{F}$ , cualquier función que vaya de  $\Omega$  a un espacio de la forma  $(S, B(S))$  donde  $S$  es un conjunto y  $B(S)$  una  $\sigma$ -álgebra<sup>2</sup> en  $S$ , deberá ser tal que  $X^{-1}(A) \in \mathbb{F}$  para todo  $A \in B(S)$ . Esto es, la imagen inversa de los conjuntos de interés (en  $B(S)$ ) debe tener una probabilidad bien definida. La formalización de tal función es la siguiente.

**Definición A.4.** Sean  $(\Omega, \mathbb{F}, \mathbb{P})$  un espacio de probabilidad y  $(S, B(S))$  un conjunto y una  $\sigma$ -álgebra en  $S$ . Se define una variable aleatoria (v.a.) como una función  $X : \Omega \rightarrow S$  tal que  $X^{-1}(A) \in \mathbb{F}$  para toda  $A \in B(S)$ .

Los siguientes son casos particulares de variables aleatorias.

---

<sup>1</sup>Si  $E$  es una clase de subconjuntos de  $\Omega$  entonces la  $\sigma$ -álgebra generada por  $E$ , la cual denotamos por  $\sigma(E)$ , es la menor  $\sigma$ -álgebra que contiene a  $E$ .

<sup>2</sup>Aquí se puede interpretar  $B(S)$  como la clase de conjuntos de interés en  $S$ .

**Definición A.5.** Decimos que una v.a. es una v.a. discreta si toma una cantidad a lo más numerable de valores<sup>3</sup> y además  $B(S)$  es el conjunto potencia de  $S$ .

**Definición A.6.** Una v.a.  $X$  es una v.a. real si  $S = \mathbb{R}$  y es tal que  $\{\omega \in \Omega : X(\omega) \leq x\}$  es un evento para toda  $x \in \mathbb{R}$ . A la función  $F : \mathbb{R} \rightarrow \mathbb{R}$  la cual está dada por

$$F(x) = \mathbb{P}(X \leq x)$$

se le conoce como función de distribución de  $X$ .

**Definición A.7.** Una v.a. real es una v.a. continua si  $P(X = x) = 0$  para toda  $x \in (-\infty, \infty)$ . Equivalentemente se dice que  $X$  es una v.a. continua si su función de distribución es continua.

De manera similar se define una v.a. compleja como una v.a. que toma valores en los complejos y su  $\sigma$ -álgebra asociada es la  $\sigma$ -álgebra generada por los intervalos abiertos inducidos por la métrica estándar<sup>4</sup>.

**Definición A.8.** Sea  $X$  una v.a. continua. Si existe una función determinista  $f : \mathbb{R} \rightarrow \mathbb{R}^+ \cup \{0\}$  Lebesgue integrable tal que para toda  $a, b \in \mathbb{R}$  con  $a \leq b$  se satisface que

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx, \quad (\text{A.2})$$

entonces a la función  $f$  se le llama la función de densidad de probabilidad de  $X$ , y se dice que  $X$  es absolutamente continua.

**Definición A.9.** Se define el valor esperado  $\mathbb{E}[X]$  de una v.a. absolutamente continua  $X$  con función de densidad  $f$ , como

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{A.3})$$

siempre y cuando  $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$ .

**Definición A.10.** Sea  $X$  una v.a. absolutamente continua, se define la varianza de  $X$  como  $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ .

---

<sup>3</sup>Es decir, si  $S$  es a lo más numerable.

<sup>4</sup> $d(x, y) = |y - x|$  con  $x, y \in \mathbb{C}$ .

**Definición A.11.** Sea  $X$  una v.a. y  $S$  su imagen. Si la cardinalidad de  $S$  es finita, entonces decimos que  $X$  tiene una distribución uniforme discreta si  $\mathbb{P}(X = a) = \mathbb{P}(X = b)$  para toda  $a, b \in S$ .

**Definición A.12.** Se dice que una v.a. absolutamente continua  $X$  tiene una distribución uniforme en el intervalo  $[a, b]$  si  $f(x) = (b - a)^{-1}$  para toda  $x \in [a, b]$  y  $f(x) = 0$  para  $x \notin [a, b]$ .

**Definición A.13.** Se dice que una v.a. absolutamente continua  $X$  tiene una distribución gaussiana (o simplemente se dice que  $X$  es gaussiana) de media  $\mu$  y varianza  $\sigma^2$  si

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (\text{A.4})$$

para toda  $x \in \mathbb{R}$ .

**Definición A.14.** Se dice que una v.a. aleatoria es multivariada compleja si y sólo si su imagen es un subconjunto de  $\mathbb{C}^n$  para algún  $n$  y todos sus componentes son v.a.s complejas.

Cuando nos referimos a una distribución univariada estamos haciendo referencia a una distribución que sólo depende de un parámetro, mientras que multivariada implica la dependencia de la distribución en varios parámetros.

**Definición A.15.** La covarianza  $\text{cov}(X, Y)$  de dos v.a.s complejas  $X$  y  $Y$  está dada por

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^*]. \quad (\text{A.5})$$

**Definición A.16.** La correlación  $\rho(X, Y)$  de dos v.a.s complejas  $X$  y  $Y$  está dada por

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \quad (\text{A.6})$$

**Definición A.17.** Sean  $\mathbf{A} = \{a_{ij}\}$  y  $\mathbf{B} = \{b_{ij}\}$  matrices de  $m \times n$  y  $p \times q$  respectivamente.

Se define su producto de Kronecker como la matriz de  $mp \times nq$  dada por

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}. \quad (\text{A.7})$$

**Definición A.18.** Un proceso estocástico es un conjunto de v.a.s  $\{X_t : \Omega \rightarrow E_X : t \in T_X\}$  definidas sobre un espacio de probabilidad común. A los conjuntos  $E_X$  y  $T_X$  se les denomina espacio de estados del proceso estocástico y conjunto índice del proceso estocástico respectivamente.

**Definición A.19.** Una cadena de Markov es un proceso estocástico  $X_1, X_2, \dots$  tal que

$$\mathbb{P}(X_{n+1} = x | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x | X_n = x_n). \quad (\text{A.8})$$

Intuitivamente, una cadena de Markov ignora el pasado en el sentido que el futuro sólo depende del estado actual. Al proceso estocástico de la definición anterior también se le denomina cadena de Markov de primer orden, mientras que una cadena de Markov de orden  $k$  es un proceso estocástico tal que

$$\mathbb{P}(X_{n+1} = x | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x | X_{n-k+1} = x_{n-k+1}, \dots, X_n = x_n). \quad (\text{A.9})$$

La convergencia de una sucesión de números complejos no contiene ambigüedad en su definición, es decir, esencialmente sólo existe una manera de convergencia. Dado que las v.a.s son funciones que van de un espacio de probabilidad a un espacio medible<sup>5</sup>, la forma en que puede converger una sucesión de v.a.s a otra v.a. puede darse en muchos sentidos. Las siguientes son algunas de tales formas de convergencia.

**Definición A.20.** Sean  $X, X_1, X_2, \dots$  v.a.s definidas en un espacio de probabilidad común

---

<sup>5</sup>Se denomina espacio medible a la pareja  $(S, B(S))$  compuesta por un conjunto  $S$  y una  $\sigma$ -álgebra  $B(S)$  de subconjuntos de  $S$ .

$(\Omega, \mathbb{F}, \mathbb{P})$ . Se dice que  $X_1, X_2, \dots$  converge casi seguramente a la v.a.  $X$  si

$$\mathbb{P} \left( \left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = 1. \quad (\text{A.10})$$

Esto se denota por  $X_n \xrightarrow{\text{c.s.}} X$  o  $X_n \rightarrow X$  con probabilidad uno.

A la convergencia casi segura también se le conoce como convergencia casi en todas partes.

**Definición A.21.** Sean  $X, X_1, X_2, \dots$  v.a.s definidas en un espacio de probabilidad común  $(\Omega, \mathbb{F}, \mathbb{P})$ . Se dice que  $X_1, X_2, \dots$  converge en probabilidad a la v.a.  $X$  si para todo  $\epsilon > 0$  se tiene que

$$\lim_{n \rightarrow \infty} \mathbb{P} (|X_n - X| > \epsilon) = 0. \quad (\text{A.11})$$

Denotamos esta convergencia por  $X_n \xrightarrow{p} X$ .

**Definición A.22.** Sean  $X, X_1, X_2, \dots$  v.a.s,  $F_X$  la función de distribución de  $X$  y  $F_{X_n}$  la función de distribución de  $X_n$  ( $n \in \mathbb{N}$ ). Se dice que  $X_1, X_2, \dots$  converge en distribución a la v.a.  $X$  si

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (\text{A.12})$$

para toda  $x$  donde  $F_X$  es continua. Este modo de convergencia se denota por  $X_n \xrightarrow{d} X$ .

A la convergencia en distribución también se le conoce como convergencia en ley o convergencia débil. A continuación se enuncia la ley débil de los grandes números, la cual ejemplifica el uso de la convergencia en probabilidad.

**Teorema A.1.** Sean  $X_1, X_2, \dots$  v.a.s iid con distribución común  $X$ . Si  $\mathbb{E}[X] < \infty$  entonces

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} \mathbb{E}[X]. \quad (\text{A.13})$$

## A.2. Breve Introducción a la Teoría Ergódica

La teoría ergódica y en especial el teorema ergódico de Birkhoff, tienen una importancia fundamental en la motivación detrás del uso de la capacidad ergódica espectral como medida de desempeño de un sistema de comunicaciones. Esta sección tiene como objetivo mostrar

brevemente dicha relevancia. Algunas pruebas se omiten<sup>6</sup>, excepto aquellas que ilustran el uso de nuevos conceptos de teoría ergódica que resulten relevantes al estudiar el comportamiento de ciertos canales de comunicación, como por ejemplo los conceptos de transformación conservadora de la medida y transformación mezclante.

**Definición A.23.** *Sea  $(\Omega, \mathbb{F}, \mathbb{P})$  un espacio de probabilidad y  $T : \Omega \rightarrow \Omega$  una transformación medible con respecto a  $\mathbb{F}$ .*

- *Si  $\mathbb{P}(T^{-1}A) = \mathbb{P}(A)$  para toda  $A \in \mathbb{F}$ , se dice que  $T$  preserva la medida.*
- *Si  $T$  es una función biyectiva y para toda  $A \in \mathbb{F}$  se tiene que  $TA \in \mathbb{F}$  se dice que  $T$  es invertible.*

La definición de preservación de la medida se da en términos de  $T^{-1}A$  en lugar de  $TA$  debido a que el último conjunto no necesariamente pertenece a  $\mathbb{F}$  (dado que  $T$  es medible con respecto a  $\mathbb{F}$ ). Y debido a que  $TA$  no necesariamente está en  $\mathbb{F}$  para  $T$  biyectiva, se define la invertibilidad con la condición adicional (a la biyectividad) que  $T^{-1}$  sea medible. Esto le da sentido a la expresión  $\mathbb{P}(TA)$  cuando  $T$  es invertible.

**Proposición A.1.** *Sea  $T$  invertible. Entonces  $\mathbb{P}(A) = \mathbb{P}(TA)$  para toda  $A \in \mathbb{F}$  si y sólo si  $T$  preserva la medida.*

**Demostración.** Supóngase que  $\mathbb{P}(A) = \mathbb{P}(TA)$ . Para toda  $A \in \mathbb{F}$  defínase  $B := T^{-1}A \in \mathbb{F}$  (dado que  $T$  es una transformación medible). Dado que  $T$  es invertible entonces  $TT^{-1}A = A$  y utilizando la hipótesis  $\mathbb{P}(A) = \mathbb{P}(TA)$  tenemos

$$\mathbb{P}(B) = \mathbb{P}(TB), \tag{A.14}$$

$$\mathbb{P}(T^{-1}A) = \mathbb{P}(TT^{-1}A), \tag{A.15}$$

$$\mathbb{P}(T^{-1}A) = \mathbb{P}(A), \tag{A.16}$$

y por lo tanto  $T$  preserva la medida.

---

<sup>6</sup>Las pruebas se omiten debido a que involucran un cierto grado de tecnicismo de teoría de la medida, lo cual se escapa del propósito de esta tesis. Estas demostraciones pueden ser encontradas en cualquier libro estándar de teoría ergódica, como por ejemplo el libro de Billingsley [Bil78].

Supóngase que  $T$  preserva la medida. Para todo  $A \in \mathbb{F}$  defínase  $B := TA \in \mathbb{F}$  (dado que  $T$  es invertible). Dado que  $T$  preserva la medida y es invertible se tiene

$$\mathbb{P}(B) = \mathbb{P}(T^{-1}B), \quad (\text{A.17})$$

$$\mathbb{P}(TA) = \mathbb{P}(T^{-1}TA), \quad (\text{A.18})$$

$$\mathbb{P}(TA) = \mathbb{P}(A), \quad (\text{A.19})$$

y por lo tanto  $\mathbb{P}(A) = \mathbb{P}(TA)$  para toda  $A \in \mathbb{F}$ .  $\square$

**Corolario A.1.** *Sea  $T$  una transformación invertible que preserva la medida, entonces  $T^{-1}$  preserva la medida.*

**Demostración.** Por la Proposición A.1 se tiene que  $\mathbb{P}(A) = \mathbb{P}(TA)$  para todo conjunto  $A \in \mathbb{F}$ , y dado que  $T$  es biyectiva (puesto que es invertible) tenemos que  $T = (T^{-1})^{-1}$  y en consecuencia

$$\mathbb{P}(A) = \mathbb{P}((T^{-1})^{-1}A). \quad (\text{A.20})$$

Por lo tanto  $T^{-1}$  también preserva la medida.  $\square$

La imagen de cualquier transformación  $T$  que conserva la medida es esencialmente  $\Omega$ , ya que si  $A \in \mathbb{F}$  es tal que  $T\Omega \subset A$ , entonces  $\mathbb{P}(A) = \mathbb{P}(T^{-1}A) = \mathbb{P}(\Omega) = 1$ .

El siguiente teorema nos facilita demostrar que una transformación preserva la medida, por lo cual será de particular importancia en la siguiente sección.

**Teorema A.2.** *Sea  $\mathbb{F}_0$  un anillo generador de  $\mathbb{F}$ . Si  $T^{-1}A \in \mathbb{F}$  y  $\mathbb{P}(T^{-1}A) = \mathbb{P}(A)$  para toda  $A \in \mathbb{F}_0$  entonces  $T$  es una transformación que preserva la medida.*

Este teorema es una consecuencia directa del teorema de la clase monótona (véase el libro de Halmos [Hal50]), por lo cual su demostración se omite. Es importante notar que una transformación que preserva la medida está compuesta por la terna  $(T, \mathbb{F}, \mathbb{P})$  y no solamente por la transformación  $T$ .

**Definición A.24.** *Un conjunto  $A$  se dice que es estrictamente invariante (respecto a  $T$ ) si  $T^{-1}A = A$ .*

**Definición A.25.** *Se dice que un conjunto  $A$  es invariante si  $\mathbb{P}(A \Delta T^{-1}A) = 0$ .*

**Definición A.26.** Se dice que una transformación es ergódica si cualquier conjunto invariante  $A$  satisface que  $\mathbb{P}(A) = 0$ , o bien  $\mathbb{P}(A) = 1$ .

**Proposición A.2.** Sea  $T$  una transformación que preserva la medida, entonces existe un conjunto estrictamente invariante no trivial si y sólo si existe un conjunto invariante no trivial.

**Demostración.** Claramente si existe un conjunto estrictamente invariante  $A$  no trivial entonces este conjunto satisface la condición de invarianza, ya que  $\mathbb{P}(A \Delta T^{-1}A) = \mathbb{P}(\emptyset) = 0$ . Si existe un conjunto invariante  $A$ , definamos  $B := \limsup_n T^{-n}A$ . Observemos que

$$B = \limsup_{n \geq 0} T^{-n}A = \limsup_{n \geq 1} T^{-n}A = T^{-1} \limsup_{n \geq 0} T^{-n}A = T^{-1}B, \quad (\text{A.21})$$

por lo tanto  $B$  es estrictamente invariante. Dado que  $A$  y  $B$  tienen la misma medida entonces  $B$  es el conjunto buscado.  $\square$

La proposición anterior implica que la definición de transformación ergódica no depende de qué invarianza se tome (estricta o no estricta). Dado que

$$\mathbb{P}(A - T^{-1}A) = \mathbb{P}(A - A \cap T^{-1}A) \quad (\text{A.22})$$

$$= \mathbb{P}(A) - \mathbb{P}(A \cap T^{-1}A) \quad (\text{A.23})$$

$$= \mathbb{P}(T^{-1}A) - \mathbb{P}(A \cap T^{-1}A) \quad (\text{A.24})$$

$$= \mathbb{P}(T^{-1}A - A \cap T^{-1}A) \quad (\text{A.25})$$

$$= \mathbb{P}(T^{-1}A - A), \quad (\text{A.26})$$

tenemos que  $\mathbb{P}(A \Delta T^{-1}A) = 2\mathbb{P}(A - T^{-1}A) = 2\mathbb{P}(T^{-1}A - A)$ . Por lo tanto, para probar la invarianza de un cierto conjunto  $A$  es suficiente demostrar que  $A - T^{-1}A$  o  $T^{-1}A - A$  tienen probabilidad cero. Este resultado nos dice que en particular, una transformación ergódica no deja invariante, contrae o expande a ningún conjunto medible.

**Teorema A.3.** Si  $T$  es invertible, entonces  $T$  es ergódica si y sólo si  $T^{-1}$  es ergódica.

**Demostración.** Supongamos que  $T$  es ergódica. Sea  $A$  un conjunto invariante con respecto a  $T^{-1}$ , es decir  $A = TA$ . Dado que  $T$  preserva la medida y es biyectiva tenemos que  $T^{-1}A = A$ ,

y por lo tanto  $A$  también es invariante con respecto a  $T$ . Al ser  $T$  ergódica se tiene que  $\mathbb{P}(A) = 0$  o  $\mathbb{P}(A) = 1$  y por lo tanto  $T^{-1}$  también es ergódica. El regreso es idéntico.  $\square$

**Definición A.27.** Decimos que una transformación es mezclante si para todos  $A, B \in \mathbb{F}$  se cumple que

$$\lim_{n \rightarrow \infty} \mathbb{P}(A \cap T^{-n}B) = \mathbb{P}(A)\mathbb{P}(B). \quad (\text{A.27})$$

En términos generales, una transformación es mezclante si  $\lim_{n \rightarrow \infty} \mathbb{P}(T^{-n}B|A) = \mathbb{P}(B)$ , lo cual intuitivamente implica que la *densidad* o abundancia<sup>7</sup> de  $T^{-n}B$  en  $A$  es la misma que la *densidad* de  $B$  en  $\Omega$  para  $n$  suficientemente grande.

**Teorema A.4.** Si  $T$  es mezclante entonces  $T$  es ergódica.

**Demostración.** Sea  $B$  un conjunto invariante, dejando  $A = B$  se tiene que

$$\lim_{n \rightarrow \infty} \mathbb{P}(B \cap T^{-n}B) = \lim_{n \rightarrow \infty} \mathbb{P}(B \cap B) = \mathbb{P}(B). \quad (\text{A.28})$$

Dado que  $T$  es mezclante entonces  $\mathbb{P}(B) = \mathbb{P}(B)^2$ , y por lo tanto  $\mathbb{P}(B) = 0$  o bien  $\mathbb{P}(B) = 1$ .  $\square$

Esta propiedad será particularmente importante en la siguiente sección, ya que bastará mostrar que una transformación es mezclante para obtener la ergodicidad como consecuencia.

El siguiente teorema afirma que es suficiente tener la propiedad mezclante en un anillo generador de  $\mathbb{F}$  para asegurar que la propiedad mezclante se mantiene en general.

**Teorema A.5.** Sea  $\mathbb{F}_0$  un anillo generador de  $\mathbb{F}$ . Si

$$\lim_{n \rightarrow \infty} \mathbb{P}(A \cap T^{-n}B) = \mathbb{P}(A)\mathbb{P}(B) \quad (\text{A.29})$$

para toda  $A, B \in \mathbb{F}_0$  entonces  $T$  es mezclante.

**Definición A.28.** Se dice que una función  $g(\omega)$  es invariante si  $g(T\omega) = g(\omega)$  casi en todas partes.

---

<sup>7</sup>Nos referimos a *densidad* en analogía al caso hipotético en el que  $\Omega$  representa un líquido compuesto por varias sustancias y los eventos representan diferentes regiones del líquido. En esta analogía hablar de la *densidad* de  $B$  en  $A$  implica hablar de la probabilidad condicional  $\mathbb{P}(B|A)$ .

Claramente un conjunto es invariante si y sólo si su función indicadora es invariante. El siguiente resultado, conocido como teorema ergódico de Birkhoff, nos permitirá calcular la cantidad de información que pasa a través de ciertos canales de comunicación. La demostración puede verse en el libro de Billingsley “Ergodic Theory and Information” [Bil78].

**Teorema A.6.** *Si  $f$  es integrable, entonces existe una función invariante integrable  $\hat{f}$  tal que  $\mathbb{E}[f] = \mathbb{E}[\hat{f}]$  y*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \hat{f}(\omega) \quad (\text{A.30})$$

*casi en todas partes. Si  $T$  es ergódica entonces  $\hat{f} = \mathbb{E}[f]$  casi en todas partes.*

### A.3. Canales Ergódicos

En esta sección se motiva el uso de transformaciones ergódicas en el modelado de canales de comunicación, así como las implicaciones que esto conlleva. Dado que estamos interesados especialmente en canales inalámbricos, la motivación se hace a través de dichos canales, sin embargo argumentos similares aplican a las contrapartes alámbricas. La ergodicidad se derivará después de motivar la propiedad de mezcla en un canal inalámbrico<sup>8</sup>.

Sea  $(\Omega, \mathbb{F}, \mathbb{P})$  un espacio de probabilidad con

$$\Omega = \{\omega = (\dots, \rho_{-1}, \rho_0, \rho_1, \dots) : \rho_i \in \rho\}, \quad (\text{A.31})$$

donde  $\rho$  es un conjunto equipado con una  $\sigma$ -álgebra  $\beta(\rho)$ . Sea  $\mathbb{F} = \sigma(\mathbb{F}_0)$  con  $\mathbb{F}_0$  conformado por todos los conjuntos de la forma  $\{\omega \in \Omega : \omega_n \in A_n, \dots, \omega_{n+k-1} \in A_{n+k-1}, A_i \in \beta(\rho)\}$  con  $n \in \mathbb{Z}$  y  $k \in \mathbb{N}$ . Además definamos  $T : \Omega \rightarrow \Omega$  por

$$T\omega = (\dots, x_0(\omega), x_1(\omega), x_2(\omega), \dots), \quad (\text{A.32})$$

donde  $x_n(\omega) = \rho_n$ , es decir,  $x_n$  devuelve la  $n$ -ésima coordenada de  $\omega$ . Claramente  $T$  es invertible, y  $T^{-1}$  está dada por

$$T^{-1}\omega = (\dots, x_{-2}(\omega), x_{-1}(\omega), x_0(\omega), \dots). \quad (\text{A.33})$$

---

<sup>8</sup>Decimos que un canal tiene la propiedad de mezcla si el canal como proceso estocástico es mezclante.

Con lo anterior, modelaremos un canal de comunicación inalámbrica como un  $\omega \in \Omega$ , del cual sólo podemos observar  $x_0(\omega)$  en un momento dado. El cambio del canal se ve representado por la evolución temporal inducida por  $T$ , es decir, si en este momento se observa  $x_0(\omega)$  para el siguiente instante de tiempo lo que observaremos será  $x_0(T\omega) = x_1(\omega)$ . Esta motivación ciertamente equivale a decir que el canal tiene un destino o trayectoria predeterminada, y que el tiempo es la transformación que lleva al canal de un estado al siguiente. En principio esto puede parecer irreal, pero más allá de esta implicación lo importante es que el modelo se ajuste y sea útil.

Que  $T$  preserve la medida es equivalente a decir que

$$\mathbb{P}(\omega \in \Omega : \omega_n \in A_n, \dots, \omega_{n+k-1} \in A_{n+k-1}, A_i \in \beta(\rho)) \quad (\text{A.34})$$

no depende de  $n$ . Esto es, que el canal sea estacionario. Si bien el canal es aleatorio, idealmente éste mantiene una cierta regularidad a lo largo del tiempo<sup>9</sup>. Por lo tanto, es en general razonable suponer que el canal es estacionario, y en consecuencia  $T$  preserva la medida. Por el Corolario A.1 tenemos que  $T^{-1}$  también preserva la medida.

Observemos que si  $A = \{\omega : \omega_n \in A_n, \dots, \omega_{n+k-1} \in A_{n+k-1}, A_i \in \beta(\rho)\}$  y  $B = \{\omega : \omega_m \in B_m, \dots, \omega_{m+l-1} \in B_{m+l-1}, B_j \in \beta(\rho)\}$  entonces

$$\begin{aligned} \mathbb{P}(A \cap T^s B) &= \mathbb{P}(\{\omega : \omega_n \in A_n, \dots, \omega_{n+k-1} \in A_{n+k-1}, \\ &\quad \omega_{m+s} \in B_{m+s}, \dots, \omega_{m+l+s-1} \in B_{m+l+s-1}, A_i \in \beta(\rho), B_j \in \beta(\rho)\}). \end{aligned} \quad (\text{A.35})$$

Es importante resaltar que cuando  $s \rightarrow \infty$  el lado derecho de la ecuación anterior requiere calcular la probabilidad de  $A$  y  $T^s B$ . Para  $s$  suficientemente grande uno esperaría que estos eventos sean prácticamente independientes, ya que lo contrario sería afirmar que dado un valor del canal en un instante específico, ese valor aportará información del canal para cualquier

---

<sup>9</sup>Sin embargo, esto puede ser no necesariamente cierto, ya que el ambiente en general puede tener evoluciones temporales que afecten esta idealización.

tiempo<sup>10</sup>. Por lo tanto, puede suponerse que

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{P}(A \cap T^s B) &= \lim_{n \rightarrow \infty} \mathbb{P}(\{\omega : \omega_n \in A_n, \dots, \omega_{n+k-1} \in A_{n+k-1}, \\
&\quad \omega_{m+s} \in B_{m+s}, \dots, \omega_{m+l+s-1} \in B_{m+l+s-1}, \\
&\quad A_i \in \beta(\rho), B_j \in \beta(\rho)\}) \\
&= \mathbb{P}(A)\mathbb{P}(T^s B) = \mathbb{P}(A)\mathbb{P}(B),
\end{aligned} \tag{A.36}$$

donde la última igualdad se sigue del hecho que tanto  $T$  como  $T^{-1}$  preservan la medida. Debe observarse que la última ecuación implica que  $T^{-1}$  es mezclante, y por el Teorema A.4 se tiene que  $T^{-1}$  es ergódica. Aplicando el Teorema A.3 se tiene que  $T$  también es ergódica y por lo tanto se cumple el teorema ergódico de Birkhoff (Teorema A.6).

Por ejemplo, si tomamos  $\rho = M_{n_R \times n_T}(\mathbb{C})$  el conjunto de las matrices complejas de  $n_R \times n_T$ , entonces los valores  $\rho_i$  no son otros que los posibles estados de un canal MIMO con  $n_R$  antenas receptoras y  $n_T$  antenas transmisoras. Dado que la eficiencia espectral para un valor del canal dado<sup>11</sup> es igual a

$$f(\omega) = \log \det(I + x_0(\omega)\Phi x_0(\omega)^*), \tag{A.37}$$

y dado que  $T^{-1}$  es ergódica, por el teorema ergódico tenemos que

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k}\omega) = \mathbb{E}[f] \right) = 1, \tag{A.38}$$

donde  $\mathbb{E}[f] = \int f(w) \mathbb{P}(dw) = \int f(x_0(\omega)) \mathbb{P}(dx_0(\omega))$ . Observemos que la eficiencia espectral ergódica sólo depende de la función de distribución de  $x_0(\omega)$ , la cual es la distribución común a todas las  $x_i(\omega)$  que componen el proceso estocástico (estacionario) bajo estudio.

En general, la ergodicidad conlleva a que la eficiencia espectral ergódica sólo dependa de la distribución estacionaria común a las v.a.s que componen el proceso estocástico. Esto nos permite calcular la eficiencia espectral ergódica utilizando únicamente dicha distribución común<sup>12</sup>.

---

<sup>10</sup>Obsérvese que para tiempos pequeños el estado actual del canal nos puede ayudar a predecir estados siguientes, pero aquí se trabaja con el límite asintótico.

<sup>11</sup>Manteniendo fija la distribución de la entrada, es decir, dejando a la matriz  $\Phi$  constante.

<sup>12</sup>Por ejemplo, la fórmula  $\mathbb{E}_{\mathbf{H}}[\log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{H}^*)]$  sólo depende de la distribución de  $\mathbf{H}$ , la cual se puede interpretar en este caso como la distribución de  $x_0(\omega)$ .



# Apéndice B

## Codificación

En este apéndice se presenta un análisis de las tareas básicas de la codificación: la compresión de datos y la transmisión segura de la información a través de un canal ruidoso. El material que se presenta en este apéndice son resultados elementales de la teoría de la información, tanto por su importancia como por su valor histórico dentro de la teoría. Estos resultados son necesarios para entender los fundamentos y límites de la compresión y transmisión de datos. Además, nos muestran la importancia del análisis de la capacidad ergódica.

### B.1. Introducción

Supóngase que se tiene una fuente a tiempo discreto, i.e. su conjunto índice como proceso estocástico es  $T = \mathbb{N}$ , la cual tiene un espacio de estados compuesto por los símbolos  $\{A_1, A_2, \dots, A_l\}$ . Para transmitir la información que produce esta fuente a través de un canal binario<sup>1</sup> sin ruido<sup>2</sup>, es necesario pasar los símbolos  $A_i$  a binario. Informalmente, a este proceso de cambiar los datos de una fuente por otros lo denominaremos codificación.

Es claro que si tomamos  $k \geq \log_2 l$  entonces a cada palabra del conjunto  $\{A_1, A_2, \dots, A_l\}$  le podremos asignar una única palabra binaria de longitud  $k$  que la represente. Así por ejemplo, si nuestra fuente toma tres posibles símbolos  $\{A_1, A_2, A_3\}$ , entonces se puede tomar  $k = 2$  y alguna asignación, por ejemplo  $A_1 \mapsto 01$ ,  $A_2 \mapsto 10$  y  $A_3 \mapsto 11$  lo cual resuelve hasta cierto

---

<sup>1</sup>Un canal donde sólo se pueden transmitir unos o ceros.

<sup>2</sup>Un canal sin ruido es un canal donde se recibe lo que se transmite. Por el contrario, en un canal con ruido existe la posibilidad de recibir algo diferente a lo que se transmitió.

punto el problema. Supóngase que los símbolos que produce la fuente son v.a.s iid tales que  $\mathbb{P}(A_1) = 1/2$ ,  $\mathbb{P}(A_2) = 1/4$  y  $\mathbb{P}(A_3) = 1/4$ . Obsérvese que bajo el esquema de codificación anterior, en promedio se necesitan dos dígitos binarios para transmitir cada símbolo (de hecho cada símbolo utiliza dos dígitos).

Si en lugar de la asignación propuesta en el párrafo anterior utilizamos la asignación  $A_1 \mapsto 1$ ,  $A_2 \mapsto 00$  y  $A_3 \mapsto 01$ , entonces la longitud promedio por símbolo será igual a

$$\mathbb{E}[L] = \mathbb{P}(A_1)L(A_1) + \mathbb{P}(A_2)L(A_2) + \mathbb{P}(A_3)L(A_3) = \frac{1}{2}1 + \frac{1}{4}2 + \frac{1}{4}2 = \frac{3}{2}, \quad (\text{B.1})$$

donde  $L(A_i)$  representa la longitud de la palabra asignada al símbolo  $A_i$ . Dos cosas deben observarse, en primer lugar el código es decodificable, es decir, para cada cadena binaria recibida existe una y sólo una cadena de símbolos de la fuente que producen dicha cadena binaria. Además, la longitud promedio se redujo en un 25 % con respecto al primer método de codificación, por lo que concluimos que no todos los métodos de codificación son igualmente buenos.

Por lo tanto, la codificación nos permite transmitir (en el caso ideal de un canal sin ruido) de manera más eficiente la información, esto debido a la compresión de datos que da a lugar durante dicho proceso. Sin embargo, surgen dos preguntas: ¿Hasta qué punto será posible comprimir la información de una fuente? ¿Qué pasa en el caso general de un canal ruidoso? La primer pregunta se responderá en las primeras dos secciones de este apéndice a través de la llamada propiedad de equipartición asintótica AEP (del inglés Asymptotic Equipartition Property). Para responder la segunda pregunta, enunciaremos y comentaremos el teorema de codificación de Shannon para canales discretos sin memoria. Este teorema, además de responder la pregunta planteada, nos mostrará la importancia de conocer la capacidad ergódica de un canal.

## B.2. Propiedad de Equipartición Asintótica

El objetivo de esta sección es estudiar los límites teóricos de la compresión de datos. Para ello, primeramente estudiaremos la propiedad de equipartición asintótica, la cual nos permitirá construir un código asintóticamente óptimo.

En esta sección supondremos que la fuente está compuesta de v.a.s iid, es decir el proceso estocástico  $\{X_n\}_{n \geq 1}$  que modela la fuente está compuesto de v.a.s independientes todas con una misma distribución común  $X$ . Además supondremos que  $X$  es una variable aleatoria discreta que toma una cantidad finita de símbolos y que el medio por el cual se transmite la información es un canal discreto<sup>3</sup> sin memoria<sup>4</sup> y sin ruido. En la siguiente sección analizaremos el caso más general en el que  $\{X_n\}_{n \geq 1}$  es un proceso estocástico estacionario y ergódico.

La siguiente propiedad, denominada propiedad de equipartición asintótica, afirma que conforme  $n \rightarrow \infty$  la probabilidad de cualquier cadena *típica* de símbolos  $x_1, \dots, x_n$  converge a  $2^{-nH(X)}$ . Esto implica que todas las cadenas típicas tienen asintóticamente la misma probabilidad.

**Teorema B.1.** *Sean  $X_1, X_2, \dots$  v.a.s iid con distribución común  $X$ . Si  $H(X) < \infty$  entonces*

$$-\frac{1}{n} \log_2 p(X_1, \dots, X_n) \xrightarrow{p} H(X). \quad (\text{B.2})$$

**Demostración.** Dado que las  $X_i$  son v.a.s independientes entonces  $-\log_2 p(X_i)$  son v.a.s iid. Por las identidades de logaritmos tenemos que

$$-\frac{1}{n} \log_2 p(X_1, \dots, X_n) = \frac{\sum_{i=1}^n -\log_2 p(X_i)}{n}. \quad (\text{B.3})$$

Aplicando la ley débil de los grandes números (Teorema A.1), la última ecuación implica que

$$-\frac{1}{n} \log_2 p(X_1, \dots, X_n) \xrightarrow{p} \mathbb{E}[-\log_2 p(X)] = H(X) \quad (\text{B.4})$$

tal y como se quería demostrar. □

La noción de cadena típica se formaliza de la siguiente manera.

**Definición B.1.** *Sean  $\epsilon > 0$  y  $n \in \mathbb{N}$  dados, se define el conjunto típico  $A_\epsilon^{(n)}$  con respecto a  $X$  como el conjunto de cadenas  $(x_1, \dots, x_n) \in X^n$  tales que*

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}. \quad (\text{B.5})$$

---

<sup>3</sup>Se define un canal discreto como un canal en el cual los conjuntos de estados de la entrada y salida son finitos.

<sup>4</sup>Un canal es un canal sin memoria si la distribución estadística de la salida sólo depende de la entrada y no de estados anteriores del canal.

En la definición anterior  $X^n$  denota al conjunto  $\{(x_1, \dots, x_n) : x_1, \dots, x_n \in E_X\}$ , donde  $E_X$  es el conjunto de valores que toma  $X$ . Las propiedades más importantes del conjunto típico están resumidas en el siguiente teorema.

**Teorema B.2.** *Sea  $\epsilon > 0$ , entonces el conjunto típico cumple las siguientes propiedades.*

- a)  $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$  si y sólo si  $H(X) - \epsilon \leq -\frac{1}{n} \log_2 p(x_1, \dots, x_n) \leq H(X) + \epsilon$ .
- b)  $\mathbb{P}\left(A_\epsilon^{(n)}\right) > 1 - \epsilon$  para  $n$  suficientemente grande.
- c)  $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$ .
- d)  $|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{n(H(X)-\epsilon)}$  para  $n$  suficientemente grande.

**Demostración.**

- a) Por definición  $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$  si y sólo si

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}. \quad (\text{B.6})$$

Dado que el  $\log_2 x \leq \log_2 y$  si y sólo  $x \leq y$  la ecuación anterior se puede reescribir como

$$-n(H(X) + \epsilon) \leq \log_2 p(x_1, \dots, x_n) \leq -n(H(X) - \epsilon), \quad (\text{B.7})$$

dividiendo por  $-n$  se llega a la desigualdad deseada.

- b) Por la AEP se tiene que para todo  $\epsilon > 0$  y para toda  $\delta > 0$  existe  $N$  tal que para toda  $n > N$  se cumple que

$$\mathbb{P}\left[\left|H(X) - \frac{1}{n} \log_2 p(X_1, \dots, X_n)\right| > \epsilon\right] < \delta. \quad (\text{B.8})$$

Por lo tanto, para toda  $n > N$

$$\mathbb{P}\left[\left|H(X) - \frac{1}{n} \log_2 p(X_1, \dots, X_n)\right| < \epsilon\right] > 1 - \delta. \quad (\text{B.9})$$

Si dejamos que  $\delta = \epsilon$  concluimos que

$$1 - \epsilon < \mathbb{P}\left[H(X) - \epsilon \leq -\frac{1}{n} \log_2 p(x_1, \dots, x_n) \leq H(X) + \epsilon\right] = \mathbb{P}\left[A_\epsilon^{(n)}\right]. \quad (\text{B.10})$$

c) Dado que  $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$  implica  $p(x_1, \dots, x_n) \geq 2^{-n(H(X)+\epsilon)}$  tenemos que

$$1 \geq \mathbb{P} \left[ A_\epsilon^{(n)} \right] \tag{B.11}$$

$$= \sum_{(x_1, \dots, x_n) \in A_\epsilon^{(n)}} p(x_1, \dots, x_n) \tag{B.12}$$

$$\geq \sum_{(x_1, \dots, x_n) \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \tag{B.13}$$

$$= 2^{-n(H(X)+\epsilon)} |A_\epsilon^{(n)}|, \tag{B.14}$$

de donde el resultado se sigue.

d) Dado que  $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$  implica  $p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$  y dado que para  $n$  suficientemente grande  $\mathbb{P} \left[ A_\epsilon^{(n)} \right] > 1 - \epsilon$  tenemos que

$$1 - \epsilon < \mathbb{P} \left[ A_\epsilon^{(n)} \right] \tag{B.15}$$

$$= \sum_{(x_1, \dots, x_n) \in A_\epsilon^{(n)}} p(x_1, \dots, x_n) \tag{B.16}$$

$$\leq \sum_{(x_1, \dots, x_n) \in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} \tag{B.17}$$

$$= 2^{-n(H(X)-\epsilon)} |A_\epsilon^{(n)}|, \tag{B.18}$$

tal y como se quería demostrar. □

El conjunto típico debe su nombre a la propiedad (b), la cual asegura que el conjunto típico tiene una probabilidad tan alta como se desee tomando un  $n$  suficientemente grande. Esto quiere decir que para  $n$  suficientemente grande, las cadenas  $(x_1, \dots, x_n)$  que tienen una entropía muestral cercana a la entropía del proceso son prácticamente todas las cadenas probables. Antes de demostrar el teorema central de esta sección tenemos que definir con precisión qué es un código.

**Definición B.2.** *Se define un código  $C$  como una función que mapea las cadenas  $(x_1, \dots, x_n)$  a cadenas binarias  $C(x_1, \dots, x_n)$  las cuales son llamadas palabras (o palabras del código). Se dice que un código es invertible si la función es uno a uno. Denotaremos por  $L$  a la función que a cada cadena  $(x_1, \dots, x_n)$  le asigna la longitud de  $C(x_1, \dots, x_n)$ , i.e.  $L(x_1, \dots, x_n)$  es*

igual longitud de la palabra  $C(x_1, \dots, x_n)$ .

**Definición B.3.** Se dice que un código  $C$  es un código de bloque si la información codificada puede ser dividida en bloques de  $n$  símbolos los cuales pueden ser decodificados independientemente. Si el rango de  $C$  tiene cardinalidad  $M$ , entonces se dice que  $C$  es un código  $(M, n)$ .

Los códigos de bloque tienen una estructura algebraica muy fuerte [Lin91], además, un error en un bloque no afecta bloques subsecuentes, lo cual los hace bastante prácticos en términos de implementación. El siguiente teorema establece que existen buenos códigos en términos de compresión de la fuente.

**Teorema B.3.** Sean  $X_1, X_2, \dots$  v.a.s iid con distribución común  $X$ . Sea  $\epsilon > 0$ , entonces existe un entero  $n$  y un código  $C$  que mapea las cadenas  $x_1, \dots, x_n$  en cadenas binarias tales que el mapeo es invertible y

$$\mathbb{E} \left[ \frac{1}{n} L(X_1, \dots, X_n) \right] \leq H(X) + \epsilon. \quad (\text{B.19})$$

**Demostración.** Para demostrar este teorema construiremos un código que cumpla las propiedades requeridas. Dado que  $|A_{\epsilon'}^{(n)}| \leq 2^{n(H(X)+\epsilon')}$  para toda  $n$  y para todo  $\epsilon'$ , a cada cadena en el conjunto típico se le puede asociar un único índice con exactamente  $\lceil n(H(X) + \epsilon') \rceil \leq n(H(X) + \epsilon') + 1$  bits. El complemento del conjunto típico tiene una cardinalidad menor a  $|X|^n$  y por lo tanto se puede asociar un único índice a cada elemento en el complemento del conjunto típico, el cual contenga exactamente  $\lceil n \log_2 |X| \rceil \leq n \log_2 |X| + 1$  bits. El código  $C$  en cuestión es tal que a cada cadena  $(x_1, \dots, x_n) \in A_{\epsilon'}^{(n)}$  se le asigna como palabra codificada el índice asociado a la cadena dentro del conjunto típico agregando previamente un uno. A cada cadena fuera de  $A_{\epsilon'}^{(n)}$  se le asignará el índice correspondiente anteponiendo un cero.

Bajo este esquema de codificación la longitud esperada por símbolo cumple que

$$\mathbb{E} \left[ \frac{1}{n} L(C(X_1, \dots, X_n)) \right] = \frac{1}{n} \sum_{(x_1, \dots, x_n) \in X^n} p(x_1, \dots, x_n) L(x_1, \dots, x_n) \quad (\text{B.20})$$

$$= \frac{1}{n} \sum_{(x_1, \dots, x_n) \in A_{\epsilon'}^{(n)}} p(x_1, \dots, x_n) L(x_1, \dots, x_n) \quad (\text{B.21})$$

$$+ \frac{1}{n} \sum_{(x_1, \dots, x_n) \notin A_{\epsilon'}^{(n)}} p(x_1, \dots, x_n) L(x_1, \dots, x_n) \quad (\text{B.22})$$

$$\leq \frac{1}{n} \sum_{(x_1, \dots, x_n) \in A_{\epsilon'}^{(n)}} p(x_1, \dots, x_n) (n(H(X) + \epsilon') + 2) \quad (\text{B.23})$$

$$+ \frac{1}{n} \sum_{(x_1, \dots, x_n) \notin A_{\epsilon'}^{(n)}} p(x_1, \dots, x_n) (n \log_2 |X| + 2) \quad (\text{B.24})$$

$$= \frac{1}{n} (n(H(X) + \epsilon') + 2) \mathbb{P} \left[ A_{\epsilon'}^{(n)} \right] \quad (\text{B.25})$$

$$+ \frac{1}{n} (n \log_2 |X| + 2) \left( 1 - \mathbb{P} \left[ A_{\epsilon'}^{(n)} \right] \right) \quad (\text{B.26})$$

$$\leq \frac{1}{n} (n(H(X) + \epsilon') + 2) + \frac{1}{n} \epsilon' (n \log_2 |X| + 2) \quad (\text{B.27})$$

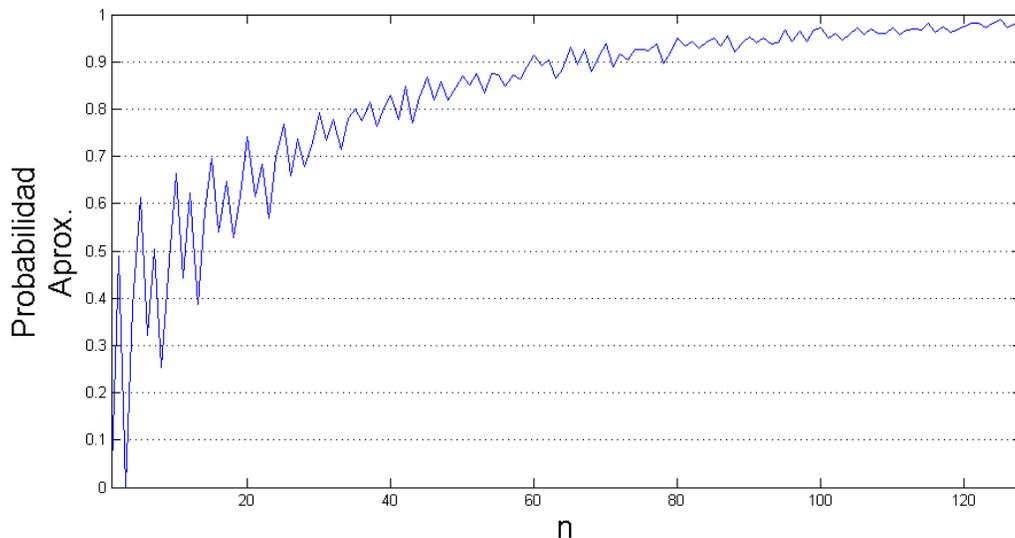
$$= H(X) + \epsilon' \log_2 |X| + \epsilon' \left( 1 + \frac{2}{n} \right) + \frac{2}{n}. \quad (\text{B.28})$$

Tomando  $\epsilon'$  suficientemente pequeño y  $n$  suficientemente grande se obtiene el resultado deseado.  $\square$

Podemos modificar el esquema de codificación de la demostración anterior de tal manera que las cadenas fuera del conjunto típico se codifiquen a una palabra de  $nH(X)$  bits (todas a la misma palabra), y las cadenas dentro del conjunto típico se codifiquen de la misma forma que en la demostración del teorema anterior. Si bien esta manera de codificar no es invertible puesto que manda todas las cadenas fuera del conjunto típico a la misma palabra, asintóticamente la probabilidad de estar fuera del conjunto típico es cero y por lo tanto el código es asintóticamente invertible casi en todas partes.

Por ejemplo, la fuente presentada al inicio de este apéndice tiene una entropía igual a  $H(X) = \frac{3}{2}$ , la cual coincide con la longitud esperada del código  $A_1 \mapsto 1$ ,  $A_2 \mapsto 00$  y  $A_3 \mapsto 01$ . Sin embargo, este código no es un código de bloque (ya que distintas cadenas tienen palabras de distintas longitudes). Con el fin de ilustrar la eficacia del código descrito en el párrafo anterior,

la siguiente gráfica muestra una aproximación<sup>5</sup> de la probabilidad del conjunto típico  $A_{0.1}^{(n)}$  (con respecto a la v.a. de la introducción de este apéndice) para valores de  $n$  en el intervalo  $[1, 128]$ .



Como se puede apreciar en la gráfica, para valores pequeños de  $n$  el conjunto típico  $A_{0.1}^{(n)}$  tiene probabilidades relativamente altas (mayores a 0.9). De hecho, de 100,000 simulaciones realizadas solamente 137 cayeron fuera del conjunto típico  $A_{0.1}^{256}$ , lo cual representa menos del 0.2% del total.

Es importante resaltar que el Teorema B.3 no afirma que no existan códigos cuya longitud promedio por símbolo sea menor que  $H(X)$ . Sin embargo, en la siguiente sección mostraremos que en efecto  $H(X)$  es una cota inferior para la mínima longitud promedio.

### B.3. AEP en Procesos Estocásticos

En la sección anterior relacionamos la longitud a la que se puede codificar una fuente, la cual es un proceso estocástico compuesto de v.a.s iid y la entropía de la v.a. que compone el proceso estocástico. En esta sección presentamos una generalización de dicho teorema y sus consecuencias.

Dado que calcular la entropía de un proceso estocástico directamente de la Definición 1.3

---

<sup>5</sup>Las probabilidades se estimaron simulando mil cadenas aleatorias, para después comparar la fracción de las cadenas que pertenecieron al conjunto típico con el total de ellas.

puede ser relativamente complicado, a continuación se enuncia<sup>6</sup> un teorema que nos permite calcular la entropía de un proceso estocástico de una manera alternativa.

**Teorema B.4.** *Si  $\{X_n\}_{n \geq 1}$  es un proceso estocástico estacionario, entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_1). \quad (\text{B.29})$$

Este teorema es particularmente útil en el caso de cadenas de Markov. Por ejemplo, supóngase que se tiene una cadena de Markov de primer orden, la cual toma valores en  $\{0, 1\}$  y cuya matriz de transición es

$$T = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix}. \quad (\text{B.30})$$

Realizando los cálculos correspondientes se obtiene que la distribución estacionaria del proceso es  $\mu_0 = 2/3$ ,  $\mu_1 = 1/3$ .

Se podría codificar la fuente pensando que las variables aleatorias  $X_n$  son v.a.s iid Bernoulli con probabilidad de éxito  $1/3$ , es decir, pensando que las v.a.s se distribuyen como la distribución estacionaria. En esta situación la entropía *aparente* del proceso es igual a

$$H(X') = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{1}{3} \log_2 \frac{1}{3} \approx 0.9183 \quad (\text{B.31})$$

lo cual nos haría pensar que la mínima longitud promedio por símbolo a la que se puede codificar la fuente es 0.9183.

Sin embargo, la probabilidad de ver un 00 en la cadena es prácticamente  $1/2$  para  $n$  suficientemente grande, mientras que para 01, 10 y 11 la probabilidad es  $1/6$ . Por lo tanto, si pensamos en la fuente como un proceso estocástico, el cual está compuesto por v.a.s iid con distribución común  $X''$ , la cual toma los valores 00, 01, 10 y 11 con probabilidades  $1/2$ ,  $1/6$ ,

---

<sup>6</sup>A partir de esta sección se omiten las demostraciones de los teoremas dado que en general son más técnicas, y desde el punto de vista de la intuición, únicamente generalizan los resultados de la sección anterior. Las demostraciones pueden encontrarse en [Cov06], la demostración del teorema de Shannon-McMillan-Breiman también puede encontrarse en [Bil78].

1/6 y 1/6 respectivamente, entonces la entropía del proceso aparentaría ser igual a

$$H(X'') = -\frac{1}{2} \log_2 \frac{1}{2} - 3 \frac{1}{6} \log_2 \frac{1}{6} \approx 1.7925. \quad (\text{B.32})$$

Esto nos lleva a pensar que la mínima longitud promedio a la cual se puede codificar la fuente es 0.8962 (por cada símbolo de  $X''$  se tiene dos símbolos del proceso original).

Además, la entropía del proceso estocástico original es

$$H(X) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1) \quad (\text{B.33})$$

$$= \lim_{n \rightarrow \infty} H(X_n | X_{n-1}) \quad (\text{B.34})$$

$$= \sum_{i,j} -\mu_i p_{ij} \log_2 p_{ij} \quad (\text{B.35})$$

$$\approx 0.8742 \quad (\text{B.36})$$

lo cual nos hace pensar que la mínima longitud promedio por símbolo a la que se puede codificar la fuente es 0.8742. Las discrepancias en los diferentes esquemas de codificación claramente se deben a la sobresimplificación que se hace en la estructura del proceso estocástico, lo cual deriva en un menor uso de las propiedades del mismo. Por esta razón es relevante estudiar la entropía de un proceso estocástico directamente, y de hecho, la entropía del proceso estocástico es la mínima longitud a la cual se puede codificar el proceso tal y como lo enuncia el siguiente teorema<sup>7</sup>.

**Teorema B.5.** *Sea  $n \in \mathbb{N}$ , entonces la mínima longitud promedio por símbolo  $L_n^*$  cumple la siguiente relación*

$$\frac{H(X_1, \dots, X_n)}{n} \leq L_n^* \leq \frac{H(X_1, \dots, X_n)}{n} + \frac{1}{n}. \quad (\text{B.37})$$

*Además, si  $X_1, X_2, \dots$  es un proceso estocástico estacionario entonces*

$$\lim_{n \rightarrow \infty} L_n^* = H(X), \quad (\text{B.38})$$

*donde  $H(X)$  es la entropía del proceso estocástico.*

---

<sup>7</sup>En este teorema se supone que el código es un código libre de prefijos. Un código es libre de prefijos si ninguna palabra del código es un prefijo de ninguna otra palabra del código. La decodificación de este tipo de códigos es relativamente simple, lo cual los hace ampliamente utilizados (véase [Lin91]).

El siguiente resultado, conocido como Teorema de Shannon-McMillan-Breiman, afirma que la AEP también se cumple para procesos estocásticos estacionarios y ergódicos, lo cual tiene implicaciones importantes en términos de codificación.

**Teorema B.6.** *Si  $H(X)$  es la entropía de un proceso estocástico  $\{X_n\}_{n \geq 1}$  estacionario y ergódico con espacio de estados finito, entonces*

$$-\frac{1}{n} \log_2 p(X_1, \dots, X_n) \xrightarrow{c.s.} H(X). \quad (\text{B.39})$$

Este último teorema nos permite definir un conjunto típico como lo habíamos hecho anteriormente y codificar de manera totalmente similar al caso en el que el proceso estaba compuesto de v.a.s iid. Por lo tanto, también es posible codificar una fuente estacionaria y ergódica a una velocidad igual a la entropía del proceso.

## B.4. Codificación en Canales Ruidosos

En las secciones anteriores analizamos la codificación en canales sin ruido, donde la compresión es el aspecto principal a analizar. Sin embargo, en la mayoría de sistemas prácticos hay ruido presente, el cual cambia algunos símbolos transmitidos por otros en una forma aleatoria.

**Definición B.4.** *Decimos que un canal binario sin memoria es un canal simétrico binario BSC (del inglés Binary Symmetric Channel) si la probabilidad de recibir un uno, dado que se mandó un cero, es la misma que la probabilidad de recibir un cero, dado que se transmitió un uno. Si  $p$  denota tal probabilidad decimos que el canal es un canal BSC( $p$ ).*

A partir de la definición de capacidad ergódica (Definición 1.6) podemos calcular la capacidad ergódica de un BSC( $p$ ). Realizando los cálculos correspondientes llegamos a que

$$C = \max_{p(X)} I(X; Y) \quad (\text{B.40})$$

$$= 1 + p \log_2 p + (1 - p) \log_2 (1 - p). \quad (\text{B.41})$$

En esta situación es de particular importancia conocer cuál es la máxima cantidad de información que puede pasar a través del canal *prácticamente* sin errores. La noción de transmisión

sin errores se formaliza en la siguiente definición.

**Definición B.5.** *Se dice que una velocidad  $R$  es alcanzable si existe una sucesión de códigos  $(2^{nR}, n)$  tales que  $\lambda_n \xrightarrow{n \rightarrow \infty} 0$ , donde  $\lambda_n$  es la máxima probabilidad de error sobre las palabras transmitidas por el código  $(2^{nR}, n)$ .*

Observemos que la probabilidad de error puede ser diferente para cada palabra transmitida dentro de un mismo código  $(2^{nR}, n)$ . Por lo tanto, para poder decir que el canal es confiable es necesario que los máximos  $\lambda_n$  tiendan a cero<sup>8</sup>. El siguiente teorema, conocido como Teorema de Shannon para la codificación de canales con ruido, establece una relación entre la capacidad del canal y las velocidades prácticas a las que se puede transmitir información a través del canal de manera confiable.

**Teorema B.7.** *Para un canal discreto sin memoria, todas las velocidades  $R$  debajo de la capacidad  $C$  son alcanzables. Específicamente, para toda velocidad  $R < C$  existe una secuencia de códigos  $(2^{nR}, n)$  con máxima probabilidad de error  $\lambda_n \rightarrow 0$ .*

*Recíprocamente, cualquier secuencia de códigos  $(2^{nR}, n)$  con  $\lambda_n \rightarrow 0$  satisface que  $R \leq C$ .*

Este teorema vincula la capacidad ergódica teórica y la cantidad de información que puede ser transmitida físicamente a través del canal de manera segura, es decir, relaciona el cálculo teórico y el límite práctico. De hecho, es este teorema el que le da valor al análisis de la capacidad ergódica.

Como consecuencia directa de este teorema, en el caso del BSC( $p$ ) se puede transmitir con una probabilidad de error arbitrariamente pequeña hasta  $1 + p \log_2 p + (1 - p) \log_2(1 - p)$  bits por símbolo.

Dado que la mayoría de transmisores envían datos de diversas fuentes<sup>9</sup>, el diseño de transmisores se suele basar en la suposición de que la fuente tiene una distribución uniforme en los valores que toma [Lin91], lo cual permite transmitir con el mismo código (para el canal) diversas fuentes.

Por lo tanto, es necesario codificar la información original para que ésta esté representada por palabras equiprobables, y después se codifica nuevamente para transmitirla. En otras

---

<sup>8</sup>Es necesario verificar que el máximo tiende a cero, ya que si se tomara el promedio por ejemplo, podrían existir símbolos no seguros para ninguna  $n$  pero aun así tener un promedio que tiende a cero, y en esta situación claramente el canal no sería confiable.

<sup>9</sup>Por ejemplo, una computadora transmite datos y video mediante el mismo transmisor.

palabras, si se tiene una fuente  $A$  ésta será codificada primeramente en una fuente  $A'$ , la cual utiliza palabras prácticamente equiprobables. Por último, la fuente  $A'$  es transmitida mediante un código realizado para el canal, el cual contiene la redundancia necesaria para la corrección de errores<sup>10</sup>.

## B.5. Notas Históricas

Los fundamentos de la teoría de la información se encuentran en el artículo de Claude E. Shannon “A Mathematical Theory of Communication”, el cual fue publicado en 1948 en el *Bell System Technical Journal*. En este artículo se analizó por primera vez los límites de la transmisión de datos en canales sin ruido y con ruido, lo cual revolucionó totalmente las comunicaciones. En primer lugar, fundó una teoría para predecir los límites de transmisión de los canales físicos. En segundo lugar, demostró que asintóticamente los buenos códigos existen, lo cual fundamenta el nacimiento de la teoría de códigos, la cual se dedica a encontrar buenos códigos en el régimen finito<sup>11</sup>. Esta teoría es igualmente útil en comunicación alámbrica como inalámbrica.

---

<sup>10</sup>En el diseño de códigos para el canal, a los bits de información se les agregan bits extras o de redundancia, con la intención de reparar posibles daños en el mensaje transmitido (véase [Lin91]).

<sup>11</sup>Si no se demostrara la existencia de tales buenos códigos, se podrían estar buscando códigos que no existen. Por ejemplo, si se buscaran códigos de prefijos con longitud promedio menor a la entropía, éstos no podrían ser encontrados.



## Apéndice C

# Teoría de Matrices Aleatorias

En este apéndice se resumen las principales definiciones y resultados de la Teoría de Matrices Aleatorias que se utilizan en esta tesis. Estos resultados fueron tomados principalmente de las notas de Robert Gallager [Gal08a], el libro de Antonia Tulino y Sergio Verdú [Tul04] y el artículo de Jack Silverstein [Sil09].

### C.1. Matrices Aleatorias Gaussianas

**Teorema C.1.** *Sea  $\mathbf{A}$  una matriz compleja de  $n \times n$ . Entonces*

$$\det \mathbf{A} = \prod_{i=1}^n \lambda_i(\mathbf{A}), \quad (\text{C.1})$$

donde  $\lambda_i(\mathbf{A})$  representa el eigenvalor  $i$  de  $\mathbf{A}$ .

**Definición C.1.** *Sea  $\mathbf{x} = [x_1 \dots x_n]^T$  un vector aleatorio complejo, se define la matriz de covarianza del vector como*

$$\mathbf{Q} = \mathbb{E} \begin{bmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_n) \\ \text{cov}(x_2, x_1) & \text{cov}(x_2, x_2) & \dots & \text{cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_n, x_1) & \text{cov}(x_n, x_2) & \dots & \text{cov}(x_n, x_n) \end{bmatrix}. \quad (\text{C.2})$$

**Definición C.2.** *Sea  $\mathbf{x} = [x_1 \dots x_n]^T$  un vector aleatorio tal que las  $x_i$  son v.a.s gaussianas*

nas complejas y sea  $\hat{\mathbf{x}} = [\text{Re}(\mathbf{x})^T \text{Im}(\mathbf{x})^T]^T$  con  $\text{Re}(\mathbf{x}) = [\text{Re}(x_1) \dots \text{Re}(x_n)]^T$  y  $\text{Im}(\mathbf{x}) = [\text{Im}(x_1) \dots \text{Im}(x_n)]^T$ . Decimos que  $\mathbf{x}$  es un vector Gaussiano complejo circular simétrico si

$$\mathbb{E} [(\hat{\mathbf{x}} - \mathbb{E}[\hat{\mathbf{x}}])(\hat{\mathbf{x}} - \mathbb{E}[\hat{\mathbf{x}}])^*] = \frac{1}{2} \begin{pmatrix} \text{Re}(\mathbf{Q}) & -\text{Im}(\mathbf{Q}) \\ \text{Im}(\mathbf{Q}) & \text{Re}(\mathbf{Q}) \end{pmatrix}$$

para alguna matriz Hermitiana no negativa definida  $\mathbf{Q} \in \mathbb{C}^{n \times n}$ .

De la definición anterior es claro que en el caso 1-dimensional una v.a. es gaussiana compleja circular simétrica si y sólo si tiene parte real e imaginaria gaussianas iid.

**Corolario C.1.** Sea  $\mathbf{x} = [x_1 \dots x_n]^T$  un vector aleatorio tal que las  $x_i$  son v.a.s gaussianas complejas circulares simétricas independientes, entonces el vector  $\mathbf{x}$  es un vector Gaussiano complejo circular simétrico.

El siguiente resultado afirma que la matriz gaussiana con coeficientes iid Gaussianos complejos circulares simétricos tiene una distribución invariante bajo multiplicaciones por matrices unitarias.

**Teorema C.2.** Sea  $\mathbf{H}$  una matriz de  $n_R \times n_T$  con coeficientes iid Gaussianos complejos circulares simétricos. Si  $\mathbf{U}$  es una matriz unitaria de  $n_R \times n_R$  entonces la distribución de  $\mathbf{UH}$  iguala la distribución de  $\mathbf{H}$ .

## C.2. Transformada de Stieltjes

**Definición C.3.** Sea  $X$  una variable aleatoria real con distribución  $F_X$ . Su transformada de Stieltjes está definida para argumentos complejos  $z \in \mathbb{C}^+$  como

$$m_X(z) = \mathbb{E} \left[ \frac{1}{X - z} \right] = \int_{-\infty}^{\infty} \frac{1}{x - z} dF_X(x). \quad (\text{C.3})$$

**Teorema C.3.** Sea  $X$  una v.a. con función de distribución  $F_X$  y transformada de Stieltjes  $m_X$ . Si  $x_0 \in \mathbb{R}$  es tal que el límite  $\lim_{z \in \mathbb{C}^+ \rightarrow x_0} \text{Im}(m_X(z))$  existe, entonces  $F_X$  es diferenciable en  $x_0$  y la función de densidad de  $X$  está dada por

$$f_X(x_0) = \frac{1}{\pi} \lim_{z \in \mathbb{C}^+ \rightarrow x_0} \text{Im}(m_X(z)). \quad (\text{C.4})$$

**Teorema C.4.** Sean  $F$  y  $G$  dos funciones de distribución de probabilidad, entonces  $F = G$  si y sólo si  $m_F(z) = m_G(z)$  para toda  $z \in \mathbb{C}^+$ .

Por simplicidad, se utiliza la notación  $m_F(z)$  para denotar la transformada de Stieltjes de la v.a. cuya función de distribución es la función  $F$ .

**Definición C.4.** Se define el conjunto  $\mathcal{M}(\mathbb{R})$  de funciones de distribución de subprobabilidad en los reales, como el conjunto de todas las funciones  $G$  tales que  $G = \alpha F$  donde  $\alpha \in [0, 1]$  y  $F$  es una función de distribución de probabilidad.

**Definición C.5.** Sea  $\{F_n\}_{n \geq 1} \subset \mathcal{M}(\mathbb{R})$ , entonces  $F_n$  converge vagamente a  $F \in \mathcal{M}(\mathbb{R})$  si para cualesquiera  $a$  y  $b$  puntos de continuidad de  $F$  se tiene que  $\lim_{n \rightarrow \infty} F_n\{[a, b]\} = F\{[a, b]\}$ . Esto se denota por  $F_n \xrightarrow{v} F$ .

**Teorema C.5.** Sea  $\{F_n\}_{n \geq 1} \subset \mathcal{M}(\mathbb{R})$ , entonces  $F_n$  converge vagamente a  $F \in \mathcal{M}(\mathbb{R})$  si y sólo si  $m_{F_n}(z) \rightarrow m_F(z)$  para toda  $z \in \mathbb{C}^+$ .

**Definición C.6.** Sea  $X$  una v.a. no negativa, entonces la transformada  $\eta$  de  $X$  se define como

$$\eta_X(\gamma) = \mathbb{E} \left[ \frac{1}{1 + \eta X} \right], \quad (\text{C.5})$$

donde  $\gamma$  es un número real no negativo.

Mediante continuación analítica la transformada  $\eta$  puede extenderse a  $\mathbb{C}^+$  (véase [Tul04]), y mediante los cálculos pertinentes se llega al siguiente corolario.

**Corolario C.2.** Sea  $X$  una v.a. real, entonces la transformada  $\eta$  de  $X$  y la transformada de Stieltjes de  $X$  satisfacen que

$$\eta_X(\gamma) = \frac{m_X\left(-\frac{1}{\gamma}\right)}{\gamma} \quad (\text{C.6})$$

para toda  $\gamma \in \mathbb{C}^+$ .

### C.3. Teoremas Tipo Marchenko-Pastur

**Definición C.7.** Sea  $\mathbf{A}$  una matriz autoadjunta de  $n \times n$ , la función de distribución empírica espectral de  $\mathbf{A}$  denotada por  $F^{\mathbf{A}}(x) : \mathbb{R} \rightarrow \mathbb{R}$  se define como

$$F^{\mathbf{A}}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\lambda_i(\mathbf{A}) \leq x\}}, \quad (\text{C.7})$$

donde  $\lambda_i(\mathbf{A})$  denota el eigenvalor  $i$  de  $\mathbf{A}$ .

Si  $\mathbf{A}$  es una matriz Hermitiana positiva semidefinida, entonces denotamos por  $m_{\mathbf{A}}(z)$  a la transformada de Stieltjes de la función de distribución empírica espectral de  $\mathbf{A}$ .

**Definición C.8.** Sea  $\beta > 0$ , se define la distribución de Marchenko-Pastur  $F_{\beta}(x)$  de parámetro  $\beta$  como la función de distribución tal que

$$F_{\beta}(dx) = (1 - \beta)^+ \delta_0(dx) + \frac{1}{2\pi x} \sqrt{(r - x)^+(x - l)^+} dx, \quad (\text{C.8})$$

donde  $(x)^+ = \max(0, x)$ ,  $l = (1 - \sqrt{\beta})^2$  y  $r = (1 + \sqrt{\beta})^2$ .

**Teorema C.6.** Sea  $\mathbf{H}$  una matriz compleja de  $n_R \times n_T$  cuyas coeficientes son independientes con media cero, varianza unitaria y momentos cuartos de orden  $O(1/n_R^2)$ . Supongamos que  $n_T/n_R \rightarrow \beta$  cuando  $n_R \rightarrow \infty$ . Denotemos la distribución empírica espectral de  $\mathbf{H}\mathbf{H}^*/n_R$  por  $F^{\mathbf{H}\mathbf{H}^*/n_R}(x)$ . Entonces, cuando  $n_R \rightarrow \infty$  se tiene que  $F^{\mathbf{H}\mathbf{H}^*/n_R}(x) \xrightarrow{\text{c.s.}} F_{\beta}(x)$ , donde  $F_{\beta}(x)$  es la distribución de Marchenko-Pastur de parámetro  $\beta$ .

**Teorema C.7.** Supóngase que

- a. Para cada  $n_R = 1, 2, \dots$  sea  $\mathbf{H}_{n_R}$  una matriz de  $n_R \times n_T$  con coeficientes complejos idénticamente distribuidos para toda  $n$ ,  $i$ ,  $j$ .
- b. Los coeficientes  $\mathbf{H}_{n_R}(i, j)$  son independientes para cada  $n_R$  y  $\text{var}(\mathbf{H}_{n_R}(i, j)) = 1$ .
- c.  $n_T = n_T(n_R)$  con  $n_T/n_R \rightarrow \beta > 0$  cuando  $n_R \rightarrow \infty$ .
- d.  $\mathbf{R}_{n_T} = \text{diag}(r_{n_T}(1), \dots, r_{n_T}(n_T))$  con  $r_{n_T}(i) \in \mathbb{R}$  y la función de distribución empírica de  $\{r_{n_T}(1), \dots, r_{n_T}(n_T)\}$  converge débilmente, con probabilidad uno, a una función de distribución no aleatoria  $R(x)$  cuando  $n_R \rightarrow \infty$ .

e.  $\mathbf{A}_{n_R}$  es una matriz aleatoria de  $n_R \times n_R$  Hermitiana tal que, con probabilidad uno,  $F^{\mathbf{A}_{n_R}} \xrightarrow{v} F^{\mathbf{A}}$  donde  $F^{\mathbf{A}}$  es una función de distribución no aleatoria.

f.  $\mathbf{H}_{n_R}$ ,  $\mathbf{R}_{n_T}$  y  $\mathbf{A}_{n_R}$  son independientes.

Sea  $\mathbf{B}_{n_R} = \mathbf{A}_{n_R} + \mathbf{H}_{n_R} \mathbf{R}_{n_T} \mathbf{H}_{n_R}^* / n_R$ , entonces con probabilidad uno  $F^{\mathbf{B}_{n_R}} \xrightarrow{v} F^{\mathbf{B}}$  cuando  $n_R \rightarrow \infty$ , donde  $F^{\mathbf{B}}$  es una función de distribución de probabilidad no aleatoria la cual satisface

$$m_{\mathbf{B}}(z) = m_{\mathbf{A}} \left( z - \beta \int \frac{x}{1 + x m_{\mathbf{B}}(z)} dR(x) \right) \quad (\text{C.9})$$

para todo  $z \in \mathbb{C}^+$ . Además,  $m_{\mathbf{B}}(z)$  es la única solución a la ecuación anterior con parte imaginaria positiva.

**Teorema C.8.** *Supóngase que*

a. Para cada  $n_R = 1, 2, \dots$  sea  $\mathbf{H}_{n_R}$  una matriz de  $n_R \times n_T$  con coeficientes complejos idénticamente distribuidos para toda  $n$ ,  $i, j$ .

b. Los coeficientes  $\mathbf{H}_{n_R}(i, j)$  son independientes para cada  $n_R$  y  $\text{var}(\mathbf{H}_{n_R}(i, j)) = 1$ .

c.  $n_T = n_T(n_R)$  con  $n_T/n_R \rightarrow \beta > 0$  cuando  $n_R \rightarrow \infty$ .

d.  $\mathbf{R}_{n_R} = \text{diag}(r_{n_R}(1), \dots, r_{n_R}(n_R))$  con  $r_{n_R}(i) \in \mathbb{R}^+ \cup \{0\}$  y la función de distribución empírica de  $\{r_{n_R}(1), \dots, r_{n_R}(n_R)\}$  converge en distribución, con probabilidad uno, a una función de distribución no aleatoria  $R(x)$  cuando  $n_R \rightarrow \infty$ .

e.  $\mathbf{H}_{n_R}$  y  $\mathbf{R}_{n_R}$  son independientes.

Sea  $\mathbf{R}_{n_R}^{1/2}$  cualquier raíz cuadrada Hermitiana de  $\mathbf{R}_{n_R}$ . Defínase  $\mathbf{B}_{n_R} = \mathbf{R}_{n_R}^{1/2} \mathbf{H} \mathbf{H}^* \mathbf{R}_{n_R}^{1/2} / n_T$ , entonces, con probabilidad uno  $F^{\mathbf{B}_{n_R}} \xrightarrow{d} F^{\mathbf{B}}$  cuando  $n_R \rightarrow \infty$  donde, para todo  $z \in \mathbb{C}^+$  la transformada  $m_{\mathbf{B}}(z)$  satisface

$$m_{\mathbf{B}}(z) = \int \frac{dR(x)}{x(1 - c - c z m_{\mathbf{B}}(z)) - z}. \quad (\text{C.10})$$

Además,  $m_{\mathbf{B}}(z)$  es la única solución a la ecuación anterior tal que  $-(1 - c)/z + c m_{\mathbf{B}}(z) \in \mathbb{C}^+$ , donde  $c := 1/\beta$ .

**Teorema C.9.** *Supóngase que*

- a. Para cada  $n_R = 1, 2, \dots$  sea  $\mathbf{H}_{n_R}$  una matriz de  $n_R \times n_T$  con coeficientes complejos idénticamente distribuidos para toda  $n, i, j$ .
- b. Los coeficientes  $\mathbf{H}_{n_R}(i, j)$  son independientes para cada  $n_R$  y  $\text{var}(\mathbf{H}_{n_R}(i, j)) = 1$ .
- c.  $n_T = n_T(n_R)$  con  $n_T/n_R \rightarrow \beta > 0$  cuando  $n_R \rightarrow \infty$ .
- d.  $\mathbf{R}_{n_R}$  es una matriz de  $n_R \times n_T$ , independiente de  $\mathbf{H}_{n_R}$ , con  $F^{\mathbf{R}_{n_R} \mathbf{R}_{n_R}^*/n_T}(x) \xrightarrow{d} R(x)$  casi seguramente cuando  $n_R \rightarrow \infty$ , donde  $R(x)$  es una función de distribución no aleatoria.
- e.  $\mathbf{H}_{n_R}$  y  $\mathbf{R}_{n_R}$  son independientes.

Sea  $\mathbf{B}_{n_R} := (\mathbf{R}_{n_R} + \sigma \mathbf{H}_{n_R})(\mathbf{R}_{n_R} + \sigma \mathbf{H}_{n_R})^*/n_T$  con  $\sigma > 0$  no aleatoria, entonces, con probabilidad uno  $F^{\mathbf{B}_{n_R}} \xrightarrow{d} F^{\mathbf{B}}$  cuando  $n_R \rightarrow \infty$  donde, para cada  $z \in \mathbb{C}^+$  se tiene que  $m_{\mathbf{B}}(z)$  satisface

$$m_{\mathbf{B}}(z) = \int \frac{dR(x)}{\frac{x}{1+\sigma^2 c m_{\mathbf{B}}(z)} - (1 + \sigma^2 c m_{\mathbf{B}}(z))z + \sigma^2(1-c)}. \quad (\text{C.11})$$

Además,  $m_{\mathbf{B}}(z)$  es la única solución a la ecuación anterior tal que  $\text{Im}(z m_{\mathbf{B}}(z)) \geq 0$ , donde  $c := 1/\beta$ .

**Definición C.9.** Sea  $\mathbf{H}$  una matriz de  $n_R \times n_T$  tal que  $n_T/n_R \rightarrow \beta$  cuando  $n_R \rightarrow \infty$ . Se dice que  $\mathbf{H}$  satisface la condición de Lindeberg si

$$\frac{1}{n_R} \sum_{i,j} \mathbb{E} \left[ \frac{|H_{ij}|^2}{n_R} \mathbb{1}_{\{|H_{i,j}|/\sqrt{n_R} \geq \delta\}} \right] \xrightarrow{n_R \rightarrow \infty} 0 \quad (\text{C.12})$$

para toda  $\delta > 0$ .

**Teorema C.10.** Sea  $\mathbf{H} = \mathbf{C}\mathbf{S}\mathbf{A}$  con  $\mathbf{S}$  una matriz de  $n_R \times n_T$  con coeficientes aleatorios independientes complejos (arbitrariamente distribuidos) que satisface la condición de Lindeberg. Además, supóngase que los coeficientes tienen medias idénticas y varianzas unitarias. Sean  $\mathbf{D} = \mathbf{C}\mathbf{C}^*$  y  $\mathbf{T} = \mathbf{A}\mathbf{A}^*$  matrices de  $n_R \times n_R$  y  $n_T \times n_T$  respectivamente tales que los espectros asintóticos de  $\mathbf{D}$  y  $\mathbf{T}$  convergen casi seguramente a medidas de soporte compacto. Si  $\mathbf{C}$ ,  $\mathbf{S}$  y  $\mathbf{A}$  son independientes, entonces cuando  $n_R \rightarrow \infty$  con  $n_T/n_R \rightarrow \beta$ , se tiene que la transformada  $\eta$  de  $\mathbf{H}\mathbf{H}^*/n_R$  está dada por

$$\eta_{\mathbf{H}\mathbf{H}^*/n_R}(\gamma) = \mathbb{E}[\Gamma_{\mathbf{H}\mathbf{H}^*/n_R}(D, \gamma)], \quad (\text{C.13})$$

donde  $z \in \mathbb{C}^+$  y  $\Gamma_{\mathbf{H}\mathbf{H}^*/n_R}$  satisface

$$\Gamma_{\mathbf{H}\mathbf{H}^*/n_R}(d, \gamma) = \left( 1 + \gamma \beta d \mathbb{E} \left[ \frac{T}{1 + \gamma T \mathbb{E}[D \Gamma_{\mathbf{H}\mathbf{H}^*/n_R}(D, \gamma)]} \right] \right)^{-1} \quad (\text{C.14})$$

con  $D$  y  $T$  v.a.s independientes con la distribución asintótica del espectro de  $\mathbf{D}$  y  $\mathbf{T}$  respectivamente.

Utilizando la relación entre la transformada  $\eta$  y la transformada de Stieltjes se obtiene el siguiente corolario al teorema anterior.

**Corolario C.3.** *Sea  $\mathbf{H} = \mathbf{C}\mathbf{S}\mathbf{A}$  con  $\mathbf{S}$  una matriz de  $n_R \times n_T$  con coeficientes aleatorios independientes complejos (arbitrariamente distribuidos) que satisface la condición de Lindeberg. Además, supóngase que los coeficientes tienen medias idénticas y varianzas unitarias. Sean  $\mathbf{D} = \mathbf{C}\mathbf{C}^*$  y  $\mathbf{T} = \mathbf{A}\mathbf{A}^*$  matrices de  $n_R \times n_R$  y  $n_T \times n_T$  respectivamente tales que los espectros asintóticos de  $\mathbf{D}$  y  $\mathbf{T}$  convergen casi seguramente a medidas de soporte compacto. Si  $\mathbf{C}$ ,  $\mathbf{S}$  y  $\mathbf{A}$  son independientes, entonces cuando  $n_R \rightarrow \infty$  con  $n_T/n_R \rightarrow \beta$ , se tiene que la transformada de Stieltjes asintótica de  $B := \mathbf{H}\mathbf{H}^*/n_R$  está dada por*

$$m_{\mathbf{B}}(z) = \mathbb{E}[\mathcal{B}(D, z)], \quad (\text{C.15})$$

donde  $z \in \mathbb{C}^+$  y  $\mathcal{B}$  satisface

$$\mathcal{B}(d, z) = \left( -z + z \beta d \mathbb{E} \left[ \frac{T}{z + z T \mathbb{E}[D \mathcal{B}(D, z)]} \right] \right)^{-1} \quad (\text{C.16})$$

con  $D$  y  $T$  v.a.s independientes con la distribución asintótica del espectro de  $\mathbf{D}$  y  $\mathbf{T}$  respectivamente.



# Bibliografía

- [Big07] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj y H. Poor, *MIMO Wireless Communication*. Cambridge University Press, 2007.
- [Bil78] P. Billingsley, *Ergodic Theory and Information*. Robert E. Krieger Publishing Company, 1978.
- [Bli04] D. Bliss, A. Chan y N. Chang, “MIMO wireless communication channel phenomenology,” *IEEE Trans. on Antennas and Propagation*, vol. 52, pág. 2073-2082, 2004.
- [Coe95] C. Coello, “Introducción a los algoritmos genéticos,” *Tecnologías de Información y Estrategias de Negocios*, año 3, No. 17, pág. 5-11, enero 1995.
- [Cou10] R. Couillet, M. Debbah y J. Silverstein, “A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels,” enviado a *IEEE Trans. on Information Theory*, 2010.
- [Cov06] T. Cover y J. Thomas, *Elements of Information Theory. Second Edition*. Wiley, 2006.
- [Fos96] G. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Laboratories Technical Journal*, pág. 41-59, octubre 1996.
- [Gal08] R. Gallager, *Principles of Digital Communication*. Cambridge University Press, 2008.

- [Gal08a] R. Gallager, “Circularly-Symmetric Gaussian random vectors,” *Notas de Clase*, pág. 1-9, enero 2008.
- [Gui77] S. Guiasu, *Information Theory with Applications*. McGraw-Hill, 1977.
- [Hal50] P. Halmos, *Measure Theory*. Springer-Verlag, 1950.
- [Hoe71] P. Hoel, S. Port y C. Stone, *Introduction to Probability Theory*. Houghton Mifflin Company, 1971.
- [Jaf05] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge University Press, 2005.
- [Jin10] S. Jin, M. McKay y K. Wong, “Ergodic capacity analysis of amplify-and-forward MIMO dual hop systems,” *IEEE Trans. on Information Theory*, vol. 56, pág. 2204-2224, mayo 2010.
- [Ker02] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen y F. Frederiksen, “A stochastic MIMO radio channel model with experimental validation,” *IEEE J. Select. Areas Commun.*, vol. 20, pág. 1211-1226, 2002.
- [Lin91] J. van Lint, *Introduction to Coding Theory. Third Revised and Expanded Edition*. Springer-Verlag, 1991.
- [Liu10] R. Liu, T. Liu, H. Poor y S. Shamai, “MIMO Gaussian broadcast channels with confidential messages,” *IEEE Trans. on Information Theory*, vol. 56, pág. 4215-4227, septiembre 2010.
- [PWJ08] Proyecto WiMAX Jalisco, *Documento Informativo del Proyecto*. 2008.  
[http://www.cudi.edu.mx/boletin/2008/Proyecto\\_Wimax.pdf](http://www.cudi.edu.mx/boletin/2008/Proyecto_Wimax.pdf)
- [Qua10] Qualcomm, *Presentación Informativa sobre los Beneficios del LTE-Advanced*. 2010.  
<http://www.qualcomm.com/documents/files/lte-advanced-benefits.pdf>
- [Sha48] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pág. 379-423 y 623-656, julio y octubre 1948.

- [Sha06] J. Sharony, "Introduction to wireless MIMO: theory and applications," *IEEE LI*, noviembre 2006.
- [Sil09] J. Silverstein, "The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices," *Random Matrix Theory and Its Applications*, vol. 18, pág. 1-25, 2009.
- [Tel99] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommunications*, vol. 10, pág. 585-595, noviembre-diciembre 1999.
- [Tse05] D. Tse y P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [Tul04] A. Tulino y S. Verdú, *Random Matrix Theory and Wireless Communications*. Now Publishers Inc, 2004.
- [Tul05] A. Tulino, A. Lozano y S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. on Information Theory*, vol. 51, No. 7, pág. 2491-2509, 2005.
- [Yu01] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson y M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2GHz measurements," *Proc. IEEE Global Telecomm. Conf.*, vol. 1, pág. 25-29, 2001.

