Departamento de
Electrónica e Sistemas
UNIVERSIDADE DA CORUÑA

PhD Thesis

Blind Channel Estimation for Space-Time Block Codes: Novel Methods and Performance Studies

Héctor José Pérez Iglesias

July 2010

PhD Advisor:

Adriana Dapena Janeiro

Dra. Adriana Dapena

CERTIFICA:

Que la memoria titulada "Blind Channel Estimation for Space-Time Block Codes: Novel Methods and Performance Studies", ha sido realizada por D. Héctor José Pérez Iglesias bajo mi dirección en el Departamento de Electrónica e Sistemas de la Universidade da Coruña y concluye la Tesis que presenta para optar al grado de Doctor con la mención de Doctor Europeo.

A Coruña, 26 de abril de 2010

Fdo: Dra. Adriana Dapena Janeiro
Directora de la Tesis Doctoral
Titular de Universidad
Dpto. de Electrónica e Sistemas Universidade da Coruña

Tesis Doctoral:	Blind Channel Estimation for Space-Time Block Codes: Novel Methods and Performance Studies
Autor:	D. Héctor José Pérez Iglesias
Director:	Dra. Adriana Dapena Janeiro
Fecha:	8 de julio de 2010

Tribunal

Presidente:

Vocal 1:

Vocal 2:

Vocal 3:

Secretario:

Agradecimientos

En primer lugar me gustaría manifestar mi agradecimiento a mi directora de tesis, Adriana Dapena. Gracias por todo lo que me has enseñado, por guiarme, apoyarme y motivarme durante toda la etapa de realización del doctorado, y hacerme tan fácil la tarea de componer los artículos que hemos publicado en congresos y revistas.

Quiero dar gracias también a Luis Castedo, quien ha sido uno de los mejores profesores que he tenido durante la realización de la carrera, y quien me ha motivado y facilitado recorrer el díficil camino de la docencia en la universidad. No puedo olvidar la grata experiencia de compartir la impartición de materias contigo.

Me gustaría agradecer a Vicente Zarzoso su calurosa acogida durante mi estancia en Niza, y la productividad de las discusiones científicas que allí tuvimos, junto con las de diversos congresos en los que hemos coincidido. Además, no me quiero olvidar de darle las gracias por su importante colaboración en los artículos que hemos publicado con Adriana y Luis.

Quiero agredecer a Carlos Escudero, gran docente que imparte geniales materias, la motivación que me ha inspirado por el mundo de las señales y la comunicación. Gracias Lamas, director de mi proyecto de fin de carrera de la ingeniería, quien tanto me ayudó y guió en la realización del mismo en plazos tan ajustados. Dani, gracias por hacer siempre un ambiente mejor en todas las reuniones de grupo y departamento, y ayudarme en la impartición de materias. Cris, gracias, contigo el departamento y el grupo ha cambiado mucho desde que llegaste, nos has hecho más fácil y divertido el día a día, y más festivas las reuniones. Gracias Jose, es genial haber compartido y compartir tantos cafés, trabajos y trabajillos del mundo de la electrónica y la mecánica. Gracias Fran, todas las conversaciones que hemos tenido en el laboratorio han sido muy productivas y amenas. Gracias Santi, por ser el Arduino Man, Josmary, Javi y Ángel, hay viajes que no se olvidarán. Miguel, gracias, por guiarme y ayudarme tanto a tomar decisiones, y por nuestras conversaciones trascendentales. Paula, gracias por la preocupación y ayuda que

ofreces siempre a todo el grupo.

Gracias Alba por ser tan nena, Carla por ser tan 80, Carlos por ser de leyenda, David por ser tan primo abuelo, Iván por ser tan Nike-Sony-Gasolina, Iván porque la dislexia colectiva no sería lo mismo sin ti, el tráfico de Internet sería la mitad y las ventanillas y puertas estarían más tranquilas, Paulo por ser una panda de flipaos y Sergio por formar parte de ella.

Gracias Diego Cifrian Bueno, Mau y Moro, si tuviera que poner el motivo sería un CRC.

Gracias Tebas por ser tan pesado, Kobe por ser la fiesta, Diego por ser tan pirotecnias y Goday por ser el mejor okupa. A los cuatro por ser la música, cada uno a su modo. Gracias Adri por ser tan chancarella, Yuli por ser tan chuleta y mimosín, Oscar por ser tan Huuuuuuu, Yan por gustarte la carne y Nan por Never. Gracias Porta y Gracia, mis buenos compañeros del café y sandwich.

Sonia, muchas gracias por estar a mi lado y como no quiero poner noñerías, no pongo nada más.

A todos mis primos, en especial a mi prima Rosa por preocuparse tanto por mí y por ser así. A todos mis tíos y tías, en especial a mi tía Manola y a mi tía Carmen.

Gracias papá, mamá y bro, que suerte he tenido con vosotros.

Por último quiero manifestar mi agradecimiento al Ministerio de Educación por el soporte económico que me ha ofrecido durante la realización de la tesis, mediante la beca FPU con referencia AP2005-5014.

Resumen

En este trabajo se realiza un estudio de técnicas de separación ciega de fuentes para la estimación de los coeficientes en sistemas de transmisión que emplean la codificación de Alamouti con 2 antenas transmisoras y 1 antena receptora. La mayoría de los estándares actuales incluyen símbolos piloto para estimar el canal en recepción. Dado que estos símbolos no transportan datos del usuario, su utilización decrementa la tasa de transferencia y degrada el rendimiento del sistema. Por otro lado, los algoritmos de separación ciega son menos precisos en la estimación de los coeficientes de canal que los supervisados pero consiguen una tasa de transferencia mayor.

En el presente trabajo, modelaremos el sistema de codificación de Alamouti como un problema típico de separación ciega de fuentes donde las señales transmitidas y los coeficientes del canal deben ser estimados a partir de mezclas lineales e instantáneas (observaciones). La estructura ortogonal impuesta por la codificación de Alamouti permite resolver este problema mediante la descomposición de autovalores y autovectores de matrices calculadas a partir de diferentes estadísticos de las observaciones. Estos algoritmos pueden ser clasificados en aquellos que utilizan estadísticos de segundo orden y aquellos que emplean estadísticos de orden superior.

Los algoritmos que emplean estadísticos de segundo orden trabajan con la matriz de correlación de las observaciones, son computacionalmente poco costosos pero requieren de un precodificador lineal para descompensar la potencia de las señales transmitidas. Una de nuestras aportaciones es la de determinar de forma empírica cómo debe realizarse la descompesación de potencia de cara a reducir la probabilidad de error del sistema.

Por otro lado, los algoritmos que trabajan con estadísticos de orden superior se basan en diagonalizar una o varias matrices de cumulantes de orden superior, lo que conlleva un mayor coste computacional en el receptor. Como ventaja debe resaltarse que no requieren incluir un precodificador lineal que realice la descompensación de potencia. En este trabajo mostraremos que el rendimiento de estas técnicas depende del grado de

dispersión de los autovalores de la matriz que se diagonaliza. Utilizaremos esta idea para obtener la matriz de cumulantes óptima y para formular un nuevo algoritmo que supera en rendimiento a los propuestos previamente por otros autores.

Otra aportación relevante del presente trabajo es presentar una detallada comparación de las técnicas de estimación de canal en entornos simulados, considerando canales con ditribución Rayleigh y Rice, y en entornos reales en la banda ISM de 2.4 GHz mediante el empleo de una plataforma de transmisión MIMO desarrollada en la Universidade da Coruña.

Summary

This work is based on a study of blind source separation techniques in order to estimate coefficients in transmission systems using Alamouti codification with two transmit antennas and one receive antenna. Most of present standards include pilot symbols to estimate the channel in reception. Since these symbols do not deliver user's data, their use decrease transferring quantity and also the system capacity. On the other hand, algorithms of blind separation are less precise when estimating channel coefficients than those supervised, but achieving a higher transferring rate.

In this work we will deal with Alamouti codification system as a typical problem of blind sources separation where the signals transmitted and the channel coefficients must be estimated according to lineal and instantaneous mixtures (observations). Orthogonal structure required by Alamouti codification allows us to solve this problem by decomposing eigenvalues and eigenvectors of matrices calculated from different statistics of the observations. These algorithms could be classified as those using second order statistics and those using higher order statistics.

Algorithms based on second order statistics work with correlation matrix of observations. They are computationally less expensive, but require a lineal precoder in order to balance the power of the signals transmitted. One of our contributions is being able to determine in an empirical way how the power decompensation should be done in order to reduce the proabibility of error in the system.

On the other hand, algorithms dealing with high level statistics are based on diagonalize one or several high level cumulant matrices deriving into a major computational cost in the receiver. As an advantage we must point out that they do not require to include a lineal precoder to do the power decompensation. In this work we will prove that the output of these techniques depends on the level of eigenvalue of the diagonalized matrix spreading. This idea will be used by us in order to achieve the optimal cumulant matrix and also to propose a new algorithm that increases the output in relation to those already

proposed by other authors.

Another important contribution of this present study is to propose a detailed comparison between channel estimation techniques in simulated scenarios, considering channels with Rayleigh and Rice distribution, and in real scenarios in ISM of 2.4 GHz band, by using a MIMO testbed developed in Universidade da Coruña.

1. Resumen

En la tesis se realiza un estudio de técnicas de separación ciega de fuentes para la estimación de los coeficientes en sistemas de transmisión que emplean la codificación de Alamouti con 2 antenas transmisoras y 1 antena receptora. La mayoría de los estándares actuales incluyen símbolos piloto para estimar el canal en recepción. Dado que estos símbolos no transportan datos del usuario, su utilización decrementa la tasa de transferencia y degrada el rendimiento del sistema. Por otro lado, los algoritmos de separación ciega son menos precisos en la estimación de los coeficientes de canal que los supervisados pero consiguen una tasa de transferencia mayor.

Nosotros modelaremos el sistema de codificación de Alamouti como un problema típico de separación ciega de fuentes donde las señales transmitidas y los coeficientes del canal deben ser estimados a partir de mezclas lineales e instantáneas (observaciones). La estructura ortogonal impuesta por la codificación de Alamouti permite resolver este problema mediante la descomposición de autovalores y autovectores de matrices calculadas a partir de diferentes estadísticos de las observaciones. Estos algoritmos pueden ser clasificados en aquellos que utilizan estadísticos de segundo orden y aquellos que emplean estadísticos de orden superior.

Los algoritmos que emplean estadísticos de segundo orden trabajan con la matriz de correlación de las observaciones, son computacionalmente poco costosos pero requieren de un precodificador lineal para descompensar la potencia de las señales transmitidas. Una de nuestras aportaciones es la de determinar de forma empírica cómo debe realizarse la descompensación de potencia de cara a reducir la probabilidad de error del sistema.

Por otro lado, los algoritmos que trabajan con estadísticos de orden superior se basan en diagonalizar una o varias matrices de cumulantes de orden superior, lo que conlleva un mayor coste computacional en el receptor. Como ventaja debe resaltarse que no requieren incluir un precodificador lineal que realice la descompensación de potencia. En la tesis mostraremos que el rendimiento de estas técnicas depende del grado de dispersión de los

autovalores de la matriz que se diagonaliza. Esta idea será utilizada por nosotros para obtener la matriz de cumulantes óptima y para formular un nuevo algoritmo que supera en rendimiento a los propuestos previamente por otros autores.

Otra aportación relevante del trabajo es el presentar una detallada comparación de las técnicas de estimación de canal en entornos simulados, considerando canales con distribución Rayleigh y Rice, y en entornos reales en la banda ISM de 2.4 GHz, mediante el empleo de una plataforma de transmisión MIMO desarrollada en la Universidade da Coruña.

2. Metodología

El primer paso en la elaboración de la tesis ha sido la revisión de la bibliografía existente en el ámbito de la separación ciega de fuentes y en el ámbito de transmisión sobre canales MIMO. Se han estudiado los algoritmos más importantes de separación ciega de fuentes, realizándose su implementación y se ha comprobado su funcionamiento en los entornos de procesado de imagen, procesado de audio y comunicaciones digitales. De forma paralela, se han estudiado los sistemas que trabajan con canales MIMO y se han implementado y utilizado los tradicionales métodos de estimación de canal para los mismos.

Posteriormente, se han estudiado las aplicaciones existentes de separación ciega de fuentes para la estimación de canal en sistemas MIMO que utilizan OSTBC (del inglés, Orthogonal Space-Time Block Codes). En un paso siguiente, estos algoritmos han sido particularizados para los sistemas que emplean la codificación de Alamouti con 2 antenas transmisoras y 1 antena receptora. Además, se han desarrollado nuevos algoritmos que utilizan estadísticos de segundo orden y nuevos algoritmos que emplean estadísticos de orden superior.

Se ha hecho una comparativa de rendimiento, mediante simulaciones de ordenador, entre los algoritmos existentes y los nuevos realizando la transmisión por canales que

siguen distribuciones Rayleigh y distribuciones Rice.

Para obtener una mayor representatividad de los resultados obtenidos, el funcionamiento de los algoritmos ha sido cotejado en un entorno interior de transmisión real de 2.4 GHz, mediante una plataforma desarrollada por el Grupo de Tecnología Electrónica y Comunicaciones (GTEC) de la Universidade da Coruña. Se han utilizado dos tipos de escenarios: con visión de línea directa y sin visión de línea directa. Los resultados indican el adecuado funcionamiento de los nuevos algoritmos desarrollados.

Los canales obtenidos han sido sometidos a un contraste estadístico para determinar su correspondencia con una determinada distribución.

Finalmente, se han obtenido las conclusiones y las posibles líneas futuras del trabajo realizado.

3. Conclusiones y Aportaciones

El principal objetivo de este trabajo es estudiar y comparar diferentes tipos de técnicas, basadas en la descomposición de autovectores y autovalores, para la estimación de la matriz y de canal (y la recuperación de las señales transmitidas) en sistemas de comunicaciones que emplean la codificación de Alamouti con 2 antenas transmisoras y 1 antena receptora.

La ortogonalidad que imponen los OSTBC ha sido utilizada para diseñar algoritmos específicos que estimen los parámetros de canal calculando los autovalores y autovectores de una matriz formada por estadísticos calculados a partir de las observaciones, esto es, calcular una EVD (del inglés, Eigenvalue Decomposition). En la tesis se clasifican estos algoritmos en dos clases teniendo en cuenta el tipo de estadísticos que se utilizan en la estimación de la matriz: métodos SOS (del inglés, Second Order Statistics) y métodos HOS (Higher Order Statistics).

El funcionamiento de los métodos basados en SOS radica la diagonalización de la matriz de correlación de las observaciones. Hemos mostrado que la matriz de canal puede ser identificada solo cuando las señales transmitidas tienen diferente varianza. En la tesis, se han estudiado diferentes métodos alternativos para evitar esta limitación. El primero de ellos, ha sido el propuesto por Shahbazpanahi y otros. Estos autores mostraron que la estimación de canal basada en SOS para la codificación de Alamouti puede realizarse mediante una simple matriz que actúe como precodificador lineal, escogida de forma que garantice que los autovalores de la matriz de autocorrelación tendrán orden 1, lo cuál puede obtenerse utilizando una simple matriz de precodificación situada antes de la esquema de codificación de Alamouti. En este sentido, Vía y otros han propuesto implementar esta procedimiento basado en SOS considerando una representación "real" en lugar de "compleja" del esquema de codificación de Alamouti. En la tesis hemos derivado una simple implementación de este algoritmo empleando notación de "complejos".

Otra importante contribución de la tesis es la presentación de los resultados de simulaciones obtenidos en diferentes escenarios que nos permiten determinar de forma empírica la matriz óptima de precodificación que ha de ser usada para los métodos SOS. Nuestros resultados de simulación muestran que el método SOS que emplea el parámetro óptimo de descompensación proporciona un adecuado rendimiento medio, aunque existe una consecuencia inmediata y no deseada: la descompensación proporcionada por el paso de precodificación, la tasa de error para una de las dos señales puede ser considerablemente incrementada.

Para los métodos basados en HOS se considera la diagonalización de una matriz formada por cumulantes de 4^o orden. Esta idea fue propuesta inicialmente por Beres y otros. Los autores proponen emplear una matriz que utiliza un número reducido de cumulantes de 4^o orden. Mostramos que el rendimiento de este método puede mejorarse haciendo una selección más adecuada de los cumulantes de 4^o orden utilizados. De hecho, se obtiene una mejora sustancial cuando la matriz diagonalizada se calcula a partir de las combinación lineal de los cumulantes de 4^o orden empleados por Beres y otros.

La contribución más importante de la tesis es la obtención de una forma cerrada para la matriz óptima de cumulantes de 4 orden que maximiza la dispersión de los autovalores. Nuestros análisis muestran que la matriz óptima depende de un parámetro γ, cuyos valores han de ser calculados considerando el valor concreto de los coeficientes de canal. Los resultados de las simulaciones verifican que el rendimiento de este método coincide con el rendimiento obtenido cuando el receptor conoce perfectamente los parámetros de canal. Hemos determinado también, una sencilla forma de estimar el parámetro γ mediante los cumulantes de 4 orden de las observaciones, pero, desafortunadamente se ha observado una degradación del rendimiento para altas SNR debido a los errores de estimación.

Otra importante contribución es el proponer un método subóptimo que selecciona la matriz con mayor dispersión de autovalores entre un conjunto de dos matrices. Este método presenta un rendimiento satisfactorio comparándolo con otras técnicas basadas en SOS o HOS. Por lo tanto, el método subóptimo es adecuada para su implementación real en FPGAs y DSPs: calcula pocos cumulantes y diagonaliza una matriz 2x2. Además, el parámetro γ es simplemente utilizado como umbral en el método subóptimo, haciéndolo menos sensible a errores de estimación.

El comportamiento de los métodos basados en SOS y HOS ha sido evaluado mediante simulaciones de ordenador considerando canales con distribución Rayleigh y distribución Rice y en escenarios de interior realistas empleando una plataforma MIMO. Esta plataforma MIMO, configurada como un sistema con 2 antenas transmisoras y 2 receptora, ha sido diseñada para operar en la banda de 2.4 GHz ISM con el objetivo de hacer pruebas y rápidos prototipados de módulos MIMO de banda base. La puesta en marcha de la plataforma consiste en el procesamiento off-line de la señales en el transmisor y el receptor mientras que los datos son enviados y adquiridos en tiempo real. Esta propiedad proporciona en la fase de generación de señal, que la modulación y codificación espacio-temporal puedan llevarse a cabo mediante Matlab. En el receptor, el flujo de datos adquirido también es procesado en Matlab: tiempo y frecuencia, sincronización, estimación de canal, decodificación espacio-temporal y, finalmente, detección símbolo a

símbolo son bloques de operación fundamentales.

Publicaciones relacionadas:

- D. Ramírez, I. Santamaría, J. Pérez, J. Vía (Dpt. de Ingeniería de Comunicaciones, Universidad de Cantabria), J. A. García-Naya, T. M. Fernández-Caramés, H. J. Pérez-Iglesias, M. González-López, L. Castedo (Dpt. de Electrónica y Sistemas, Universidade da Coruña), J. M. Torres-Royo (IMS System Engineering, Motorola Inc.), "A Comparative Study of STBC Transmissions at 2.4 GHz over Indoor Channels Using a 2 x 2 MIMO Testbed", Wireless Communications and Mobile Computing, John Wiley and Sons, vol. 8, no. 9, pp. 1449–1164, November 2008.

- H. J. Pérez-Iglesias, J. A. García-Naya, A. Dapena, L. Castedo (Dpt. de Electrónica y Sistemas, Universidade da Coruña), V. Zarzoso (Laboratoire d'Informatique, Signaux et Systmes de Sophia Antipolis), "Blind Channel Identification in Alamouti Coded Systems: A Comparative Study of Eigendecomposition Methods in Indoor Transmissions at 2.4 GHz", in European Transactions on Telecomunications. Special Issue: European Wireless 2007, vol. 19, Issue 7, pp. 751–759, November 2008.

- A. Dapena, H. J. Pérez-Iglesias (Dpt. de Electrónica y Sistemas, Universidade da Coruña), V. Zarzoso (Laboratoire d'Informatique, Signaux et Systmes de Sophia Antipolis), "Blind Channel Estimation Based on Maximizing the Eigenvalue Spread of Cumulant Matrices in (2 x 1) Alamouti's Coding Schemes", accepted in Wireless Communications and Mobile Computing, John Wiley and Sons.

- D. Ramírez, I. Santamaría, J. Pérez, J. Vía, A. Tazón, J. A. García-Naya, T. M. Fernández-Caramés, M. González López, H. J. Pérez-Iglesias, L. Castedo, "A Flexible Testbed for the Rapid Prototyping of MIMO Baseband Modules", in ISWCS 2006, Valencia, Spain, September 2006.

- H. J. Pérez-Iglesias, A. Dapena, J. A. García-Naya, "Using BSS Algorithms in Alamouti Space-Time Block Coding Schemes", 2006 ICA Research Network International Workshop, Liverpool, United Kingdom, September 2006.

- H. J. Pérez-Iglesias, A. Dapena, L. Castedo, V. Zarzoso, "Blind Channel Identification for Alamouti's coding System Based on Eigenvector decomposition", 13th European Wireless Conference, Paris, France, April 2007.

- J. A. García-Naya, T. Fernández-Caramés, H. Pérez-Iglesias, M. González-López, L. Castedo, D. Ramírez, I. Santamaría, J. Pérez-Arriaga, J. Vía, J. M. Torres-Royo, "Performance of STBC Transmissions with Real Data", 16th IST Mobile and Wireless Communications Summit, Budapest, Hungary, July 2007.

- H. J. Pérez-Iglesias, A. Dapena, "Channel Estimation for O-STBC MISO Systems Using Fourth-Order Cross-Cumulants", ICA 2007, London, England, September 2007.

- J. A. García Naya, T. M. Fernández Caramés, H. J. Pérez Iglesias, M. González López, L. Castedo, "A Flexible MIMO Testbed Developed at the University of A Coruña", 1st Workshop on CMCS 2007, Duisburg, Germany, September 2007.

- H. J. Pérez-Iglesias, J. A. García-Naya, A. Dapena, "A Blind Channel Estimation Strategy for the 2x1 Alamouti System Based on Diagonalising 4th-order Cumulant Matrices", ICASSP 2008, Las Vegas (Nevada), U.S.A. March 2008.

- J. A. García-Naya, H. J. Pérez-Iglesias, A. Dapena, L. Castedo, "A Comparative Sutdy of Blind Channel Identification Methods for Alamouti Coded Systems over Indoor Transmissions at 2.4 GHz", SAM 2008, Darmstadt, Germany, July 2008.

- J. A. García-Naya, H. J. Pérez-Iglesias, T. M. Fernández-Caramés, M. González-López, L. Castedo, "A Distributed Multilayer Architecture Enabling End-User Access to MIMO Testbeds", PIMRC 2008, Cannes, France, September 2008.

- H. J. Pérez-Iglesias, D. Iglesia, A. Dapena, V. Zarzoso, "Blind Channel Identification in (2x1) Alamouti Coded Systems Based on Maximizing the Eigenvalue Spread of Cumulant Matrices", ICA 2009, Paraty, Brazil, March 2009.

"Siempre que te equivoques

probablemente será por uno"

Contents

1 Introduction **1**

1.1 Space-Time Diversity 2

1.2 Blind Channel Estimation 6

1.3 Work Overview 8

2 Orthogonal Space-Time Block Codes **11**

2.1 The Alamouti Code 12

2.2 Maximum Likelihood Decoder 16

2.3 Supervised Decoder 18

2.4 Blind Source Separation 19

2.4.1 Adaptive Algorithms 21

2.4.2 Batch Algorithms 30

2.5 Comparison between Classical and BSS Estimation Algorithms 32

3 SOS-based Approaches **35**

3.1 Real-valued Representation of Alamouti Code 37

3.2 The Shahbazpanahi et al. Approach 39
3.2.1 Simulation Results 43
3.3 The Via et al. Adaptive Implementation 45
3.3.1 Complex-valued Implementation 47
3.3.2 Simulation Results 52
3.4 Choosing the Optimal Precoding Matrix 55

4 HOS-based Approaches 61

4.1 The Beres and Adve Approach 63
4.1.1 Simulation Results 63
4.2 The Beres and Adve Approach Improvement I: Linear Combination of Cumulants 65
4.2.1 Simulation Results 68
4.3 The Beres and Adve Approach Improvement II: Cross-slot Cumulants . . . 69
4.3.1 Simulation Results 70
4.4 Maximum Eigenvalue Spread Approach 70
4.4.1 Maximum Eigenvalue Spread Suboptimal Approach 75
4.4.2 Comparison Among Eigenvalue Spreads 76
4.4.3 Blind Estimation of Parameter β 78
4.4.4 Simulation Results 79
4.5 Joint-Diagonalization Approach 83
4.5.1 Simulation Results 86

5 Performance Comparison **87**

5.1 Symbol Error Rate Comparison 88

5.2 Computational Cost 91

6 MIMO Testbed **95**

6.1 Description of the MIMO Testbed 96

6.2 Experiment Description 98

6.2.1 Channel Characterization 101

6.2.2 SNR Estimation 105

6.3 Experimental Results 106

6.3.1 Scenario 1: LOS 107

6.3.2 Scenario 2: NLOS 109

7 Conclusions and Future Work **113**

7.1 Related Publications 116

7.1.1 Journals 116

7.1.2 International Conferences 117

7.2 Further work 119

A Capacity of MIMO Channels **123**

A.1 Capacity of (2×1) Alamouti Code 124

A.2 Capacity of (2×2) Alamouti Code 125

B Computation of a Whitening Matrix **129**

C EVD Computation **131**

C.1 EVD of a 2×2 Matrix 132

C.2 EVD of a 3×3 Matrix 134

D Demonstration: Equations of Shahbazpanahi et al. Approach **137**

D.1 The Kronecker Product 137

D.2 Demonstration of Equation (3.29) 139

D.3 Demonstration of Equation (3.35) 140

D.4 Demonstration of Equation (3.38) 142

E HOS Matrices for (2×1) Alamouti Code **145**

E.1 Higher Order Statistics (HOS) definition 145

E.2 HOS Properties 146

E.3 4^{th} Order Cumulants for 2 Variables 147

E.4 HOS Matrices for (2×1) Alamouti Code 149

F Joint Diagonalization Procedure **153**

List of Figures

1.1 Scheme of a digital communication system 3

1.2 Block diagram of a STC 5

2.1 The (2×1) Alamouti code 13

2.2 The (2×2) Alamouti code 14

2.3 Adaptive BSS algorithms 21

2.4 Separation system employed by Herault and Jutten 22

2.5 Comparison between supervised and blind algorithms for Rayleigh channel distribution: SER versus SNR 33

2.6 Comparison between supervised and blind algorithms for Rice channel distribution: SER versus SNR 34

3.1 Performance of the Shahbazpanahi et al. SOS algorithm employing different precoding matrices with Rayleigh channel distribution: SER versus SNR 44

3.2 Performance of the Shahbazpanahi et al. SOS Algorithm employing different precoding matrices with Rice channel distribution: SER versus SNR 44

3.3 Performance of the Via et al. adaptive SOS algorithm employing different precoding matrices with Rayleigh channel distribution: SER versus SNR 54

3.4 Performance of the Via et al. adaptive SOS algorithm employing different precoding matrices with Rice channel distribution: SER versus SNR 54

3.5 Finding the optimum γ^2 parameter for Rayleigh channel distribution: SER versus SNR 56

3.6 Finding the optimum γ^2 parameter for Rice channel distribution: SER versus SNR 57

3.7 Performance of the general SOS algorithm employing the optimal encoding matrix with Rayleigh channel distribution: SER versus SNR 58

3.8 Performance of the general SOS algorithm employing the optimal encoding matrix with Rice channel distribution: SER versus SNR 58

3.9 Comparison between SOS algorithms for Rayleigh channel distribution: SER versus SNR 59

3.10 Comparison between SOS algorithms for Rice channel distribution: SER versus SNR 60

4.1 Performance of the Beres and Adve HOS algorithm employing $\mathbf{C}_{\text{HOS}}^{[1,1]}$ (or $\mathbf{C}_{\text{HOS}}^{[2,2]}$) with Rayleigh channel distribution: SER versus SNR 64

4.2 Performance of the Beres and Adve HOS algorithm employing $\mathbf{C}_{\text{HOS}}^{[1,1]}$ (or $\mathbf{C}_{\text{HOS}}^{[2,2]}$) with Rice channel distribution: SER versus SNR 64

4.3 Finding the optimum λ parameter for Rayleigh channel distribution: SER versus SNR 67

4.4 Finding the optimum λ parameter for Rice channel distribution: SER versus SNR 67

4.5 Performance of the Beres and Adve improvement I HOS algorithm with matrix $\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$ for Rayleigh channel distribution: SER versus SNR 68

4.6 Performance of the Beres and Adve improvement I HOS algorithm with Matrix $\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$ for Rice channel distribution: SER versus SNR 69

4.7 Performance of the Beres and Adve improvement II HOS algorithm with matrix $\mathbf{C}_{\text{HOS}}^{[1,2]}$ for Rayleigh channel distribution: SER versus SNR 71

4.8 Performance of the Beres and Adve improvement II HOS algorithm with matrix $\mathbf{C}_{\text{HOS}}^{[1,2]}$ for Rice channel distribution: SER versus SNR 71

4.9 Cumulative probability distribution for different matrices $\mathbf{M}$: $L(\mathbf{M}_1)$, $L(\mathbf{M}_2)$, $\max(L(\mathbf{M}_1), L(\mathbf{M}_2))$ and $L(\mathbf{M}_{\text{opt}})$ with Rayleigh channel distribution 76

4.10 Cumulative probability distribution for different matrices $\mathbf{M}$: $L(\mathbf{M}_1)$, $L(\mathbf{M}_2)$, $\max(L(\mathbf{M}_1), L(\mathbf{M}_2))$ and $L(\mathbf{M}_{\text{opt}})$ with Rice channel distribution 77

4.11 Normalized error between the theoretical and the estimated value of the 4^{th} order cumulants $\text{cum}(x_1, x_1^*, x_1, x_2^*)$, $\text{cum}(x_1, x_2^*, x_1, x_2^*)$ and $\text{cum}(x_1, x_1^*, x_2, x_2^*)$. 81

4.12 Normalized error in the estimation of $|\beta|$ using $|\beta| = \frac{|\text{cum}(x_1, x_2^*, x_1, x_2^*)|}{|\text{cum}(x_1, x_1^*, x_1, x_2^*)|}$ and $|\beta| = \frac{\text{cum}(x_1, x_1^*, x_2, x_2^*)}{|\text{cum}(x_1, x_1^*, x_1, x_2^*)|}$. 81

4.13 Performance of the eigenvalue spread HOS algorithms for Rayleigh channel distribution: SER versus SNR 82

4.14 Performance of the eigenvalue spread HOS algorithms for Rice channel distribution: SER versus SNR 83

4.15 Performance of Joint-Diagonalization HOS algorithm for Rayleigh channel distribution: SER versus SNR 85

4.16 Performance of Joint-Diagonalization HOS algorithm for Rice channel distribution: SER versus SNR 85

5.1 Performance comparison of algorithms for Rayleigh channel distribution: SER versus SNR 89

5.2 Performance comparison of algorithms for Rice channel distribution: SER versus SNR 89

5.3 Performance comparison of algorithms for Rayleigh channels: SER versus block size 90

5.4 Performance comparison of algorithms for Rice channels: SER versus block size 91

6.1 MIMO testbed 96

6.2 MIMO testbed picture 97

6.3 Experiment locations 98

6.4 (2×1) Alamouti complete scheme 99

6.5 Scenario 1: LOS - probability density function 101

6.6 Scenario 1: LOS - cumulative distribution function 102

6.7 Scenario 2: NLOS - probability density function 103

6.8 Scenario 2: NLOS - cumulative distribution function 103

6.9 SNR estimation 105

6.10 Scenario LOS: performance for the SOS-based approach according to the unbalance power factor γ^2. 107

6.11 Scenario 1: LOS - performance results measuring SER according to the SNR . . 108

6.12 Scenario 1: LOS - performance results measuring SER according to block size for an SNR of 10 dB 109

6.13 Scenario NLOS: performance for the SOS-based approach according to the unbalance power factor γ^2. 110

6.14 Scenario 2: NLOS - performance results measuring SER according to the SNR . 110

6.15 Scenario 2: NLOS - performance results measuring SER according to block size for an SNR of 10 dB 111

List of Tables

2.1 FastICA algorithm 29

2.2 JADE algorithm 32

3.1 General SOS algorithm: PCA 36

3.2 The Shahbazpanahi et al. SOS algorithm 43

3.3 The Via et al. adaptive SOS algorithm 48

3.4 The complex-valued adaptive SOS algorithm 53

4.1 General HOS Algorithm: ICA 62

4.2 *Maximum Eigenvalue Spread Optimal Approach* (MESOA) 79

4.3 *Maximum Eigenvalue Spread Suboptimal Approach* (MESSA) 80

4.4 Joint-Diagonalization HOS algorithm 84

5.1 Computational load corresponding to the computation of 2^{nd} and 4^{th} order cumulants and matrix diagonalization 92

5.2 Computational load of the approaches studied 92

5.3 Ranking classification in terms of performance and computational cost . . 93

6.1 χ^2 Pearson goodness-of-fit test - scenario 1: LOS 104

6.2 χ^2 Pearson goodness-of-fit test - scenario 2: NLOS 104

Chapter 1

Introduction

This work focuses on the problem of estimating the channel coefficients in an Alamouti coded system [1]. To this end, the received signals (observations) will be modeled as instantaneous mixtures of the transmitted signals (sources). This interpretation allows us to estimate the channel by using *Blind Source Separation* (BSS) algorithms [2]. More specifically, we will investigate those algorithms based on computing the *Eigenvalue Decomposition* (EVD) of matrices computed from the observation statistics.

EVD-based channel estimation algorithms can be classified into SOS-based and HOS-based methods depending on the information used to compute the matrix to be diagonalized. SOS-based approaches compute the correlation matrix of the observations which requires less computer operation, but they need to include a linear precoder before the Alamouti encoder to unbalance the source power or to color the sources [3, 4]. On the contrary, HOS-based methods are based on diagonalizing one or several matrices containing higher order cumulants. The best-known method has been proposed in the context of BSS, although not specifically for the Alamouti code, by Cardoso et al [5] with the name of *Joint Approximate Diagonalization of Eigenmatrices* (JADE). More recently, Beres et al [6] have particularized JADE for Alamouti coded systems. Following these ideas, in this work we will present novel HOS-based methods which present a good performance in

different environments with a considerably low computational load.

The performance of the blind channel estimation techniques will be evaluated both in computer simulations carried out assuming Rayleigh and Rice distributed channels, and realistic indoor scenarios using a *Multiple Input Multiple Output* (MIMO) [7] hardware testbed. This MIMO testbed, configured as a *Two Input Single Output* (TISO) system, has been designed to operate at the 2.4 GHz *Industrial, Scientific and Medical* (ISM) band, and is intended for the testing and rapid prototyping of MIMO baseband modules. The testbed operation consists of performing the signal processing off-line both at the transmitter and at the receiver while the data are sent and acquired in real-time. This property enables, at the signal generation stage, modulation and space-time coding operations to be carried out off-line using `MATLAB`. At the receiver, the acquired data stream is also processed in `MATLAB`: time and frequency synchronization, channel estimation, space-time decoding and, finally, symbol-by-symbol detection are the fundamental operational blocks.

This work is based on two important issues. Section 1.1 focuses on the concept of Space-Time Diversity and Section 1.2 motivates the importance of using blind channel estimation methods. Finally, Section 1.3 presents an overview of this work.

1.1 Space-Time Diversity

Information theory investigations have demonstrated that the capacity of wireless channels can be considerably increased, at no extra bandwidth or power consumption, if the multipath is sufficiently rich and properly exploited by means of multi-element antennas at both transmission and reception points [7, 8, 9, 10, 11]. The basic idea is to interpret the transmission through multipath channels from a new perspective in which multipath signal propagation is no longer viewed as an impairment. Instead of this standpoint, the multipath is considered as a phenomenon that provides spatial diversity and can be successfully exploited to improve reception. To this end, the signals on the transmit antennas

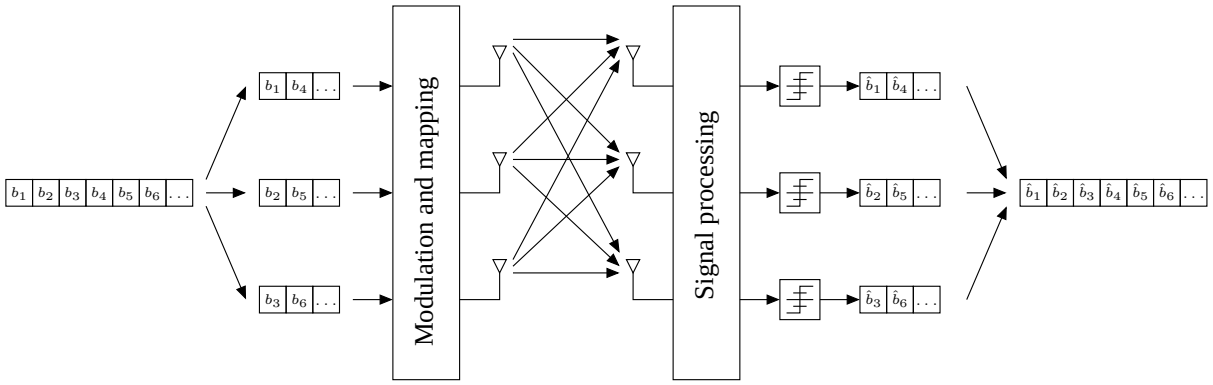


Figure 1.1: Scheme of a digital communication system

at one end and the receive antennas at the other end are "combined" in such way that the quality, *Bit Error Rate* (BER), or the data rate for each user will be improved [12, 13].

Figure 1.1 shows the scheme of a digital communication system. A high-rate bit stream is decomposed into several bit sequences which are then transmitted simultaneously using multiple antennas. At the receiver side, after identifying the channel, the individual bit streams are separated and estimated. This kind of system can be expressed using linear equations. In fact, the received signals (observations), $\mathbf{x}$, are a linear combination of the transmitted signals (sources), $\mathbf{s}$, i.e., the signal model obeys the following equation

$$\mathbf{x} = \mathcal{H}\mathbf{s} \tag{1.1}$$

where $\mathcal{H}$ contains the channel coefficients (the input h_{ij} denotes the path from the i-th transmit antenna to the j-th receive antenna). Depending on the number of elements in transmission and reception, communication systems can be classified into four groups: *Single Input Single Output* (SISO), *Multiple Input Single Output* (MISO), *Single Input Multiple Output* (SIMO) and MIMO.

A SISO system refers to configurations that employ one antenna in transmission and one antenna in reception. The capacity of this system is

$$C(\mathcal{H}) = \log_2(1 + \frac{\sigma_s^2}{\sigma_v^2}|h|^2) \tag{1.2}$$

where h is the normalized channel coefficients and $\frac{\sigma_s^2}{\sigma_v^2}$ is the *Signal to Noise Ratio* (SNR)

at the receive antenna, i.e. the quotient between the signal power and noise power.

The capacity can be improved by using diversity on reception obtained by including n_R receive antennas, such that we have a SIMO system with

$$C(\mathcal{H}) = \log_2(1 + \frac{\sigma_s^2}{\sigma_v^2} \sum_{i=1}^{n_R} |h_i|^2) \tag{1.3}$$

Note that the fact of increasing the value of n_R only results in a logarithmic increase in the average capacity. On the contrary, if we opt for transmission diversity by using a MISO system which employs n_T antennas in transmission and one antenna in reception, the capacity is given by

$$C(\mathcal{H}) = \log_2(1 + \frac{\sigma_s^2}{\sigma_v^2} \frac{1}{n_T} \sum_{i=1}^{n_T} |h_i|^2) \tag{1.4}$$

where the normalization of the SNR by n_T ensures a fixed total transmitter power. Again the capacity has a logarithmic relationship with n_T.

Consider now a MIMO system which uses both transmission and reception diversity. For the case of n_T transmit antennas and n_R received ones, we have the famous capacity equation [14]

$$C(\mathcal{H}) = \log_2 \left(\det(\mathbf{I}_{n_R} + \frac{\sigma_s^2}{\sigma_v^2} \frac{1}{n_T} \mathcal{H}\mathcal{H}^{\mathrm{H}}) \right) \tag{1.5}$$

where $(\cdot)^{\mathrm{H}}$ denotes the hermitian operator. Foschini [10] and Telatar [8] have demonstrated that the capacity of a MIMO system grows linearly with $m = \min(n_R, n_T)$ rather than with the logarithm function.

The performance of systems with several antennas in transmission or reception can be substantially improved by including codes specifically designed to take both the spatial and temporal dimensions [15] into account. These techniques are collectively known as *Space-Time Coding* (STC) [12] and obey the scheme shown in Figure 1.2. In this scheme, a number of code symbols equal to the number of transmit antennas are generated and transmitted simultaneously, one symbol from each antenna. These symbols are generated

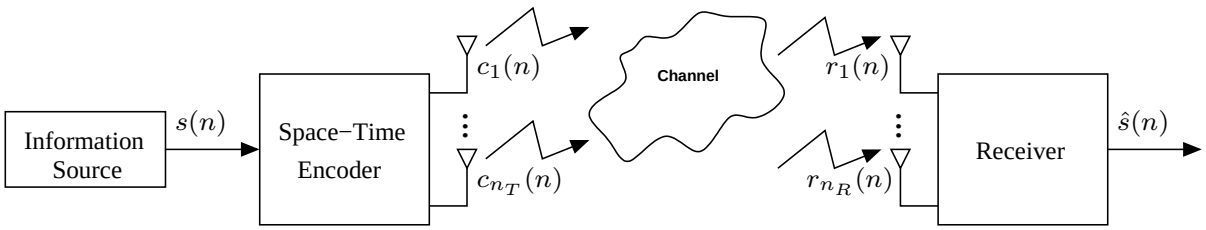


Figure 1.2: Block diagram of a STC

by the space-time encoder such that by using the appropriate signal processing and decoding algorithm at the receiver, the diversity gain and/or the coding gain is maximized.

The key development of the STC concept was originally revealed in [15] in the form of *Space-Time Trellis Codes* (STTC), which require a multidimensional Viterbi algorithm at the receiver for decoding. These codes provide, without any loss in bandwith efficiency, a diversity gain equal to the number of transmit antennas in addition to a coding gain that depends on the number of code complexity (i.e. trellis states). The popularity of STC was due to the discovery of the so-called *Space-Time Block Codes* (STBC) that can be decoded using simple linear processing at the receiver instead of a Viterbi trellis. STBC provides the same diversity gain as STTC but the coding benefit is zero or minimal.

In addressing the issue of decoding complexity, Alamouti [1] discovered a remarkable STBC considering two antennas at transmission. This scheme represents the first example of the *Orthogonal Space-Time Block Coding* (OSTBC) [16, 17]. The basic premise of OSTBC is the encoding of the transmitted symbols into an orthogonal matrix which reduces the optimum *Maximum Likelihood* (ML) decoder to a matrix matched filter followed by a symbol-by-symbol detector. Other OSTBC have been proposed for more than two transmit antennas but they suffer from severe spatial rate loss [17, 18]. The Alamouti code can be used in systems with one or multiple antennas at reception. At first glance, it seems that using several receive antennas is beneficial because this increases the diversity gain and provides array gain. However, the signal structure imposed by the Alamouti code reduces the constrained channel capacity limit when there is more than one receive an-

tenna [19]. This means that when used in concatenation with capacity approaching codes (such as turbo or LDPC codes), Alamouti systems with more than one receive antenna exhibit a degradation in performance. Thus, the (2×1) Alamouti coded systems are most attractive in wireless communications due to their simplicity and their ability to provide maximum diversity gain while preserving channel capacity. Because of these advantages, the Alamouti code has been incorporated in the IEEE 802.11 and IEEE 802.16 standards [20].

In this work we will study and propose methods considering the (2×1) Alamouti code, some of which can be generalized for other kind of OSTBC.

1.2 Blind Channel Estimation

Coherent detection in the (2×1) Alamouti coded systems requires the identification of a 2×2 unitary channel matrix. The standard way of estimating this channel matrix is through the transmission of pilot symbols, also referred to as training sequences. However, the inclusion of pilot symbols reduces the system throughput (equivalently, it also reduces the system spectral efficiency) and wastes transmission energy because training sequences do not convey information. Strategies that avoid this limitation include the so-called Differential STBC (DSTBC) [21], which is a signalling technique that generalizes differential modulations to the transmission over MIMO channels. DSTBC can be incoherently decoded without the aid of channel estimates but they incur a 3 dB performance penalty when compared to coherent detection.

Training sequences can also be avoided by the use of BSS techniques to identify the channel, which is the main theme of discussion in this work. The transmitted symbol substreams can be considered as unknown sources to be recovered from their mixtures observed at the receive antenna output, whereas the channel matrix can be seen as the mixing transformation between the sources and the observations [22, 23, 24, 25, 26]. The term *blind* (or unsupervised) refers to the fact that little or nothing is known or assumed

about the sources and the mixing matrix structure in a general BSS scenario. Under the assumption of statistical independence between the transmitted symbol substreams, *Independent Component Analysis* (ICA) techniques can be used to tackle this problem. Hence, many existing ICA algorithms (e.g., [22, 23, 24, 25]) would be able to identify the channel matrix and recover the transmitted symbols. However, in order to reduce the computational load, specific algorithms taking advantage of the special structure of these codes can be designed instead [4, 6, 27, 28].

A property commonly exploited in BSS is the statistical independence of the sources. Depending on the degree of independence considered, two main groups of techniques can be distinguished: *Principal Component Analysis* (PCA), which are based on *Second Order Statistics* (SOS), and ICA, which exploits *Higher Order Statistics* (HOS). A number of PCA and ICA approaches rely on the eigendecomposition of certain matrix or tensor structures.

PCA approaches are based on diagonalizing the correlation matrix of the observed signals. However, it is well known that this operation can be done only when the associated eigenvalues are different [29]. In order to guarantee this condition, several authors have proposed the use of a linear precoder before the Alamouti encoder with the aim of unbalancing the source power or coloring the sources [3, 4]. In this work, we will show that the global performance is degraded when the power source is unbalanced because, although the mean probability is adequate, the error probability of the sources with lower power is excessively high for some real applications. Another way of guarantying the identifiable condition consists of transmitting an odd number of real symbols at each block [27]. This approach, however, produces a loss in the transmission rate because some symbols must be ruled out.

The higher order independence of the source signals is exploited by the ICA approach. Independence is typically measured by means of HOS such as the higher order cumulants: the absolute value of the marginal cumulants is to be maximized or, equivalently, that of the cross-cumulants minimized, subject to the appropriate constraints. In Comon's

pioneering ICA contribution [23], the initial source estimations provided by PCA are further processed via Givens rotations aiming at maximizing the 4^{th} order independence of the transformed signals. The optimal rotation angles are obtained by rooting a low-degree polynomial whose coefficients are computed from the 4^{th} order cumulants of the signal pair. Several sweeps over all signal pairs are necessary for convergence. This pairwise scheme can be seen as the generalization to 4^{th} order cumulant tensors (higher order arrays) of the well-known Jacobi technique for matrix diagonalization. Research into higher order eigen-based approaches began with Cardoso's early work on the so-called quadricovariance, a folded version of the 4^{th} order moment array, and culminated in the popular JADE method [5].

The orthogonality property imposed in OSTBC makes it possible to propose HOS approaches consisting of computing the eigenvalue decomposition (EVD) of matrices containing 4^{th} order statistics of the observations [6, 28]. These algorithms can be considered as particular cases of JADE but present a reduced computational cost. In this work, we will study this kind of BSS algorithms in detail.

1.3 Work Overview

This work is organized as follows.

Chapter 2 focuses on *Space-Time Block Codes* (STBC). We present the signal model of the (2×1) Alamouti system that in other OSTBC takes advantage of their orthogonality to allow an easy, linear and optimal decoding at the receiver. In this context we explain classical decoding strategies and well-known BSS algorithms.

Chapter 3 presents SOS-based channel estimation algorithms based on finding the eigenvector and eigenvalues of the correlation matrix obtained from the observed signals. In this sense, we make a review of the methods proposed by Shahbazpanahi et al. [3] and Via et al. [4] considering the general real-valued model of an STBC. In addition, we

present a new formulation of these methods for the (2×1) Alamouti system considering a complex-valued signal model. An important contribution of this chapter is to determine empirically the form of the matrix that can be used to code the original signals.

Chapter 4 is devoted to present HOS-based BSS approaches which diagonalize a linear combination of a sensible choice from a set of different 4^{th} order cumulant matrices. We begin with the method proposed by Beres et al. [6] and then go on to propose a set of novel methods. This chapter contains the most important contribution of this work, that is to propose a closed form to obtain the optimal cumulant matrix.

Chapter 5 shows a comparison of performance for the different approaches through a set of computer simulations.

In Chapter 6 a MIMO Testbed developed at the Universidade da Coruña is shown and employed to test the approaches shown in the previous chapters.

Finally, Chapter 7 is devoted to the conclusions and future work.

Chapter 2

Orthogonal Space-Time Block Codes

Space-Time Codes (STC) make it possible to improve the performance of data transmission in wireless communications that employ multiple transmit and/or receive antennas [12]. STC consist of transmitting multiple, redundant copies of a data stream across a number of antennas at the transmitter to the multiple antennas at the receiver with the aim of overcoming the impairments of the physical paths between transmission and reception. The transmitted signal must cross a potentially difficult environment with scattering, reflection, refraction and thermal noise presence in the receiver, which means that some of the received copies of the data will be better than others. In fact, STC combine all the copies of the received signals in an optimal way to extract as much information from each of them as possible.

STC can be classified into two main groups according to their function:

- *Space-Time Trellis Codes* (STTC) distribute a trellis code over multiple antennas and multiple time slots providing both coding gain and diversity gain.

- *Space-Time Block Codes* (STBC) act on a block of data simultaneously (similarly to block codes). Unlike STTC, the STBC only provide diversity gain, but their implementation complexity is less.

A particular and interesting case of STBC is *Orthogonal Space-Time Block Codes* (OSTBC), which takes advantage of their orthogonality to allow an easy, linear and optimal decoding at the receiver. Their most important disadvantage is that the redundancy imposed by the orthogonality can punish the data rate and, as a consequence, they are unable to attain the full capacity of the MIMO channel.

This chapter is structured as follows. Section 2.1 describes the Alamouti code, which is the simplest and the pioneer OSTBC. Section 2.2 presents the *Maximum Likelihood* (ML) criterion. Subsequently, we present two different strategies to estimate the channel matrix: Supervised methods in Section 2.3 and *Blind Source Separation* (BSS) techniques in Section 2.4. Finally, Section 2.5 compares the performance of classical and BSS algorithms to estimate the channel matrix for the Alamouti coded scheme.

2.1 The Alamouti Code

In 1998 [1], Alamouti designed a code considering 2 transmit antennas and 1 or 2 receive antennas. Although he did not make any reference to the term OSTBC, this method is considered today the first example of this kind of code. For the case that employs 1 receive antenna, this is the only OSTBC that achieves full diversity for complex constellations.

In order to explain the Alamouti code, we define the sources $s_1(n)$ and $s_2(n)$ as independent equiprobable discrete random variables, where $n \in \mathbb{N}$ is the independent variable, the time instant. Henceforth, we prefer to employ the notation s_1 and s_2, without the independent variable for reasons of simplicity. The signals take values from a finite set of symbols belonging to a real or complex modulation (PAM, PSK, QAM...). Hence, we will employ the operator $(\cdot)^*$ to denote the conjugate of a symbol.

The (2×1) Alamouti code is shown in Figure 2.1. In this scheme, each pair of symbols $\{s_1, s_2\}$ is transmitted in two adjacent time slots using a simple strategy: in the first time slot s_1 and s_2 are transmitted from the first and the second antenna, respectively, and

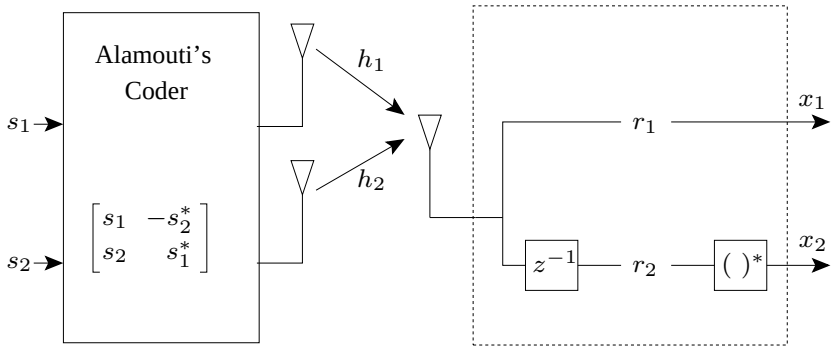


Figure 2.1: The (2×1) Alamouti code

in the second time slot $-s_2^*$ is transmitted from the first antenna and s_1^* from the second one. The transmitted symbols arrive at the receive antenna through the fading paths h_1 and h_2 and the signal is perturbed by an additive noise v_1, i.e, the received value in the first time slot has the form

$$r_1 = h_1 s_1 + h_2 s_2 + v_1 \tag{2.1}$$

where h_i denotes the path from the i-th transmit antenna to the receive one. If the channel remains constant during two time slots, the received value in the second time slot is given by

$$r_2 = h_2 s_1^* - h_1 s_2^* + v_2^* \tag{2.2}$$

Thus, from equations (2.1) and (2.2) we obtain the received vector $\mathbf{r} = [r_1 \ r_2]^{\mathrm{T}}$. The (2×1) Alamouti code can be also expressed in matrix form as follows

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} h_1 s_1 + h_2 s_2 \\ h_2 s_1^* - h_1 s_2^* \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2^* \end{bmatrix} \tag{2.3}$$

where v_1 and v_2 are two independent random variables following the distribution $N(0, \sigma_v)$, i.e. additive white Gaussian noise.

A more convenient form of writing this coding strategy consists of considering the source vector $\mathbf{s} = [s_1 \ s_2]^{\mathrm{T}}$, the noise vector $\mathbf{v} = [v_1 \ v_2]^{\mathrm{T}}$, and the observation vector

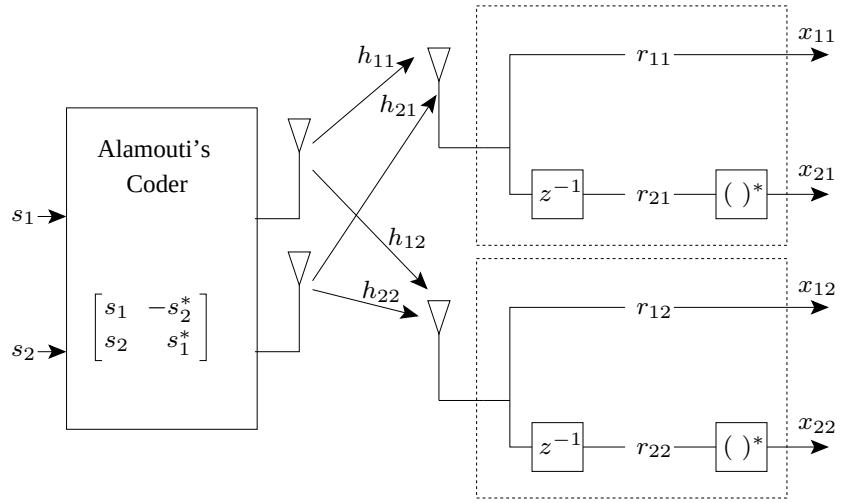


Figure 2.2: The (2×2) Alamouti code

$\mathbf{x} = [x_1 \; x_2]^{\mathrm{T}} = [r_1 \; r_2^*]^{\mathrm{T}}$. Considering the 2×2 channel matrix

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \tag{2.4}$$

the relationship between observations and sources is given by

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v} \iff \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{2.5}$$

It is important to point out that the (2×1) Alamouti code is the unique STBC that achieves the maximum capacity of the equivalent MIMO channel. In Appendix A, we demonstrate that the capacity is given by

$$C(\mathbf{H}) = \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2} (|h_1|^2 + |h_2|^2) \right) \tag{2.6}$$

where σ_s^2 is the variance of the sources s_1 and s_2, and σ_v^2 is the variance of the noise.

Alamouti [1] has also proposed the code for 2 transmit antennas and 2 receive antennas, as shown in Figure 2.2. In this case, we make the same transmission as in the (2×1) system but the 2 signals are received simultaneously (1 at each receive antenna). Thus, the source vector is the same $\mathbf{s} = [s_1 \; s_2]^{\mathrm{T}}$ and there are 4 fading paths: h_{11}, h_{21}, h_{12} and

h_{22} where h_{ij} denotes the path from i-th transmit antenna to the j-th receive antenna. Since there are 2 receive antennas and 2 time slots, we have 4 observations. The noise vector is $\mathbf{v} = [v_{11}\ v_{21}\ v_{12}\ v_{22}]^{\mathrm{T}}$ and the observation vector is $\mathbf{x} = [x_{11}\ x_{21}\ x_{12}\ x_{22}]^{\mathrm{T}}$. The relationship between these terms is given by

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v} \iff \begin{bmatrix} x_{11} \\ x_{21} \\ x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{21} \\ h_{21}^* & -h_{11}^* \\ h_{12} & h_{22} \\ h_{22}^* & -h_{12}^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} v_{11} \\ v_{21} \\ v_{12} \\ v_{22} \end{bmatrix} \tag{2.7}$$

Denoting by k the number of symbols that the encoder takes as its input at each operation, and p the number of transmission time slots required to transmit the *Space-Time Coded* symbols through the multiple transmit antennas, the relationship between these two terms is known as rate, i.e.

$$R = \frac{k}{p} \tag{2.8}$$

In particular, the rate is $R = k/p = 2/2 = 1$ for the Alamouti code with 1 and 2 receive antennas. In Appendix A we show that the capacity of the (2×2) Alamouti code is

$$C(\mathbf{H}) = \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2} (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2) \right) \tag{2.9}$$

This value is lower than the maximum capacity of a (2×2) MIMO system.

Tarokh et al. [17, 30] coined the concept of STBC and developed a theory to design OSTBC which support an easy, linear and optimal decoding at the receiver. For any number of transmit antennas, these codes achieve the maximum possible transmission rate when the symbols correspond to any arbitrary real constellation and 1/2 for complex constellations. For the specific case of 2, 3 or 4 transmit antennas, it is possible to achieve, respectively, full, 3/4 and 3/4 of the maximum transmission using arbitrary complex constellation.

2.2 Maximum Likelihood Decoder

Maximum Likelihood (ML) is a statistical method for estimating parameters from sample data that selects as estimates those parameter values maximizing the probability of obtaining the observed data. For the case of the (2×1) Alamouti code, assuming that the sources take values from a complex modulation with an equiprobable distribution, an ML decoder [1, 17] chooses a pair of parameters (s_i, s_j) from the modulation constellation C in order to minimize the Euclidean distance. The rule is to take the pair $(s_i, s_j) \in C \times C$ that satisfies

$$\begin{aligned} &d^2(r_1, h_1 s_i + h_2 s_j) + d^2(r_2, h_2 s_i^* - h_1 s_j^*) \leq \\ &\quad d^2(r_1, h_1 s_k + h_2 s_l) + d^2(r_2, h_2 s_k^* - h_1 s_l^*), \qquad \forall\ (s_k, s_l) \in C \times C \end{aligned} \tag{2.10}$$

where $d^2(a, b)$ is Euclidean distance between two complex-valued numbers. This is calculated by employing the equation

$$d^2(a, b) = |a - b|^2 = (a - b)(a - b)^* = (a - b)(a^* - b^*) \tag{2.11}$$

Hence, we can expand the first term of the equation (2.10) and obtain

$$\begin{aligned} &d^2(r_1, h_1 s_i + h_2 s_j) = (r_1 - h_1 s_i - h_2 s_j)(r_1^* - h_1^* s_i^* - h_2^* s_j^*) = \\ &|r_1|^2 + |h_1|^2 |s_i|^2 + |h_2|^2 |s_j|^2 - h_1 r_1^* s_i - h_1^* r_1 s_i^* - h_2 r_1^* s_j - h_2^* r_1 s_j^* + h_1 h_2^* s_i s_j^* + h_1^* h_2 s_i^* s_j \end{aligned} \tag{2.12}$$

Doing the same for the second term, we have

$$\begin{aligned} &d^2(r_2, h_2 s_i^* - h_1 s_j^*) = (r_2 - h_2 s_i^* + h_1 s_j^*)(r_2^* - h_2^* s_i + h_1^* s_j) = \\ &|r_2|^2 + |h_2|^2 |s_i|^2 + |h_1|^2 |s_j|^2 - h_2^* r_2 s_i - h_2 r_2^* s_i^* + h_1^* r_2 s_j + h_1 r_2^* s_j^* - h_1 h_2^* s_i s_j^* - h_1^* h_2 s_i^* s_j \end{aligned} \tag{2.13}$$

Thus, from the addition of the equations (2.12) and (2.13), we obtain the following expression

$$\begin{aligned} &d^2(r_1, h_1 s_i + h_2 s_j) + d^2(r_2, h_2 s_i^* - h_1 s_j^*) = \\ &|r_1|^2 + |r_2|^2 + (|h_1|^2 + |h_2|^2)(|s_i|^2 + |s_j|^2) + \\ &-(h_1 r_1^* + h_2^* r_2) s_i - (h_1^* r_1 + h_2 r_2^*) s_i^* - (h_2 r_1^* - h_1^* r_2) s_j - (h_2^* r_1 - h_1 r_2^*) s_j^* \end{aligned} \tag{2.14}$$

Hence, if we employ the decision statistics

$$\tilde{s}_1 = h_1^* r_1 + h_2 r_2^* \tag{2.15}$$

$$\tilde{s}_2 = h_2^* r_1 - h_1 r_2^* \tag{2.16}$$

we can express the equation (2.14) as

$$\begin{aligned} & d^2(r_1, h_1 s_i + h_2 s_j) + d^2(r_2, h_2 s_i^* - h_1 s_j^*) = \\ & |r_1|^2 + |r_2|^2 + (|h_1|^2 + |h_2|^2)(|s_i|^2 + |s_j|^2) - \tilde{s}_1^* s_i - \tilde{s}_1 s_i^* - \tilde{s}_2^* s_j - \tilde{s}_2 s_j^* = \\ & (|h_1|^2 + |h_2|^2 - 1)(|s_i|^2 + |s_j|^2) + d^2(\tilde{s}_1, s_i) + d^2(\tilde{s}_2, s_j) + \\ & |r_1|^2 + |r_2|^2 - |\tilde{s}_1|^2 - |\tilde{s}_2|^2 \end{aligned} \tag{2.17}$$

Substituting in both sides of the inequation (2.10), we have the ML decoding rule

$$\begin{aligned} & (|h_1|^2 + |h_2|^2 - 1)(|s_i|^2 + |s_j|^2) + d^2(\tilde{s}_1, s_i) + d^2(\tilde{s}_2, s_j) \leq \\ & (|h_1|^2 + |h_2|^2 - 1)(|s_k|^2 + |s_l|^2) + d^2(\tilde{s}_1, s_k) + d^2(\tilde{s}_2, s_l), \qquad \forall\ (s_k, s_l) \in C \times C \end{aligned} \tag{2.18}$$

In this rule, we distinguish two parts corresponding to s_i and s_j

$$(|h_1|^2 + |h_2|^2 - 1)|s_i|^2 + d^2(\tilde{s}_1, s_i) \leq (|h_1|^2 + |h_2|^2 - 1)|s_k|^2 + d^2(\tilde{s}_1, s_k),\ \forall\ s_k \in C \tag{2.19}$$

$$(|h_1|^2 + |h_2|^2 - 1)|s_j|^2 + d^2(\tilde{s}_2, s_j) \leq (|h_1|^2 + |h_2|^2 - 1)|s_l|^2 + d^2(\tilde{s}_2, s_l),\ \forall\ s_l \in C \tag{2.20}$$

In conclusion, for any constellation C employed in transmission, the ML rule consists of choosing the symbol s_i that satisfies equation (2.19) and the symbol s_j that satisfies equation (2.20). For constant modulus constellations, the term $(|h_1|^2 + |h_2|^2 - 1)|s_i|^2$ is constant and the expressions (2.19) and (2.20) can be simplified to

$$d^2(\tilde{s}_1, s_i) \leq d^2(\tilde{s}_1, s_k),\ \forall\ s_k \in C \tag{2.21}$$

$$d^2(\tilde{s}_2, s_j) \leq d^2(\tilde{s}_2, s_l),\ \forall\ s_l \in C \tag{2.22}$$

Note that in order to employ the statistics needed to apply the ML rule, we need to know the fading paths h_1 and h_2, i.e. the channel realization.

2.3 Supervised Decoder

The Wiener filter [31] can be considered the main supervised decode method in an OSTBC scheme. The idea is to obtain the desired output $s(n)$ from a set of observations $\mathbf{x}(n) = [x(n)\ x(n+1)\ \ldots\ x(n+M-1)]^{\mathrm{T}}$ employing a *Finite Impulse Response* (FIR) filter $\mathbf{w} = [w_1\ w_2\ \ldots\ w_M]^{\mathrm{T}}$. Thus, the obtained output is

$$y(n) = \sum_{i=1}^{M} w_i^* x(n+i-1) = \mathbf{w}^{\mathrm{H}}\mathbf{x}(n) \tag{2.23}$$

And the error between the desired and the estimated signals is

$$e(n) = s(n) - y(n) = s(n) - \mathbf{w}^{\mathrm{H}}\mathbf{x}(n) \tag{2.24}$$

Considering the *Minimum Mean Square Error* (MMSE) criterion, we define the following cost function

$$J(\mathbf{w}) = \mathrm{E}[|e(n)|^2] = \mathrm{E}[e(n)e^*(n)] \tag{2.25}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator. Substituting (2.23) and (2.24) in (2.25), we obtain the following expanded expression

$$\begin{aligned} J(\mathbf{w}) &= \mathrm{E}[|s(n)|^2 - s(n)\mathbf{x}^{\mathrm{H}}(n)\mathbf{w} - s^*(n)\mathbf{w}^{\mathrm{H}}\mathbf{x}(n) + \mathbf{w}^{\mathrm{H}}\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\mathbf{w}] \\ &= \mathrm{E}[|s(n)|^2] + \mathbf{w}^{\mathrm{H}}\mathrm{E}[\mathbf{x}(n)\mathbf{x}(n)^{\mathrm{H}}]\mathbf{w} - \mathrm{E}[s(n)\mathbf{x}^{\mathrm{H}}(n)]\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathrm{E}[s^*(n)\mathbf{x}(n)] \\ &= \sigma_s^2 + \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}}\mathbf{w} - \mathbf{p}^{\mathrm{H}}\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{p} \end{aligned} \tag{2.26}$$

where $\mathbf{R}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}(n)\mathbf{x}(n)^{\mathrm{H}}]$ is the correlation matrix of the observations and $\mathbf{p} = \mathrm{E}[s^*(n)\mathbf{x}(n)]$ is the correlation between the desired signal and the observations. In order to determine the minimum $J(\mathbf{w})$, we will compute the points where the gradient vanishes

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \left[\frac{\partial J}{\partial w_1^*}\ \frac{\partial J}{\partial w_2^*}\ \cdots\ \frac{\partial J}{\partial w_M^*}\right]^{\mathrm{T}} = \frac{\partial(\sigma_s^2 + \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}}\mathbf{w} - \mathbf{p}^{\mathrm{H}}\mathbf{w} - \mathbf{w}^{\mathrm{H}}\mathbf{p})}{\partial \mathbf{w}^*} = \mathbf{R}_{\mathbf{x}}\mathbf{w} - \mathbf{p} = 0 \tag{2.27}$$

Hence, the Wiener filter $\mathbf{w}$ that minimizes (2.25) is

$$\mathbf{R}_{\mathbf{x}}\mathbf{w} = \mathbf{p} \Rightarrow \mathbf{w}_{\mathrm{opt}} = \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p} \tag{2.28}$$

For the (2×1) Alamouti code, where we need to recover the signals $s_1(n)$ and $s_2(n)$, and therefore we have to compute two Wiener filters

$$\mathbf{w}_1 = \mathbf{R}_{\mathbf{x}}^{-1}\mathrm{E}[s_1(n)\mathbf{x}(n)] = \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p}_1 \tag{2.29}$$

$$\mathbf{w}_2 = \mathbf{R}_{\mathbf{x}}^{-1}\mathrm{E}[s_2(n)\mathbf{x}(n)] = \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p}_2 \tag{2.30}$$

with $\mathbf{p}_i = \mathrm{E}[s_i^*(n)\mathbf{x}(n)]$. We can express the two Wiener filters jointly in a single matrix as

$$\mathbf{W}_{\text{opt}} = [\mathbf{w}_1 \ \mathbf{w}_2] = \mathbf{R}_{\mathbf{x}}^{-1}\,[\mathbf{p}_1 \ \mathbf{p}_2] \tag{2.31}$$

In order to employ the Wiener filter, we need to know the correlation matrix of observations $\mathbf{R}_{\mathbf{x}}$ and the cross-correlation between the observations and the sources $\mathbf{p}$. This information is in most cases unknown but $\mathbf{R}_{\mathbf{x}}$ can be estimated by means of a sample correlation matrix of L observations, i.e

$$\hat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{L}\sum_{n=1}^{L}\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n) \tag{2.32}$$

and $\mathbf{p}_i$ can be estimated employing pilot symbols.

2.4 Blind Source Separation

Blind Source Separation (BSS) algorithms have been widely used in digital communication to estimate the propagation coefficients (and transmitted signals) from the signals received in the antennas without using pilot symbols [32]. This lack of prior knowledge may limit the achievable performance, but makes blind approaches more robust to calibration errors (i.e. deviations of model assumptions from reality) than conventional array processing techniques [5]. A property commonly exploited in BSS is the statistical independence of the sources. Depending on the degree of independence considered, two main groups of techniques can be distinguished: *Principal Component Analysis* (PCA), which is based on *Second Order Statistics* (SOS) and *Independent Component Analysis* (ICA), which exploits *Higher Order Statistics* (HOS).

The classical problem to represent BSS is the Cocktail Party. It consists of a party with people talking at different places in a room and, also, several microphones placed at different points of the room collecting mixtures of the voices. Each voice is a source and the received signals in the microphones are the observations, which consist of a combination of the voices. The aim is to recover the sources from the collected signals in the microphones.

The simplest model in BSS considers that the mixtures are instantaneous combinations of the sources, i.e. we have a memoryless mixing system that can be expressed as

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v} \tag{2.33}$$

where $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_N]^{\mathrm{T}}$ is the source vector, $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_M]^{\mathrm{T}}$ is the observation vector, $\mathbf{H}$ is the mixing matrix, which has dimension $M \times N$, and $\mathbf{v} = [v_1 \ v_2 \ \dots \ v_M]^{\mathrm{T}}$ is the noise vector. Note that this model corresponds to the same structure as in the Alamouti scheme in the equation (2.5). The aim is to obtain a separation matrix $\hat{\mathbf{H}}^{-1}$ such as

$$\hat{\mathbf{H}}^{-1}\mathbf{H} = \mathbf{\Delta}\mathbf{P} \tag{2.34}$$

where $\mathbf{\Delta}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix. Thus, the sources can be achieved using

$$\hat{\mathbf{s}} = \mathbf{y} = \hat{\mathbf{H}}^{-1}\mathbf{x} \tag{2.35}$$

To achieve the sources from a set of observations, it is necessary for all the sources to be statistically independent and non-Gaussian. In fact, one of the sources can be Gaussian but it is usually coined noise instead of source. The Darmois-Skitovich Theorem [33, 34] proves that these constraints are necessary for achieving the separation. Note that these ideas are valid for real and complex sources. Taking this into account and for reasons of simplicity, we henceforth assume that the sources $\mathbf{s}$ will be independent and identically distributed (i.i.d.) real or complex random variables with variance equal to

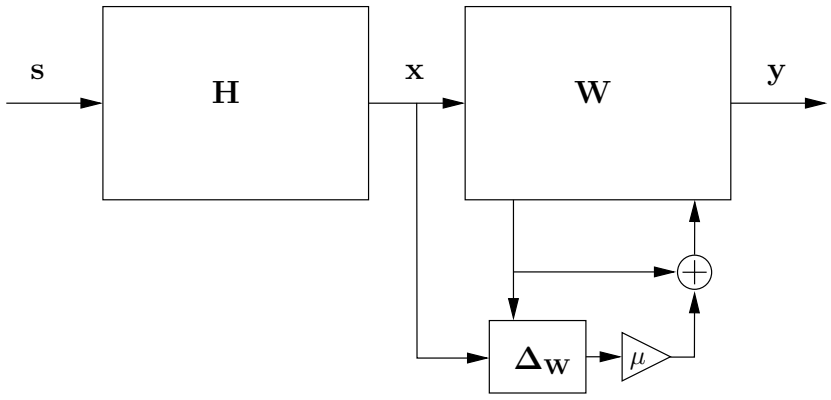


Figure 2.3: Adaptive BSS algorithms

one, i.e $\sigma_s^2 = \sigma_{s_i}^2 = 1,\ i = 1, 2, \ldots, N$. Thus, their correlation matrix will be

$$\mathbf{R_s} = \mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = \mathbf{I}_N \tag{2.36}$$

BSS algorithms can be classified into two main types according to their function:

- Adaptive algorithms obtain the result through several iterations. They have one or more adaptive parameters that influence on algorithm speed and the accuracy of the result.
- Batch algorithms achieve the separation system directly from a block input data.

2.4.1 Adaptive Algorithms

Adaptive algorithms are useful to tracking applications because they are able to adapt their parameters to changes in the environment [35]. In general, they work better with smaller data blocks than batch algorithms, and so they need less memory. Therefore, they are attractive in cases where the data input suffers frequent updatings. Their work can be summarized in Figure 2.3. The input is the observation vector $\mathbf{x} = [x_1\ x_2\ \ldots\ x_M]^{\mathrm{T}}$ and the output is the vector $\mathbf{y} = [y_1\ y_2\ \ldots\ y_N]^{\mathrm{T}}$. The matrix $\mathbf{W}$ has dimension $M \times N$ and contains the free coefficients of the separation system. The system is updated in each

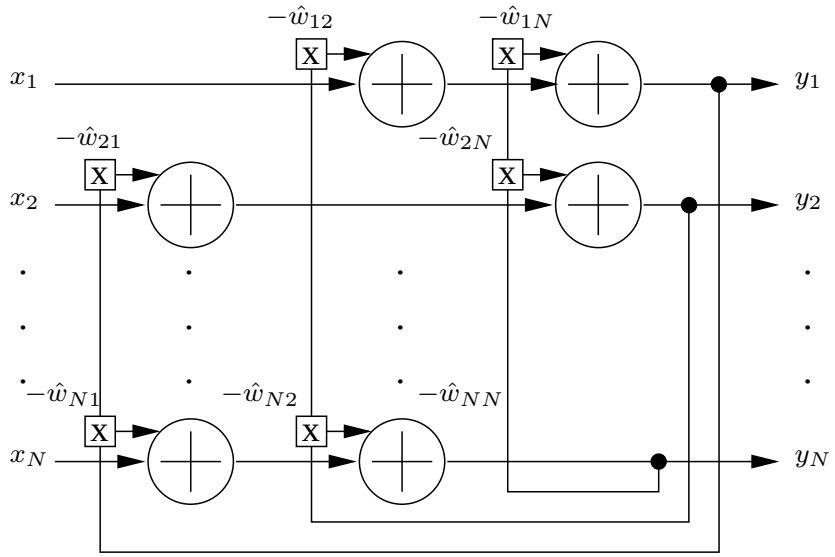


Figure 2.4: Separation system employed by Herault and Jutten

iteration following the rule

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \mathbf{\Delta}(\mathbf{W}(n), \mathbf{x}(n)) \tag{2.37}$$

where $\mu \in \mathbb{R}$ is the speed learning parameter, $\mathbf{\Delta}$ is a matrix usually composed of non-linear terms dependent on $\mathbf{W}$ and the input vector $\mathbf{x}$. The concrete form of matrix $\mathbf{\Delta}$ depends on the specific separation criterion.

We will now explain the basis of several of the best-known algorithms of this kind: Herault and Jutten, Infomax and FastICA.

Herault and Jutten's Algorithm

The first BSS algorithm was proposed in 1985 by Herault and Jutten [36]. The authors developed their work in the context of neural networks and they employed a separation system like the one shown in Figure 2.4. This separation system assumes that the number of inputs is the same as the number of outputs, and there are feedback paths which substract the inputs from the outputs on a weighted basis. The coefficient $\hat{\omega}_{ij}$ represents the factor which is multiplied by the output y_i before being subtracted from the input

x_i. Also, we assume that there are no self-feedbacks (i.e. $\hat{w}_{ii} = 0$). Mathematically, the relationship between input and output follows the rule

$$\mathbf{y}(n) = (\mathbf{I} + \hat{\mathbf{W}}^{\mathrm{T}}(n))^{-1}\mathbf{x}(n) \tag{2.38}$$

The decision criterion used to choose the matrix $\mathbf{W}$ is to minimize the statistical independence between the outputs of the network. Herault and Jutten propose maximizing the statistical independence between y_i and y_j by means of the following learning rule

$$\mathbf{W}(n+1) = \hat{\mathbf{W}}(n) + \mu\mathbf{f}(\mathbf{y}(n))\mathbf{g}^{\mathrm{T}}(\mathbf{y}(n)) \tag{2.39}$$

where $\mathbf{f}(\mathbf{y}) = [f(y_1)\ f(y_2)\ \ldots\ f(y_N)]^{\mathrm{T}}$ and $\mathbf{g}(\mathbf{y}) = [g(y_1)\ g(y_2)\ \ldots\ g(y_N)]^{\mathrm{T}}$ are two vectors composed of non-linear functions of the outputs $f(\cdot)$ and $g(\cdot)$. Considering that the signals to be separated have symmetric distributions, it is necessary for these non-linearities to be odd (i.e. $f(y_i) = -f(-y_i)$ and $g(y_i) = -g(-y_i)$) for their Taylor expansions to have odd powers. For example, $f(y_i) = y_i^m$ and $g(y_i) = y_i^n$ can be a possible choice.

The convergence point of this algorithm verifies

$$\frac{d\mathbf{W}}{dt} = \mathrm{E}[\mathbf{f}(\mathbf{y})\mathbf{g}^{\mathrm{T}}(\mathbf{y})] = 0 \tag{2.40}$$

Note that the following equality is verified when the outputs are statistically independent,

$$\mathrm{E}[\mathbf{f}(\mathbf{y})\mathbf{g}^{\mathrm{T}}(\mathbf{y})] = \mathrm{E}[\mathbf{f}(\mathbf{y})]\mathrm{E}[\mathbf{g}^{\mathrm{T}}(\mathbf{y})] = 0 \tag{2.41}$$

and for equation (2.40) this is a convergence point. Moreover, the theorem of Darmois-Skitovich guarantees that each output corresponds to a single and different source.

Bell and Sejnowski's Algorithm: Infomax

An important learning paradigm in neural networks is the principle of information preservation (infomax), which was proposed by Linsker [37]. According to this paradigm, the parameters of a neural network are adjusted with the aim of maximizing the transferred information between the input and the output. It is believed that concrete learning mechanisms of living beings work in this manner.

The method proposed by Bell and Sejnowski [38] is a gradient algorithm that maximizes the information transfer between the input and the output of a non-linear neural network with only one layer and the same number of inputs and outputs. The network is excited by the observations in order to achieve the activation vector of states, $\mathbf{y} = \mathbf{W}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{W}$ is a $N \times N$ square matrix that contains the synaptic weights. Final output is a non-linear function of state $\mathbf{u} = \mathbf{h}(\mathbf{y})$.

Bell and Sejnowski's algorithm pursues the maximization of the transfer between the input $\mathbf{x}$ and the output after the non-linearity of $\mathbf{u}$. If we use $H(\mathbf{u})$ to denominate the entropy of $\mathbf{u}$ and $H(\mathbf{u}|\mathbf{x})$ to denominate the uncertainty about $\mathbf{u}$ that is not solved when $\mathbf{x}$ is observed, we can thence write the mutual information, $I(\mathbf{u}, \mathbf{x})$, as

$$I(\mathbf{u}, \mathbf{x}) = H(\mathbf{u}) - H(\mathbf{u}|\mathbf{x}) \tag{2.42}$$

In [38], it has been demonstrated that the maximization of the mutual information $I(\mathbf{u}, \mathbf{x})$ is equivalent to maximizing the following cost function

$$J_{MI}(\mathbf{M}) \stackrel{def}{=} \ln(\det(\mathbf{W}^{\mathrm{T}})) + \sum_{i=1}^{N} \mathrm{E}[\ln(h_i'(y_i))] \tag{2.43}$$

The next step is to obtain an adaptive algorithm that maximizes $J_{MI}(\mathbf{M})$. With this aim, we can employ either a steepest descent algorithm [38] or a relative gradient algorithm [39, 40]. It can be demonstrated from equation (2.43) that

$$\nabla_{\mathbf{W}} J = \nabla_{\mathbf{W}} \left(\ln(\det \mathbf{W}^{\mathrm{T}}))\right) + \nabla_{\mathbf{W}} \left(\sum_{i=1}^{N} E\left[\ln(h_i'(y_i))\right] \right) = \frac{\mathrm{adj}(\mathbf{W}^{\mathrm{T}})}{\det(\mathbf{W}^{\mathrm{T}})} - \mathrm{E}[\mathbf{x}\mathbf{g}^{\mathrm{T}}(\mathbf{y})] = \mathbf{W}^{-\mathrm{T}} - \mathrm{E}[\mathbf{x}\mathbf{g}^{\mathrm{T}}(\mathbf{y})] \tag{2.44}$$

where $\mathbf{g}(\mathbf{y}) = [-h_1''(y_1)/h_1'(y_1) \ \ldots \ -h_N''(y_N)/h_N'(y_N)]^{\mathrm{T}}$ is a vector of non-linearities that depends on the activation function that has been employed. Estimating the expectations in the equation (2.44), we obtain the following learning rules:

- Gradient algorithm

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \left(\mathbf{W}^{-\mathrm{T}}(n) - \mathbf{x}(n)\mathbf{g}^{\mathrm{T}}(\mathbf{y}(n))\right) \tag{2.45}$$

- Relative gradient algorithm. This is obtained by multiplying the above algorithm by $\mathbf{W}(n)\mathbf{W}^{\mathrm{T}}(n)$.

$$\begin{aligned}\mathbf{W}(n+1) &= \mathbf{W}(n) + \mu\mathbf{W}(n)\mathbf{W}^{\mathrm{T}}(n)\left(\mathbf{W}^{-\mathrm{T}}(n) - \mathbf{x}(n)\mathbf{g}^{\mathrm{T}}(\mathbf{y}(n))\right) \\ &= \mathbf{W}(n) + \mu\mathbf{W}(n)\left(\mathbf{I} - \mathbf{x}(n)\mathbf{g}^{\mathrm{T}}(\mathbf{y}(n))\right) \end{aligned} \tag{2.46}$$

FastICA

FastICA is an efficient and popular ICA algorithm developed by Aapo Hyvärinen at Helsinki University of Technology in [41]. The algorithm, initially proposed for real-valued signals, is based on a fixed-point iteration scheme maximizing non-Gaussianity as a measure of statistical independence. It can also be derived as an approximative Newton iteration.

In order to explain the first step of the FastICA algorithm jointly with the first step of the JADE algorithm, which will be shown in Section 2.4.2, the hermitian operator $(\cdot)^{\mathrm{H}}$ will be employed instead of the transpose operator $(\cdot)^{\mathrm{T}}$, even though we are speaking of real matrices in this case.

So, the noiseless observation vector of dimension $M \times 1$ is defined as

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{2.47}$$

Recall that $\mathbf{H}$ is the mixing matrix of dimension $M \times N$, and $\mathbf{s}$ is the source vector of dimension $N \times 1$ which obeys equation (2.36) due to its i.i.d.

The first step is to find a whitening matrix $\mathbf{B}$ of dimension $N \times M$ carried out by the method described in Appendix B such that

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathbf{I}_N \tag{2.48}$$

Note that this new observation has the form

$$\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{H}\mathbf{s} = \mathbf{U}\mathbf{s} \tag{2.49}$$

Thus, we define a new matrix

$$\mathbf{U} = \mathbf{BH} \tag{2.50}$$

and from equations (2.48) and (2.49), it is straightforward to conclude that $\mathbf{U}$ is orthogonal, i.e.

$$\mathbf{I}_N = \mathbf{R}_{\tilde{\mathbf{x}}} = \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathrm{E}[\mathbf{U}\mathbf{s}\mathbf{s}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}] = \mathbf{U}\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}]\mathbf{U}^{\mathrm{H}} = \mathbf{U}\mathbf{U}^{\mathrm{H}} \tag{2.51}$$

Taking this into account, from equations (2.50) and (2.51)

$$\mathbf{U}^{\mathrm{H}}\tilde{\mathbf{x}} = \mathbf{U}^{\mathrm{H}}\mathbf{B}\mathbf{H}\mathbf{s} = \mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{s} = \mathbf{s} \tag{2.52}$$

Hence, with the aim of achieving the sources $\mathbf{s}$, the problem is reduced to finding the matrix $\mathbf{U}$.

The second step, like most suggested solutions to the ICA problem, is to use the 4^{th} order cumulant (*kurtosis*) of the signals, defined in Appendix E for a zero-mean random variable x as

$$\mathrm{kurt}(x) = \mathrm{E}[x^4] - 3\mathrm{E}^2[x^2] \tag{2.53}$$

The kurtosis is zero for Gaussian random variables, positive when the variables have densities peaked at zero and negative when the density is flatter. Note that for two independent variables x_1 and x_2 and for a scalar α, it obeys $\mathrm{kurt}(x_1 + x_2) = \mathrm{kurt}(x_1) + \mathrm{kurt}(x_2)$ and $\mathrm{kurt}(\alpha x_1) = \alpha^4\mathrm{kurt}(x_1)$.

Hence, we will find a linear combination of the sphered observations $\tilde{x}_i$, i.e. $\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}$, such that it has maximal or minimal kurtosis, considering that $||\mathbf{w}|| = 1$. Bearing this in mind, we can define $\mathbf{z} = \mathbf{U}^{\mathrm{H}}\mathbf{w}$, and then also $||\mathbf{z}|| = 1$. Using equation (2.49) and the properties of the kurtosis, we obtain the expression

$$\mathrm{kurt}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}) = \mathrm{kurt}(\mathbf{w}^{\mathrm{H}}\mathbf{U}\mathbf{s}) = \mathrm{kurt}(\mathbf{z}^{\mathrm{H}}\mathbf{s}) = \sum_{i=1}^{N} z_i^4\mathrm{kurt}(s_i) \tag{2.54}$$

Under the constraint $||\mathbf{w}|| = ||\mathbf{z}|| = 1$, it is straightforward to see that the function (2.54) has a number of local minima and maxima. For reasons of simplicity, we assume that there is at least one source with positive kurtosis and at least one source with negative

kurtosis. Hence, as was shown by Delfosse and Loubaton in [42], the extremal points of equation (2.54) are the canonical base vectors $\mathbf{z} = \pm\mathbf{e}_j$, i.e. vectors whose components are all zero except the j-th component, which equals 1. The corresponding weight vectors are $\mathbf{w} = \mathbf{U}\mathbf{z} = \pm\mathbf{U}\mathbf{e}_j = \pm\mathbf{u}_j$, for any j-th column of the orthogonal mixing matrix $\mathbf{U}$. So, under the given constraint, by minimizing or maximizing the kurtosis in the equation (2.54) the columns of $\mathbf{U}$ are obtained as solutions for $\mathbf{w}$, and the linear combination: $\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}} = \mathbf{u}_i^{\mathrm{H}}\tilde{\mathbf{x}} = s_i$ is one of the independent sources. Recall that equation (2.54) also shows that Gaussian components cannot be estimated in this way, because for them $\mathrm{kurt}(s_i)$ is zero.

In order to minimize or maximize equation (2.54), a neural algorithm based on gradient descent or ascent can be used, such as the one shown by Delfosse and Loubaton in [42] or that shown by Hyvärinen and Oja in [43]. Then, $\mathbf{w}$ is interpreted as the weight vector of a neuron input vector $\tilde{\mathbf{x}}$ and the objective function can be simplified due to the fact that the inputs are sphered: it obeys

$$\mathrm{kurt}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}) = \mathrm{E}[(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^4] - 3\mathrm{E}^2[(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^2] = \mathrm{E}[(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^4] - 3||\mathbf{w}||^4 \tag{2.55}$$

The constraint $||\mathbf{w}|| = 1$ must also be taken into account, e.g. by a penalty term [43]. The final objective function is

$$J(\mathbf{w}) = \mathrm{E}[(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^4] - 3||\mathbf{w}||^4 + F(||\mathbf{w}||^2) \tag{2.56}$$

where F is a penalty term due to the constraint. Henceforth, the exact form of F is not important. Denoting the sequence of the observations by $\tilde{\mathbf{x}}(n)$, the learning rate sequence by $\mu(n)$, and the derivative of $F/2$ by f, the on-line learning algorithm then has the form

$$\mathbf{w}(n+1) = \mathbf{w}(n) \pm \mu(n)[\tilde{\mathbf{x}}(n)(\mathbf{w}(n)^{\mathrm{H}}\tilde{\mathbf{x}}(n))^3 - 3||\mathbf{w}(n)||^2\mathbf{w}(n) + f(||\mathbf{w}(n)||^2)\mathbf{w}(n)] \tag{2.57}$$

The first two terms in brackets are obtained from the gradient of $\mathrm{kurt}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})$ when instantaneous values are used instead of the expectation. The third term in brackets is obtained from the gradient of $F(||\mathbf{w}||^2)$; note that as long as this is a function of $||\mathbf{w}||^2$ only, its gradient has the form $\alpha\mathbf{w}$, where $\alpha \in \mathbb{R}$. A Positive sign before the brackets means finding the local maxima, negative sign corresponds to local minima.

The convergence of this kind of algorithms can be proved by using the principles of stochastic approximation [44]. The advantage of such neural learning rules is that the inputs $\tilde{\mathbf{x}}(n)$ can be used in the algorithm immediately, thus enabling fast adaptation in a non-stationary environment. A resulting trade-off, however, is that convergence is slow and depends on a good choice of the learning rate sequence $\mu(n)$. A bad choice of the learning rate can, in practice, destroy convergence. Therefore, some ways of making the learning radically faster and more reliable may be needed: fixed-point iteration algorithms are such an alternative.

The fixed points $\mathbf{w}$ of the learning rule (2.57) are obtained by taking the expectations and equating the change in the weight to 0:

$$\mathrm{E}[\tilde{\mathbf{x}}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^3] - 3||\mathbf{w}||^2\mathbf{w} + f(||\mathbf{w}||^2)\mathbf{w} = 0 \tag{2.58}$$

The time index n has been dropped. A deterministic iteration could be formed from equation (2.58) in a number of ways, e.g. by standard numerical algorithms for solving such equations. A very fast iteration is obtained, as shown in the next section, if we write (2.57) in the form

$$\mathbf{w} = \alpha(\mathrm{E}[\tilde{\mathbf{x}}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^3] - 3||\mathbf{w}||^2\mathbf{w}) \tag{2.59}$$

Actually, because the norm of $\mathbf{w}$ is irrelevant, it is the direction of the right hand side that is important. The *scalar* in equation (2.59) is therefore not significant and its effect can be replaced by explicit normalization or the projection of $\mathbf{w}$ onto the unit sphere.

These ideas have been extended in [45] for complex-valued signals. In this case

$$\mathbf{w} = \alpha(\mathrm{E}[\tilde{\mathbf{x}}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^* g(|\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}|^2)] - \mathrm{E}[g(|\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}|^2) + |\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}|^2 g'(|\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}|^2)]\mathbf{w}) \tag{2.60}$$

For the particular case, when $g(x) = x$ and $g'(x) = 1$, we can rewrite equation (2.60) as

$$\mathbf{w} = \alpha(\mathrm{E}[\tilde{\mathbf{x}}(\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}})^* |\mathbf{w}^{\mathrm{H}}\tilde{\mathbf{x}}|^2] - 2||\mathbf{w}||^2\mathbf{w}) \tag{2.61}$$

Note that the above expression for complex-valued signals $\mathbf{w}$ is close to the expression (2.59) for real-valued signals.

- **Step 1**. Let $i = 1$.
- **Step 2**. Let $\mathbf{w}(0)$ be a random vector where $||\mathbf{w}(0)|| = 1$. Let $k = 1$.
- **Step 3**. Let

$$\begin{aligned}\mathbf{w}(k) &= \mathrm{E}[\tilde{\mathbf{x}}(\mathbf{w}(k-1)^{\mathrm{H}}\tilde{\mathbf{x}})^* + g(|\mathbf{w}(k-1)^{\mathrm{H}}\tilde{\mathbf{x}}|^2)] + \\ &- \mathrm{E}[g(|\mathbf{w}(k-1)^{\mathrm{H}}\tilde{\mathbf{x}}|^2) + |\mathbf{w}(k-1)^{\mathrm{H}}\tilde{\mathbf{x}}|^2 g'(|\mathbf{w}(k-1)^{\mathrm{H}}\tilde{\mathbf{x}}|^2)]\mathbf{w}(k-1)\end{aligned}$$

  The expectation can be estimated using a large sample of $\mathbf{x}$ vectors (say $1,000$ points).
- **Step 4**. Let $\mathbf{w}(k) = \frac{\mathbf{w}(k)}{||\mathbf{w}(k)||}$.
- **Step 5**. If $|\mathbf{w}(k)^{\mathrm{H}}\mathbf{w}(k-1)|$ is not close enough to 1, let $k = k + 1$ and return to step 3. Otherwise, output the vector $\mathbf{w}_i = \mathbf{w}(k)$.

Table 2.1: FastICA algorithm

Assuming that we have collected a sample of the sphered (or prewhitened) random vector $\tilde{\mathbf{x}}$ and using the derivation of the equation (2.60), we get the *fixed-point algorithm for ICA* shown in Table 2.1. The final vector $\mathbf{w}_i$ given by the algorithm equals one of the columns of the (orthogonal) mixing matrix $\mathbf{U}$. In the case of blind source separation, this means that $\mathbf{w}_i$ separates *one* of the non-Gaussian source signals in the sense that $\mathbf{w}_i^{\mathrm{H}}\tilde{\mathbf{x}}$ equals one of the source signals.

A remarkable property of this algorithm is that a very small number of iterations, usually 5-10, seems to be enough to obtain the maximal accuracy allowed by the sample data. This is due to the cubic convergence.

In order to achieve the remaining $N - 1$ sources and guarantee the separation of the sources, it is necessary to decorrelate each output $\mathbf{w}_i^{\mathrm{H}}$ after every one is obtained. There are several methods of resolving this situation. A simple way is a deflation scheme

based on a Grand-Schmidt-like decorrelation. This consists of estimating the independent components one by one. After obtaining i vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_i$, in order to achieve the next vector $\mathbf{w}_{i+1}$, it has to let $i = i+1$, go to step 2, and when we have the desired vector, it has to subtract from $\mathbf{w}_i$ the projections $\mathbf{w}_j \mathbf{w}_j^{\mathrm{H}} \mathbf{w}_i,\ j = 1, 2, \ldots, i-1$ of the previously $i-1$ estimated vectors

$$\mathbf{w}_i = \mathbf{w}_i - \sum_{j=1}^{i-1} \mathbf{w}_j \mathbf{w}_j^{\mathrm{H}} \mathbf{w}_i \tag{2.62}$$

and then renormalize $\mathbf{w}_i$

$$\mathbf{w}_i = \frac{\mathbf{w}_i}{||\mathbf{w}_i||} \tag{2.63}$$

2.4.2 Batch Algorithms

Batch Algorithms achieve the separation system directly from a block input data. The *Joint Approximate Diagonalization of Eigenmatrices* (JADE) is one of the best-known algorithms of this kind. This algorithm, developed by Cardoso and Souloumiac in 1993 [5], is based on the simultaneous diagonalization of cumulant matrices. 'Good' statistical performance is achieved by involving all the 2^{nd} order and 4^{th} order cumulants while a fast optimization is obtained by the device of joint diagonalization.

JADE

In order to explain the principles of the JADE algorithm, we consider N i.i.d. sources that obey equation (2.36) and M noiseless observations that have the form

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{2.64}$$

As in the case of FastICA explained in Section 2.4.1, JADE needs a first step in which to work with sphered observations $\tilde{\mathbf{x}}$, and hence we assume all equations from (2.47) to (2.52) for this algorithm. Recall that the sphered observations are

$$\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{H}\mathbf{s} = \mathbf{U}\mathbf{s},\ \mathbf{R}_{\tilde{\mathbf{x}}} = \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathbf{I}_N \tag{2.65}$$

Note that $\mathbf{U} = \mathbf{BH}$ and it has dimension $N \times N$. For explaining the following equations, it is interesting to denote its columns by

$$\mathbf{U} = \left[\begin{array}{cccc} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_N \end{array} \right] \tag{2.66}$$

Before explaining the second step, we define the kurtosis of the i-th source as

$$\rho_i = \text{kurt}(s_i) = \text{cum}(s_i, s_i^*, s_i, s_i^*) \tag{2.67}$$

where $\text{cum}(\cdot)$ represents the cumulant function (see Appendix E for more details about definition and properties). For any $N \times N$ matrix $\mathbf{M}$, it is possible to associate it with a cumulant matrix $\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})$, defined by

$$[\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})]_{ij} = \sum_{k,l=1}^{N} \text{cum}(\tilde{x}_i, \tilde{x}_j^*, \tilde{x}_k, \tilde{x}_l^*) m_{lk}, \ 1 \leq i, j \leq N \tag{2.68}$$

Taking into the cumulant properties account, it has been demonstrated in [5] that $\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})$ can be decomposed in

$$\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M}) = \mathbf{U} \boldsymbol{\Delta}_{\mathbf{M}} \mathbf{U}^{\text{H}} \tag{2.69}$$

where

$$\boldsymbol{\Delta}_{\mathbf{M}} = \text{diag}(\rho_1 \mathbf{u}_1^{\text{H}} \mathbf{M} \mathbf{u}_1, \rho_2 \mathbf{u}_2^{\text{H}} \mathbf{M} \mathbf{u}_2, \dots, \rho_N \mathbf{u}_N^{\text{H}} \mathbf{M} \mathbf{u}_N) \tag{2.70}$$

Hence, from equation (2.69) above it is possible to recover the vector $\mathbf{s}$ of N sources and it is possible to achieve the mixing matrix $\mathbf{H}$ by means of the following equation

$$\mathbf{H} = \mathbf{B}^{\sharp} \mathbf{U} \tag{2.71}$$

where $(\cdot)^{\sharp}$ is the pseudoinverse operator.

Summarizing, JADE can be described by the steps shown in Table 2.2. Actually, the JADE algorithm presented in [5] is more complicated because the selection of the matrices to be diagonalized is done by computing the EVD of an $N^2 \times N^2$ matrix whose columns correspond to the matrices $\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})$ for any matrix $\mathbf{M}$. In Appendix F, we describe an optimized version of the Jacobi technique when $N = 2$, which is particularly attractive for the (2×1) Alamouti code.

- **Step 1.** Compute a whitening matrix $\mathbf{B}$ of dimension $N \times M$ making the *Eigenvalue Decomposition* (EVD) of the correlation matrix $\mathbf{R}_{\mathbf{x}}$. To this end, some of the methods described in Appendix B can be used. Thus, you will have achieved a white vector $\tilde{\mathbf{x}}$ of dimension N.
- **Step 2.** Compute the N^2 4^{th} order cumulants in matrices $\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})$ of the whitened process $\tilde{\mathbf{x}}(n) = \mathbf{B}\mathbf{x}(n)$
- **Step 3.** Jointly diagonalize the N most significant 4^{th} order cumulants in matrices $\mathbf{Q}_{\tilde{\mathbf{x}}}(\mathbf{M})$.
- **Step 4.** The estimation of $\mathbf{H}$ is $\hat{\mathbf{H}} = \mathbf{B}^{\sharp}\mathbf{U}$.

Table 2.2: JADE algorithm

2.5 Comparison between Classical and BSS Estimation Algorithms

In this section we make a performance comparison between the channel estimation algorithms described above:

- The Wiener filter with different size of the pilot symbol set: 2%, 10% and 20% of the total symbols.
- The complex FastICA with the non-linearity function $g(y) = 1/(\epsilon + y)$ with $\epsilon = 0.1$.
- The JADE algorithm.

These algorithms have been tested through the following computer simulated scenario. Blocks of 1000 symbols were generated from an equiprobable distribution which symbols belongs a 4-QAM which have been coded using the (2×1) Alamouti code. For that,

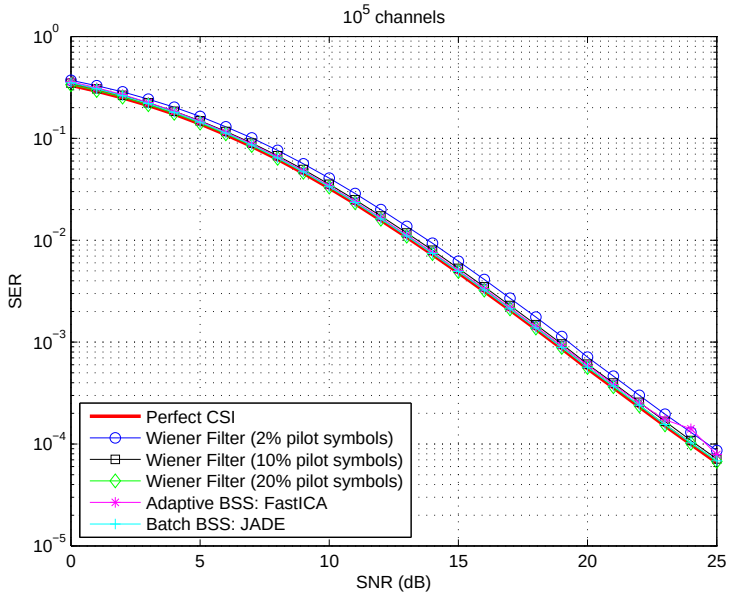


Figure 2.5: Comparison between supervised and blind algorithms for Rayleigh channel distribution: SER versus SNR

each generated block has been rearranged as 500 symbols for each source. The transmission of the symbols is carried out through block fading channels with Rayleigh or Rice distribution.

The performance has been measured in terms of the *Symbol Error Rate* (SER) for different values of the *Signal to Noise Ratio* (SNR). The results have been averaged over 10^5 channel realizations and transmitted symbols. Perfect CSI obtained using the theoretical channel matrix is used as a reference to compare performance.

Employing these parameters, Figure 2.5 shows the performance for Rayleigh channel distribution. We can see that the Wiener filter differs by only 0.5 dB when the pilot symbols are 2% of the total and achieves Perfect CSI when they are over 10%. FastICA and JADE performance is also close to Perfect CSI.

Figure 2.6 shows the results obtained for Rice channel distribution. All the algorithms achieve a SER of 10^{-4} with 10 dB, while for the Rayleigh case 25 dB are needed. Note also that FastICA presents a poor performance for SNRs when are greater than 12 dBs.

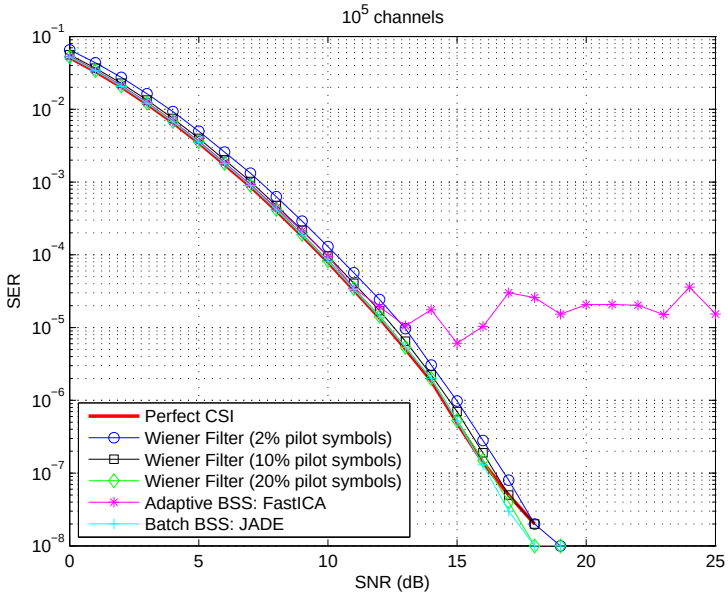


Figure 2.6: Comparison between supervised and blind algorithms for Rice channel distribution: SER versus SNR

The reason for this undesirable situation may be the non-linearity function selection.

Chapter 3

SOS-based Approaches

One of the best known SOS-based approaches is *Principal Component Analysis* (PCA) [46], introduced by Pearson in 1901 [47] in a biological context to recast linear regression analysis into a new form. This technique is based on finding the eigenvector and eigenvalues of the correlation matrix obtained from the original data. PCA is widely used in data compression where the eigenvectors represent the basis used to transform the original data into a reduced set of uncorrelated components. It is also employed for whitening the data before applying other algorithms such as FastICA and JADE.

Taking $\mathbf{x}$ to be the original data vector, the correlation matrix is defined by $\mathbf{R_x} = \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]$. In the particular case where the observations are noised linear mixtures of sources $\mathbf{s}$, i.e. $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v}$ where $\mathbf{v}$ is the noise vector, the correlation matrix has the form

$$\mathbf{R_x} = \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \mathbf{H}\mathbf{R_s}\mathbf{H}^{\mathrm{H}} + \sigma_v^2\mathbf{I} \tag{3.1}$$

The term σ_v^2 represents the noise variance and $\mathbf{R_s} = \mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}]$ is the correlation matrix of the sources. For the particular case of the (2×1) Alamouti code defined in Section 2.1 where the channel matrix is orthogonal, $\mathbf{H}\mathbf{H}^{\mathrm{H}} = ||\mathbf{h}||^2\mathbf{I}_2$, the correlation matrix given in the equation (3.1) can be rewritten as

$$\mathbf{R_x} = \mathbf{H}\mathbf{R_s}\mathbf{H}^{\mathrm{H}} + \frac{\sigma_v^2}{||\mathbf{h}||^2}\mathbf{H}\mathbf{H}^{\mathrm{H}} = \mathbf{H}\left(\mathbf{R_s} + \frac{\sigma_v^2}{||\mathbf{h}||^2}\mathbf{I}_2\right)\mathbf{H}^{\mathrm{H}} = \mathbf{H}\boldsymbol{\Delta}_{\mathrm{SOS}}\mathbf{H}^{\mathrm{H}} \tag{3.2}$$

- **Step 1**. Compose the correlation observation matrix $\mathbf{R_x} = \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]$.
- **Step 2**. Let $\mathbf{C} = \mathbf{R_x}$.
- **Step 3**. Compute the EVD of $\mathbf{C}$. The result is $\mathbf{C} = \hat{\mathbf{H}}\mathbf{\Delta}_{\mathrm{SOS}}\hat{\mathbf{H}}^{\mathrm{H}}$.
- **Step 4**. Return $\hat{\mathbf{H}}$.

Table 3.1: General SOS algorithm: PCA

So, we can think that the matrix $\mathbf{H}$ can be identified by computing the eigenvectors of the correlation matrix, applying the procedure described in Table 3.1. Note that if the two sources have the same variance, $\sigma_s^2 = \mathrm{E}[|s_1|^2] = \mathrm{E}[|s_2|^2]$, the above expression has the form

$$\mathbf{R_x} = (\sigma_s^2 + \frac{\sigma_v^2}{||\mathbf{h}||^2})\mathbf{H}\mathbf{H}^{\mathrm{H}} = (||\mathbf{h}||^2\sigma_s^2 + \sigma_v^2)\mathbf{I}_2 \tag{3.3}$$

This is true for any orthogonal matrix and, hence, it is impossible to identify $\mathbf{H}$.

The limitations of using SOS in OSTBC were first focused on by Shahbazpanahi et al. in [3]. Using specific properties of OSTBC, including (2×1) Alamouti code, these authors proposed a closed form to make a blind estimation of the channel matrix and show that the SOS-based channel estimation in the (2×1) Alamouti code can be realized by using a simple linear precoding matrix selected in order to guarantee that the eigenvalues of the correlation matrix have an order equal to one, which implies that the encoded sources have different variances that can be obtained using a precoder matrix before the Alamouti code. In this sense, Via et al. [4] have proposed an adaptive procedure which makes it possible to estimate the channel matrix when the eigenvalues of $\mathbf{R_x}$ are different. More recently, in [28] we have proposed a compact form of interpreting the approach of Shahbazpanahi et al., and we have studied how to choose the best precoding matrix for Rayleigh and Rice distributed channels which keeps the orthogonality property of the channel matrix.

This chapter is structured as follows. In Section 3.1 we present a real-valued notation to express the (2×1) Alamouti code. Using this notation, Section 3.2 presents the basis of the blind SOS approach proposed by Shahbazpanahi et al. In Section 3.3 an adaptive implementation of Shahbazpanahi et al.'s approach is shown, as proposed by Via et al. in [4]. In this section, we also propose an alternative implementation for this algorithm considering a complex-valued model. Finally, Section 3.4 shows how to choose the best precoding matrix for SOS-based approaches, keeping orthogonality coding, in the context of Rayleigh and Rice channels.

3.1 Real-valued Representation of Alamouti Code

In order to formulate the SOS-based techniques, we will introduce the real-valued notation of the (2×1) Alamouti code. To this end we define the operator $\underline{(\cdot)}$ for any vector $\mathbf{a}$ and for any matrix $\mathbf{A}$ as $\underline{\mathbf{a}} = \left[\Re\{\mathbf{a}\}^{\mathrm{T}}\ \Im\{\mathbf{a}\}^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\underline{\mathbf{A}} = \left[\Re\{\mathbf{A}\}^{\mathrm{T}}\ \Im\{\mathbf{A}\}^{\mathrm{T}}\right]^{\mathrm{T}}$, respectively.

As in the case of the complex notation introduced in Section 2.1, the source vector is $\mathbf{s} = [s_1\ s_2]^{\mathrm{T}}$ and the noise vector is redefined as $\mathbf{v} = [v_1\ v_2^*]^{\mathrm{T}}$. Remember that the received signals have the form

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} h_1 s_1 + h_2 s_2 + v_1 \\ h_2 s_1^* - h_1 s_2^* + v_2^* \end{bmatrix} \tag{3.4}$$

Hence, in order to work separately with the real and imaginary parts of the sources, we define a set formed by 4 coding matrices

$$\mathcal{C}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\ \mathcal{C}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\ \mathcal{C}_3 = \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}\ \mathcal{C}_4 = \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \tag{3.5}$$

which satisfy

$$\mathcal{C}_k^{\mathrm{H}}\mathcal{C}_l = \begin{cases} \mathbf{I}_2, & k = l \\ -\mathcal{C}_l^{\mathrm{H}}\mathcal{C}_k, & k \neq l \end{cases} \tag{3.6}$$

The real-valued source vector is $\underline{\mathbf{s}} = [\Re\{s_1\}\ \Re\{s_2\}\ \Im\{s_1\}\ \Im\{s_2\}]^{\mathrm{T}}$ which is formed by 4 elements. We will define a source matrix whose columns represent the antenna used to

transmit and whose rows represent the time slot,

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \tag{3.7}$$

Denoting by $\underline{s}_k$ the k-th entry in $\underline{\mathbf{s}}$, we can interpret $\mathcal{C}_k$ as the matrix that indicates the antenna and time slot used to transmit the symbol $\underline{s}_k$. As a result, the equation above can be rewritten as follows

$$\mathbf{S} = \begin{bmatrix} \Re\{s_1\} + j\Im\{s_1\} & \Re\{s_2\} + j\Im\{s_2\} \\ -(\Re\{s_2\} - j\Im\{s_2\}) & \Re\{s_1\} - j\Im\{s_1\} \end{bmatrix} = \sum_{k=1}^{4} \mathcal{C}_k \underline{s}_k \tag{3.8}$$

Thus we can rewrite the received signal vector in the equation (3.4) as

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{v} \tag{3.9}$$

where the channel vector is defined by

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \tag{3.10}$$

Using the form of matrices $\mathcal{C}_k$ in the equation (3.5), the vector above can be rewritten as

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{v} = \sum_{k=1}^{4} \mathcal{C}_k \underline{s}_k \mathbf{h} + \mathbf{v} = \sum_{k=1}^{4} \mathcal{C}_k \mathbf{h} \underline{s}_k + \mathbf{v} \tag{3.11}$$

Defining the vector $\mathbf{w}_k(\mathbf{h}) = \mathcal{C}_k\mathbf{h}$ and the matrix

$$\mathbf{W}(\mathbf{h}) = [\mathbf{w}_1(\mathbf{h})\ \mathbf{w}_2(\mathbf{h})\ \mathbf{w}_3(\mathbf{h})\ \mathbf{w}_4(\mathbf{h})] = [\mathcal{C}_1\mathbf{h}\ \mathcal{C}_2\mathbf{h}\ \mathcal{C}_3\mathbf{h}\ \mathcal{C}_4\mathbf{h}] \tag{3.12}$$

we can rewrite equation (3.11) as

$$\mathbf{r} = \sum_{k=1}^{4} \mathbf{w}_k(\mathbf{h}) \underline{s}_k + \mathbf{v} = \mathbf{W}(\mathbf{h})\underline{\mathbf{s}} + \mathbf{v} \tag{3.13}$$

Finally, using the operator $\underline{(\cdot)}$, we can obtain the real-valued representation of the receiving signals, $\underline{\mathbf{r}} = \left[\Re\{\mathbf{r}\}^{\mathrm{T}}\ \Im\{\mathbf{r}\}^{\mathrm{T}}\right]^{\mathrm{T}}$, which obeys the following expression

$$\underline{\mathbf{r}} = \sum_{k=1}^{4} \underline{\mathbf{w}}_k(\mathbf{h}) \underline{s}_k + \underline{\mathbf{v}} = \underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{s}} + \underline{\mathbf{v}} \tag{3.14}$$

We will prove that the matrix $\underline{\mathbf{W}}(\mathbf{h})$ is orthogonal. Directly, from equations (3.10) and (3.12), it can be obtained that the matrix $\mathbf{W}(\mathbf{h})$ has the form

$$\mathbf{W}(\mathbf{h}) = \begin{bmatrix} h_1 & h_2 & jh_1 & jh_2 \\ h_2 & -h_1 & -jh_2 & jh_1 \end{bmatrix} \tag{3.15}$$

and the matrix $\underline{\mathbf{W}}(\mathbf{h}) = [\Re\{\mathbf{W}(\mathbf{h})\}^{\mathrm{T}}\ \Im\{\mathbf{W}(\mathbf{h})\}^{\mathrm{T}}]^{\mathrm{T}}$ is

$$\underline{\mathbf{W}}(\mathbf{h}) = \begin{bmatrix} \Re\{h_1\} & \Re\{h_2\} & -\Im\{h_1\} & -\Im\{h_2\} \\ \Re\{h_2\} & -\Re\{h_1\} & \Im\{h_2\} & -\Im\{h_1\} \\ \Im\{h_1\} & \Im\{h_2\} & \Re\{h_1\} & \Re\{h_2\} \\ \Im\{h_2\} & -\Im\{h_1\} & -\Re\{h_2\} & \Re\{h_1\} \end{bmatrix} \tag{3.16}$$

By performing simple operations and considering that $|h_i|^2 = \Re\{h_i\}^2 + \Im\{h_i\}^2$, we obtain that

$$\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h})\underline{\mathbf{W}}(\mathbf{h}) = (|h_1|^2 + |h_2|^2) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = ||\mathbf{h}||^2 \mathbf{I}_4 \tag{3.17}$$

This property allows us to estimate the sources from equation (3.14) by means of

$$\hat{\underline{\mathbf{s}}} = \frac{\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h})\underline{\mathbf{r}}}{||\mathbf{h}||^2} \tag{3.18}$$

which is the ML criterion.

The complex source vector $\mathbf{s}$ can then be obtained as

$$\hat{\mathbf{s}} = [\mathbf{I}_2\ j\mathbf{I}_2]\,\hat{\underline{\mathbf{s}}} \tag{3.19}$$

where $\mathbf{I}_2$ is the 2×2 identity matrix. In order to apply equation (3.18) we need to know the vector channel $\mathbf{h}$ or an estimation of it.

3.2 The Shahbazpanahi et al. Approach

One of the most significant SOS-based techniques for channel identification for OSTBC has been proposed by Shahbazpanahi et al. in [3]. In this section, we will formulate this

technique using the real-valued notation introduced in the previous section.

The technique is based on unbalancing the source variance by using a diagonal encoder matrix

$$\mathbf{D} = \begin{bmatrix} \sigma_{z_1} & 0 \\ 0 & \sigma_{z_2} \end{bmatrix} \tag{3.20}$$

Thus, the new encoded sources are

$$\mathbf{z} = \mathbf{D}\mathbf{s} \tag{3.21}$$

In order to obtain the relation between $\underline{\mathbf{z}}$ and $\underline{\mathbf{s}}$, we define an alternative encoder matrix

$$\mathbf{D}' = \begin{bmatrix} \sigma_{z_1} & 0 & 0 & 0 \\ 0 & \sigma_{z_2} & 0 & 0 \\ 0 & 0 & \sigma_{z_1} & 0 \\ 0 & 0 & 0 & \sigma_{z_2} \end{bmatrix} \tag{3.22}$$

Thus, it is possible to establish the following relationship

$$\underline{\mathbf{z}} = \mathbf{D}'\underline{\mathbf{s}} \tag{3.23}$$

The received signals are linear combinations of $\mathbf{z}$ and they follow the model in equation (3.14) using $\mathbf{z}$ instead of $\mathbf{s}$, i.e

$$\underline{\mathbf{r}} = \sum_{k=1}^{4} \underline{\mathbf{w}}_k(\mathbf{h})\underline{z}_k + \underline{\mathbf{v}} = \underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{z}} + \underline{\mathbf{v}} \tag{3.24}$$

and their correlation matrix is given by

$$\mathbf{R}_{\underline{\mathbf{r}}} = \mathrm{E}[\underline{\mathbf{r}}\underline{\mathbf{r}}^{\mathrm{T}}] = \underline{\mathbf{W}}(\mathbf{h})\mathbf{R}_{\underline{\mathbf{z}}}\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h}) + \frac{\sigma_{\mathbf{v}}^2}{2}\mathbf{I}_4 = \underline{\mathbf{W}}(\mathbf{h})\mathbf{D}'\mathbf{R}_{\underline{\mathbf{s}}}\mathbf{D}'\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h}) + \frac{\sigma_{\mathbf{v}}^2}{2}\mathbf{I}_4 \tag{3.25}$$

Directly from equation (3.17), we obtain

$$\frac{\sigma_{\mathbf{v}}^2}{2}\mathbf{I}_4 = \frac{\sigma_{\mathbf{v}}^2}{2||\mathbf{h}||^2}\underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h}) \tag{3.26}$$

and the correlation matrix can be expressed as

$$\mathbf{R}_{\underline{\mathbf{r}}} = \underline{\mathbf{W}}(\mathbf{h})\left(\mathbf{D}'\mathbf{R}_{\underline{\mathbf{s}}}\mathbf{D}' + \frac{\sigma_{\mathbf{v}}^2}{2||\mathbf{h}||^2}\right)\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h}) \tag{3.27}$$

Since $\underline{\mathbf{W}}(\mathbf{h})$ is an orthogonal matrix and the sources have the same variance, the channel can be identified using an EVD if the values in the diagonal of $\mathbf{D}'$ are different, i.e, $\sigma_{z_1} \neq \sigma_{z_2}$. In fact, Shahbazpanahi et al. have proved that this identification problem can be solved using the following optimization criterion

$$\arg\max_{\hat{\mathbf{h}}} \frac{\mathrm{Tr}(\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\mathbf{R}_{\underline{\mathbf{r}}}\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \tag{3.28}$$

In Appendix D, using the properties of the Kronecker operator $\otimes$, it is shown that the numerator of this expression can be written as

$$\mathrm{Tr}(\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\mathbf{R}_{\underline{\mathbf{r}}}\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') = \mathrm{vec}^{\mathrm{T}}(\underline{\mathbf{W}}(\hat{\mathbf{h}}))(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') \tag{3.29}$$

We will express this criterion in a compact form by defining the following matrices

$$\mathcal{D}_k = \begin{bmatrix} \mathcal{C}_k & j\mathcal{C}_k \end{bmatrix}, \ k = 1, \ldots, 4 \tag{3.30}$$

where $\mathcal{C}_k, \ k = 1, \ldots, 4$ have been defined in equation (3.5). Applying the operator $(\underline{\cdot})$, expression (3.16) can also be rewritten using the matrices defined in [27]

$$\underline{\mathcal{D}}_k = \begin{bmatrix} \Re\{\mathcal{C}_k\} & -\Im\{\mathcal{C}_k\} \\ \Im\{\mathcal{C}_k\} & \Re\{\mathcal{C}_k\} \end{bmatrix}, \ k = 1, \ldots, 4 \tag{3.31}$$

which satisfy

$$\underline{\mathcal{D}}_k^{\mathrm{H}}\underline{\mathcal{D}}_l = \begin{cases} \mathbf{I}_4, & k = l \\ -\underline{\mathcal{D}}_l^{\mathrm{H}}\underline{\mathcal{D}}_k, & k \neq l \end{cases} \tag{3.32}$$

We also define the 16×4 matrix

$$\mathbf{\Phi} = \begin{bmatrix} \underline{\mathcal{D}}_1 \\ \underline{\mathcal{D}}_2 \\ \underline{\mathcal{D}}_3 \\ \underline{\mathcal{D}}_4 \end{bmatrix} \tag{3.33}$$

Since the matrices $\underline{\mathcal{D}}_k$ are orthogonal, this matrix obeys the following relationship

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} = 4\mathbf{I}_4 \tag{3.34}$$

In Appendix D, the following expression is demonstrated

$$\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})) = \mathbf{\Phi}\hat{\underline{\mathbf{h}}} \tag{3.35}$$

where the matrix $\underline{\mathbf{W}}(\hat{\mathbf{h}})$ is related to the matrices $\underline{\mathcal{D}}_k,\ k = 1, .., 4$ through the expression

$$\underline{\mathbf{W}}(\hat{\mathbf{h}}) = \left[\begin{array}{cccc} \underline{\mathcal{D}}_1\underline{\hat{\mathbf{h}}} & \underline{\mathcal{D}}_2\underline{\hat{\mathbf{h}}} & \underline{\mathcal{D}}_3\underline{\hat{\mathbf{h}}} & \underline{\mathcal{D}}_4\underline{\hat{\mathbf{h}}} \end{array} \right] \tag{3.36}$$

The matrix $\mathbf{\Psi}$ from equations (3.22) and (3.33) is therefore defined as

$$\mathbf{\Psi} = (\mathbf{D}' \otimes \mathbf{I}_4)\mathbf{\Phi} = \left[\begin{array}{c} \sigma_{z_1}\underline{\mathcal{D}}_1 \\ \sigma_{z_2}\underline{\mathcal{D}}_2 \\ \sigma_{z_1}\underline{\mathcal{D}}_3 \\ \sigma_{z_2}\underline{\mathcal{D}}_4 \end{array} \right] \tag{3.37}$$

In Appendix D, it is also proved that it is possible to establish the following relationship

$$\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') = \mathbf{\Psi}\underline{\hat{\mathbf{h}}} \tag{3.38}$$

Substituting equations (3.35) and (3.38) in the equation (3.29), we obtain the equation

$$\begin{aligned} \mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}}))^{\mathrm{T}}(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') &= \underline{\hat{\mathbf{h}}}^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathbf{\Psi}\underline{\hat{\mathbf{h}}} \\ &= \underline{\hat{\mathbf{h}}}^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})(\mathbf{D}' \otimes \mathbf{I}_4)\mathbf{\Phi}\underline{\hat{\mathbf{h}}} \end{aligned} \tag{3.39}$$

Since $(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})(\mathbf{D}' \otimes \mathbf{I}_4) = \mathbf{D}' \otimes \mathbf{R}_{\underline{\mathbf{r}}}$, the criterion (3.28) can be expressed as

$$\arg\max_{\hat{\mathbf{h}}} \frac{\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{D}' \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathbf{\Phi}\underline{\hat{\mathbf{h}}}}{||\hat{\mathbf{h}}||^2} \tag{3.40}$$

Observing this expression, we conclude that the channel can be obtained by computing the main eigenvector of the following matrix

$$\mathbf{C} = \mathbf{\Phi}^{\mathrm{T}}(\mathbf{D}' \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathbf{\Phi} \tag{3.41}$$

Finally, we introduce the operator $\mathcal{P}\{\cdot\}$ to denote the normalized principal eigenvector of a matrix ($||\mathcal{P}\{\cdot\}|| = 1$). Thus, we obtain the normalized solution to equation (3.40) which can be expressed as

$$\underline{\hat{\mathbf{h}}}_{\mathrm{opt}} = \mathcal{P}\{\mathbf{C}\} = \mathcal{P}\{\mathbf{\Phi}^{\mathrm{T}}(\mathbf{D}' \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathbf{\Phi}\} \tag{3.42}$$

Therefore, the algorithm to estimate the channel vector $\mathbf{h}$ from observations, after selecting an adequate precoding matrix $\mathbf{D}'$, can be summarized in the steps shown in Table 3.2. Note that the decoder needs to know the precoding matrix $\mathbf{D}'$ which has been applied to the sources before transmission.

- **Step 1**. Compose the real correlation of the receiving signals matrix $\mathbf{R}_{\underline{\mathbf{r}}} = \mathrm{E}[\underline{\mathbf{r}}\underline{\mathbf{r}}^{\mathrm{H}}]$.
- **Step 2**. Let $\mathbf{C} = \mathbf{\Phi}^{\mathrm{T}}(\mathbf{D}' \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathbf{\Phi}$.
- **Step 3**. Compute the normalized principal eigenvector of $\mathbf{C}$, i.e $\mathcal{P}\{\mathbf{C}\}$ and name it $\hat{\underline{\mathbf{h}}}$.
- **Step 4**. Return $\hat{\underline{\mathbf{h}}}$.

Table 3.2: The Shahbazpanahi et al. SOS algorithm

3.2.1 Simulation Results

We will consider the computer scenario described in Section 2.5. Blocks of 1000 symbols were generated from an equiprobable distribution whose symbols belong to a 4-QAM employing the (2×1) Alamouti code. The transmission was carried out through block fading channels with Rayleigh and Rice distribution. Bearing this in mind, we have tested the Shahbazpanahi et al. algorithm shown in Table 3.2 considering the following precoding matrix

$$\mathbf{D}_{\gamma} = \begin{bmatrix} \sigma_{z_1} & 0 \\ 0 & \sigma_{z_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{2}{1+\gamma^2}} & 0 \\ 0 & \gamma\sqrt{\frac{2}{1+\gamma^2}} \end{bmatrix} \tag{3.43}$$

giving γ the values 0.5, 0.6, 0.8 and 0.9. Note that thanks to the factor $\sqrt{\frac{2}{1+\gamma^2}}$, the mean variance between z_1 and z_2 is the same for every value of γ. Thus, with this set of values, we attain different precoding matrices which produce an unbalancing in the sources variance. The 4×4 matrix $\mathbf{D}'$ employed in Step 2 can be directly obtained from $\mathbf{D}_{\gamma}$ using equation (3.32).

We can see the results obtained for the Rayleigh channel distribution in Figure 3.1. It is shown that when the value of γ is 0.8 the performance is very close to Perfect CSI. For the remaining values of γ the performance is worse due to the high source variance unbalancing.

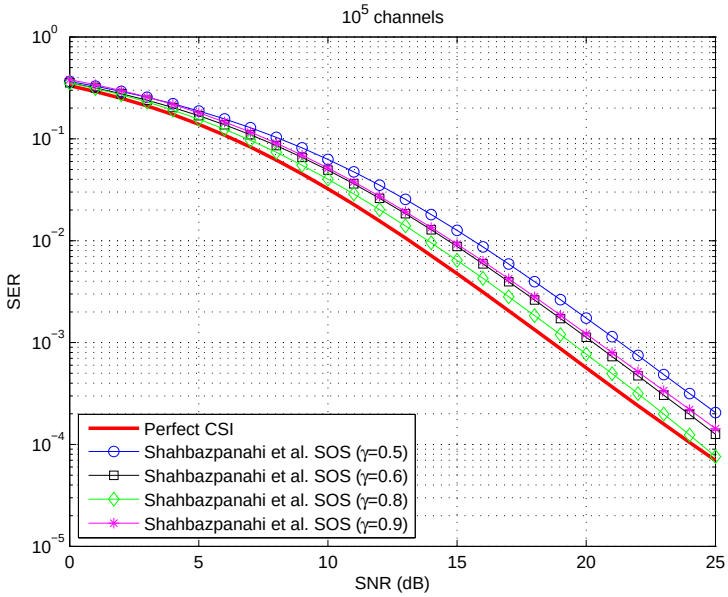


Figure 3.1: Performance of the Shahbazpanahi et al. SOS algorithm employing different precoding matrices with Rayleigh channel distribution: SER versus SNR

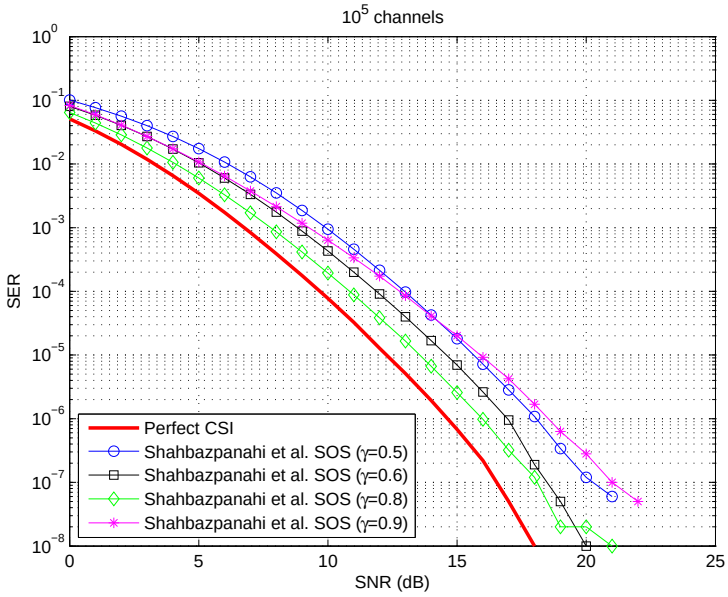


Figure 3.2: Performance of the Shahbazpanahi et al. SOS Algorithm employing different precoding matrices with Rice channel distribution: SER versus SNR

Figure 3.2 shows the results obtained for the transmission through Rice channel distribution. We can see that the performance for Perfect CSI in terms of SER is equal to 10^{-4} for 10 dB instead of 25 dB as in the case of Rayleigh channel distribution. The performance obtained for the Shahbazpanahi et al. algorithm with different values of γ is equivalent to that described for the Rayleigh case.

3.3 The Via et al. Adaptive Implementation

An important contribution of Via et al. in [4] is to formulate an adaptive algorithm to implement the ideas presented by Shahbazpanahi et al. [3]. The basis of this implementation is to find the vector $\hat{\mathbf{h}}$ that maximizes the equation (3.28),

$$J(\hat{\mathbf{h}}) = \frac{\mathrm{Tr}(\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\mathbf{R}_{\underline{\mathbf{r}}}\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \tag{3.44}$$

We will obviate the time instant n for reasons of simplicity, and considering that the unbalanced sources are $\underline{\mathbf{z}} = \mathbf{D}'\underline{\mathbf{s}}$, the received signals have the form

$$\underline{\mathbf{r}} = \underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{z}} + \underline{\mathbf{v}} \tag{3.45}$$

Thus, if we obviate the noise term $\mathbf{v}$, the cost function to be minimized could be rewritten as

$$\begin{aligned} J(\hat{\mathbf{h}}) &= \frac{\mathrm{Tr}(\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\mathrm{E}[\underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{z}}\underline{\mathbf{z}}^{\mathrm{T}}\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h})]\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \\ &= \frac{\mathrm{Tr}(\mathrm{E}[\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{z}}\underline{\mathbf{z}}^{\mathrm{T}}\underline{\mathbf{W}}^{\mathrm{T}}(\mathbf{h})\underline{\mathbf{W}}(\hat{\mathbf{h}})]\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \end{aligned} \tag{3.46}$$

Now, it is interesting to simplify the expression considering as the estimation of $\mathbf{z}$ the ML criterion shown in the following relationship

$$\hat{\underline{\mathbf{z}}} = \frac{\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\underline{\mathbf{r}}}{||\hat{\mathbf{h}}||||\mathbf{h}||} = \frac{\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\underline{\mathbf{W}}(\mathbf{h})\underline{\mathbf{z}}}{||\hat{\mathbf{h}}||||\mathbf{h}||} \tag{3.47}$$

And, introducing the term $\underline{\mathbf{W}}(\mathbf{h})$ inside the expectation, we obtain a new expression of the criterion to be maximized from equation (3.46), i.e.

$$\begin{aligned} J(\hat{\mathbf{h}}) &= \frac{\mathrm{Tr}(||\hat{\mathbf{h}}||||\mathbf{h}||\mathrm{E}[\underline{\hat{\mathbf{z}}}\underline{\hat{\mathbf{z}}}^{\mathrm{T}}]||\mathbf{h}||||\hat{\mathbf{h}}||\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \\ &= \frac{||\hat{\mathbf{h}}||^2||\mathbf{h}||^2\mathrm{Tr}(\mathrm{E}[\underline{\hat{\mathbf{z}}}\underline{\hat{\mathbf{z}}}^{\mathrm{T}}]\mathbf{D}')}{||\hat{\mathbf{h}}||^2} \\ &= ||\mathbf{h}||^2\mathrm{Tr}(\mathrm{E}[\mathbf{D}'\underline{\hat{\mathbf{z}}}\underline{\hat{\mathbf{z}}}^{\mathrm{T}}]) \\ &= ||\mathbf{h}||^2\mathrm{E}[\underline{\hat{\mathbf{z}}}^{\mathrm{T}}\mathbf{D}'\underline{\hat{\mathbf{z}}}] \end{aligned} \tag{3.48}$$

By considering $\mathbf{D}'$ to be a diagonal matrix, we have removed the trace by using $\mathrm{Tr}(\mathrm{E}[\mathbf{D}'\underline{\hat{\mathbf{z}}}\underline{\hat{\mathbf{z}}}^{\mathrm{T}}]) = \mathrm{E}[\underline{\hat{\mathbf{z}}}^{\mathrm{T}}\mathbf{D}'\underline{\hat{\mathbf{z}}}]$. Employing the matrices $\mathcal{D}_k$ defined in equation (3.31) and the equation (3.36), it is straightforward to obtain from equation (3.47) the following relationship

$$\underline{\hat{\mathbf{z}}} = \begin{bmatrix} \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_1 \\ \vdots \\ \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4 \end{bmatrix} \frac{\underline{\hat{\mathbf{h}}}}{||\hat{\mathbf{h}}||||\mathbf{h}||} \tag{3.49}$$

Substituting this expression of $\hat{\mathbf{z}}$ in the equation (3.48) a new expression is obtained for the maximization criterion shown in equation (3.44), i.e

$$\begin{aligned} J(\hat{\mathbf{h}}) = ||\mathbf{h}||^2\mathrm{E}[\underline{\hat{\mathbf{z}}}^{\mathrm{T}}\mathbf{D}'\underline{\hat{\mathbf{z}}}] &= \frac{||\mathbf{h}||^2}{||\hat{\mathbf{h}}||^2||\mathbf{h}||^2}\mathrm{E}\left[\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\begin{bmatrix} \underline{\mathcal{D}}_1^{\mathrm{T}}\underline{\mathbf{r}} & \cdots & \underline{\mathcal{D}}_4^{\mathrm{T}}\underline{\mathbf{r}} \end{bmatrix}\mathbf{D}'\begin{bmatrix} \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_1 \\ \vdots \\ \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4 \end{bmatrix}\underline{\hat{\mathbf{h}}}\right] \\ &= \frac{1}{||\hat{\mathbf{h}}||^2}\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\mathrm{E}\left[\begin{bmatrix} \underline{\mathcal{D}}_1^{\mathrm{T}}\underline{\mathbf{r}} & \cdots & \underline{\mathcal{D}}_4^{\mathrm{T}}\underline{\mathbf{r}} \end{bmatrix}\mathbf{D}'\begin{bmatrix} \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_1 \\ \vdots \\ \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4 \end{bmatrix}\right]\underline{\hat{\mathbf{h}}} \\ &= \frac{1}{||\hat{\mathbf{h}}||^2}\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\mathrm{E}\left[\sum_{k=1}^{4}\underline{\mathcal{D}}_k^{\mathrm{T}}d'_{kk}\underline{\mathbf{r}}\underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_k\right]\underline{\hat{\mathbf{h}}} \\ &= \frac{1}{||\hat{\mathbf{h}}||^2}\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\left(\sum_{k=1}^{4}\underline{\mathcal{D}}_k^{\mathrm{T}}d'_{kk}\mathrm{E}\left[\underline{\mathbf{r}}\underline{\mathbf{r}}^{\mathrm{T}}\right]\underline{\mathcal{D}}_k\right)\underline{\hat{\mathbf{h}}} \\ &= \frac{1}{||\hat{\mathbf{h}}||^2}\underline{\hat{\mathbf{h}}}^{\mathrm{T}}\left(\sum_{k=1}^{4}d'_{kk}\underline{\mathcal{D}}_k^{\mathrm{T}}\mathbf{R}_{\underline{\mathbf{r}}}\underline{\mathcal{D}}_k\right)\underline{\hat{\mathbf{h}}} \end{aligned} \tag{3.50}$$

The estimation problem is thus reduced to a PCA problem where the true channel is obtained, with scale and sign indeterminacy, making the extraction of the main eigenvector of the following correlation matrix

$$\sum_{k=1}^{4} d'_{kk} \underline{\mathcal{D}}_k^{\mathrm{T}} \mathbf{R}_{\underline{\mathbf{r}}} \underline{\mathcal{D}}_k \tag{3.51}$$

Applying Oja's rule [48], we obtain, for every k-th component of $\underline{\hat{z}}_k$ and for every n time instant, the following adaptive rule

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu d'_{kk} \underline{\mathcal{D}}_k^{\mathrm{T}} \underline{\mathbf{r}}(n) \underline{\hat{z}}_k(n) \tag{3.52}$$

where $\underline{\hat{z}}_k(n)$ is

$$\underline{\hat{z}}_k(n) = \underline{\mathbf{r}}^{\mathrm{T}}(n) \underline{\mathcal{D}}_k \hat{\underline{\mathbf{h}}}(n-1) \tag{3.53}$$

It is interesting to note that the equation (3.53) corresponds with the ML criterion. This will be shown in Section 3.3.1.

Now, considering the equations (3.52) and (3.53), assuming that there are L source vectors and L observations in the context of Alamouti code, the adaptive algorithm can be summarized in Table 3.3. It is important to note that by employing this algorithm the estimation of $\mathbf{h}$ is always normalized. However, thanks to the precoding matrix and this implementation, the estimation of sources is always non-scaled and non-permuted. Nevertheless, just as for any blind source estimation method, we will not know the phase of the channel.

3.3.1 Complex-valued Implementation

In this section, we present how to obtain the complex-valued representation of the Via et al. procedure shown in Table 3.3 in order to achieve a simpler implementation of this algorithm.

We will consider the relationship between the matrix $\mathcal{D}_k$ from $\mathcal{C}_k$ given in equation

- **Step 1**. Initialize the learning rate μ with a positive real value and $\hat{\underline{\mathbf{h}}}$ with random values. Let $n = 1$ and let $k = 1$.

- **Step 2**. Compute the estimation of the k-th unbalanced source for the time instant n

$$\hat{\underline{z}}_k(n) = \underline{\mathbf{r}}^{\mathrm{T}}(n)\underline{\mathcal{D}}_k\hat{\underline{\mathbf{h}}}$$

- **Step 3**. Update the estimated channel $\hat{\underline{\mathbf{h}}}$ employing

$$\hat{\underline{\mathbf{h}}} = \hat{\underline{\mathbf{h}}} + \mu d'_{kk}\underline{\mathcal{D}}_k^{\mathrm{T}}\underline{\mathbf{r}}(n)\hat{\underline{z}}_k(n)$$

  Note that d'_{kk} is the k-th element from the diagonal of $\mathbf{D'}$ from the precoding matrix shown in the equation (3.22).

- **Step 4**. Normalize the estimated channel

$$\hat{\underline{\mathbf{h}}} = \frac{\hat{\underline{\mathbf{h}}}}{||\hat{\underline{\mathbf{h}}}||}$$

- **Step 5**. Update the k-th estimated source

$$\hat{\underline{z}}_k(n) = \underline{\mathbf{r}}^{\mathrm{T}}(n)\underline{\mathcal{D}}_k\hat{\underline{\mathbf{h}}}$$

- **Step 6**. Let $k = k + 1$. If $k \leq 4$ then go back to step 2, otherwise continue.

- **Step 7**. Let $n = n + 1$. If $n \leq L$ then let $k = 1$ and go back to step 2. Otherwise finalize.

Table 3.3: The Via et al. adaptive SOS algorithm

(3.31). Replacing by the values of $\mathcal{C}_k$, we obtain

$$\underline{\mathcal{D}}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_4 \quad \underline{\mathcal{D}}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$
$$\underline{\mathcal{D}}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \qquad \underline{\mathcal{D}}_4 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.54}$$

Applying these matrices over the $\hat{\mathbf{h}}$ vector channel, the following expressions are obtained

$$\underline{\mathcal{D}}_1\underline{\hat{\mathbf{h}}} = \mathbf{I}_4\underline{\hat{\mathbf{h}}} = \underline{\hat{\mathbf{h}}} \quad \underline{\mathcal{D}}_2\underline{\hat{\mathbf{h}}} = \begin{bmatrix} \Re\{\hat{h_2}\} \\ -\Re\{\hat{h_1}\} \\ \Im\{\hat{h_2}\} \\ -\Im\{\hat{h_1}\} \end{bmatrix} \quad \underline{\mathcal{D}}_3\underline{\hat{\mathbf{h}}} = \begin{bmatrix} -\Im\{\hat{h_1}\} \\ \Im\{\hat{h_2}\} \\ \Re\{\hat{h_1}\} \\ -\Re\{\hat{h_2}\} \end{bmatrix} \quad \underline{\mathcal{D}}_4\underline{\hat{\mathbf{h}}} = \begin{bmatrix} -\Im\{\hat{h_2}\} \\ -\Im\{\hat{h_1}\} \\ \Re\{\hat{h_2}\} \\ \Re\{\hat{h_1}\} \end{bmatrix} \tag{3.55}$$

From now onwards, for simplicity reasons, we will remove the times index n. From the equation (3.55), we obtain the following expression

$$\begin{aligned} \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_1\underline{\hat{\mathbf{h}}} + \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_3\underline{\hat{\mathbf{h}}}j &= \Re\{r_1\}\Re\{\hat{h}_1\} + \Re\{r_2\}\Re\{\hat{h}_2\} + \Im\{r_1\}\Im\{\hat{h}_1\} + \Im\{r_2\}\Im\{\hat{h}_2\} + \\ &\quad \Big(-\Re\{r_1\}\Im\{\hat{h}_1\} + \Re\{r_2\}\Im\{\hat{h}_2\} + \Im\{r_1\}\Re\{\hat{h}_1\} - \Im\{r_2\}\Re\{\hat{h}_2\}\Big)j \\ &= \Re\{\hat{h}_1\}(\Re\{r_1\} + \Im\{r_1\}j) - \Im\{\hat{h}_1\}j(\Re\{r_1\} + \Im\{r_1\}j) + \\ &\quad \Re\{\hat{h}_2\}(\Re\{r_2\} + \Im\{r_2\}j) + \Im\{\hat{h}_2\}j(\Re\{r_2\} + \Im\{r_2\}j) \\ &= \hat{h}_1^* r_1 + \hat{h}_2 r_2^* \end{aligned} \tag{3.56}$$

where $\underline{\mathbf{r}}$ is $\underline{\mathbf{r}} = \left[\Re\{\mathbf{r}\}^{\mathrm{T}}\ \Im\{\mathbf{r}\}^{\mathrm{T}}\right]^{\mathrm{T}}$. Equivalently, for $\underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_2\underline{\hat{\mathbf{h}}}$ and $\underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4\underline{\hat{\mathbf{h}}}j$, we obtain

$$\underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_2\underline{\hat{\mathbf{h}}} + \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4\underline{\hat{\mathbf{h}}}j = \hat{h}_2^* r_1 - \hat{h}_1 r_2^* \tag{3.57}$$

Hence, stacking the terms obtained from equations (3.56) and (3.57) in an array, produces

the following expression

$$\begin{bmatrix} \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} + \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}}j \\ \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} + \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}}j \end{bmatrix} = \underbrace{\begin{bmatrix} \hat{h}_1^* & \hat{h}_2 \\ \hat{h}_2^* & -\hat{h}_1 \end{bmatrix}}_{\hat{\mathbf{H}}^{\mathrm{H}}} \underbrace{\begin{bmatrix} r_1 \\ r_2^* \end{bmatrix}}_{\mathbf{x}} = \hat{\mathbf{H}}^{\mathrm{H}}\mathbf{x} \tag{3.58}$$

which is equivalent to Step 2 of the Via et al. algorithm with $k = 1, 2, 3$ and 4, joining the real and the imaginary part as a complex number.

Note that this expression is equivalent to the ML criterion, as can be seen in the following expression

$$\begin{aligned} \hat{\mathbf{H}}^{\mathrm{H}}\mathbf{x} &= \begin{bmatrix} \hat{h}_1^* r_1 + \hat{h}_2 r_2^* \\ \hat{h}_2^* r_1 - \hat{h}_1 r_2^* \end{bmatrix} = \begin{bmatrix} \hat{h}_1^*(h_1 z_1 + h_2 z_2) + \hat{h}_2(h_2^* z_1 - h_1^* z_2) \\ \hat{h}_2^*(h_1 z_1 + h_2 z_2) - \hat{h}_1(h_2^* z_1 - h_1^* z_2) \end{bmatrix} \\ &\simeq \begin{bmatrix} |h_1|^2 z_1 + \hat{h}_1^* h_2 z_2 + |h_2|^2 z_1 - \hat{h}_2 h_1^* z_2 \\ \hat{h}_2^* h_1 z_1 + |h_2|^2 z_2 - \hat{h}_1 h_2^* z_1 + |h_1|^2 z_2 \end{bmatrix} = \begin{bmatrix} (|h_1|^2 + |h_2|^2) z_1 \\ (|h_2|^2 + |h_1|^2) z_2 \end{bmatrix} \\ &= ||\mathbf{h}||^2 \mathbf{z} \end{aligned} \tag{3.59}$$

Referring to Step 3 of the Via et al. algorithm, employing the values $k = 1, 2, 3$, and 4 and summing the obtained terms, it is possible to achieve the following expression

$$\sum_{k=1}^{4} d'_{kk}\underline{\mathcal{D}}_k^{\mathrm{T}}\underline{\mathbf{r}}\hat{\underline{z}}_k = \sum_{k=1}^{4} \underline{\mathcal{D}}_k^{\mathrm{T}}\underline{\mathbf{r}}\underbrace{d'_{kk}\hat{\underline{z}}_k}_{\hat{\underline{z}}'_k} = \sum_{k=1}^{4} \underline{\mathcal{D}}_k^{\mathrm{T}}\underline{\mathbf{r}}\hat{\underline{z}}'_k \tag{3.60}$$

Now, we can obtain from equations (3.14) and (3.54) the following relationships

$$\underline{\mathcal{D}}_1^{\mathrm{T}}\underline{\mathbf{r}} = \mathbf{I}_4\underline{\mathbf{r}} = \underline{\mathbf{r}} \qquad \underline{\mathcal{D}}_2^{\mathrm{T}}\underline{\mathbf{r}} = \begin{bmatrix} -\Re\{r_2\} \\ \Re\{r_1\} \\ -\Im\{r_2\} \\ \Im\{r_1\} \end{bmatrix} \qquad \underline{\mathcal{D}}_3^{\mathrm{T}}\underline{\mathbf{r}} = \begin{bmatrix} \Im\{r_1\} \\ -\Im\{r_2\} \\ -\Re\{r_1\} \\ \Re\{r_2\} \end{bmatrix} \qquad \underline{\mathcal{D}}_4^{\mathrm{T}}\underline{\mathbf{r}} = \begin{bmatrix} \Im\{r_2\} \\ \Im\{r_1\} \\ -\Re\{r_2\} \\ -\Re\{r_1\} \end{bmatrix} \tag{3.61}$$

Hence, substituting the expressions from equation (3.61) in equation (3.60), we obtain

the following vector

$$\underline{\mathbf{h}}^{+} = \sum_{k=1}^{4} \underline{\hat{z}}'_k \underline{\mathcal{D}}_k^{\mathrm{T}} \underline{\mathbf{r}} = \begin{bmatrix} \Re\{\hat{z}'_1\}\Re\{r_1\} - \Re\{\hat{z}'_2\}\Re\{r_2\} + \Im\{\hat{z}'_1\}\Im\{r_1\} + \Im\{\hat{z}'_2\}\Im\{r_2\} \\ \Re\{\hat{z}'_1\}\Re\{r_2\} + \Re\{\hat{z}'_2\}\Re\{r_1\} - \Im\{\hat{z}'_1\}\Im\{r_2\} + \Im\{\hat{z}'_2\}\Im\{r_1\} \\ \Re\{\hat{z}'_1\}\Im\{r_1\} - \Re\{\hat{z}'_2\}\Im\{r_2\} - \Im\{\hat{z}'_1\}\Re\{r_1\} - \Im\{\hat{z}'_2\}\Re\{r_2\} \\ \Re\{\hat{z}'_1\}\Im\{r_2\} + \Re\{\hat{z}'_2\}\Im\{r_1\} + \Im\{\hat{z}'_1\}\Re\{r_2\} - \Im\{\hat{z}'_2\}\Re\{r_1\} \end{bmatrix} \tag{3.62}$$

From the 1^{st} and 3^{rd} component of the equation (3.62), we obtain the 1^{st} component of $\mathbf{h}^{+}$, which is the complex version of $\underline{\mathbf{h}}^{+}$. It can be expressed as

$$\begin{aligned} h_1^{+} &= \underline{h}_1^{+} + \underline{h}_3^{+} j \\ &= \Re\{\hat{z}'_1\}(\Re\{r_1\} + \Im\{r_1\}j) - \Im\{\hat{z}'_1\}j(\Re\{r_1\} + \Im\{r_1\}j) + \\ &\quad -\Re\{\hat{z}'_2\}(\Re\{r_2\} + \Im\{r_2\}j) - \Im\{\hat{z}'_2\}j(\Re\{r_2\} + \Im\{r_2\}j) \\ &= \hat{z}'^{*}_1 r_1 - \hat{z}'_2 r_2 \end{aligned} \tag{3.63}$$

Equivalently, from the 2^{nd} and 4^{th} component of the equation (3.62), we obtain the 2^{nd} component of $\mathbf{h}^{+}$, i.e.

$$h_2^{+} = \underline{h}_2^{+} + \underline{h}_4^{+} j = \hat{z}'_1 r_2 - \hat{z}'^{*}_2 r_1 \tag{3.64}$$

Stacking these two terms, we obtain the following vector

$$\mathbf{h}^{+} = \begin{bmatrix} h_1^{+} \\ h_2^{+} \end{bmatrix} = \begin{bmatrix} \hat{z}'^{*}_1 & -\hat{z}'_2 \\ \hat{z}'^{*}_2 & \hat{z}'_1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{3.65}$$

Note that this is equivalent to Step 3 of the Via et. al Algorithm, but it is first necessary to compute the sources vector estimation $\hat{\mathbf{z}}'$ by means of the following expression

$$\hat{\mathbf{z}}' = \begin{bmatrix} \hat{z}'_1 \\ \hat{z}'_2 \end{bmatrix} = \mathbf{D} \begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \end{bmatrix} \tag{3.66}$$

The matrix $\mathbf{D}$ is the same introduced in equation (3.20) where the terms in its diagonal are related with the values of matrix $\mathbf{D}'$ used in the equation of the real-valued model (3.22) by $d_{11} = d'_{11} = d'_{33}$ and $d_{22} = d'_{22} = d'_{44}$.

In the complex-valued case, it is also needed to perform the normalization in Step 4. Finally, the estimated sources are updated using 2 again, i.e

$$\mathbf{z}(n) = \hat{\mathbf{H}}^{\mathrm{H}}\mathbf{x}(n) \tag{3.67}$$

Assuming that there are L source vectors and, therefore, L observations in the context of Alamouti code, the SOS complex-valued adaptive algorithm is composed of the steps shown in Table 3.4. It is important to note that this implementation is fully equivalent to the original implementation shown in Table 3.3.

3.3.2 Simulation Results

Considering the scenario described in Section 3.2.1 and taking into account the precoding matrix of equation (3.43) with γ equal to 0.5, 0.6, 0.8 and 0.9, we have performed several simulations using the Via et al. adaptive implementation presented in Table 3.3. Note that if we had employed the complex-valued implementation shown in Table 3.4, the performance achieved would have been the same because these implementations are equivalent. The learning step parameter employed is $\mu = 1$ for the Rayleigh case, and is $\mu = 0.1$ for the Rice case. This parameter has been chosen empirically in order to obtain the best performance by means of a dichotomy procedure.

We can see in Figure 3.3 the performance of the Via et al. adaptive SOS algorithm with different values of γ for Raleigh channel distribution. The case of $\gamma = 0.8$ obtains the best performance, similar to that of the batch Shahbazpanahi et al. algorithm shown in Figure 3.1, but there is a flooring effect when the SNRs are higher. Note that when the SNR is 25 dB the performance obtained for $\gamma = 0.8$ is a SER close to 10^{-3}, instead of around 10^{-4} for Perfect CSI. Hence, it is somewhat distant from the batch procedure with the same γ.

Figure 3.4 shows the performance when the Rice channel distribution is employed in transmission for the Via et al. adaptive SOS algorithm. Again, the case of $\gamma = 0.8$ obtains

- **Step 1**. Initialize the learning rate μ with a real value between 0 and 1, and randomly initialize the estimated channel vector $\hat{\mathbf{h}}$. Let $n = 1$.

- **Step 2**. Let

$$\hat{\mathbf{z}}(n) = \hat{\mathbf{H}}^{\mathrm{H}}\mathbf{x}(n)$$

  Note that the matrix $\hat{\mathbf{H}}$ is

$$\hat{\mathbf{H}} = \begin{bmatrix} \hat{\mathbf{h}}_1 & \hat{\mathbf{h}}_2 \\ \hat{\mathbf{h}}_2^* & -\hat{\mathbf{h}}_1^* \end{bmatrix}$$

- **Step 3**. Obtain a new sources vector estimation

$$\hat{\mathbf{z}}'(n) = \mathbf{D}\hat{\mathbf{z}}(n)$$

  Note that $\mathbf{D}$ is the precoding matrix (3.20) to unbalance the variance of the sources.

- **Step 4**. Update the estimated channel $\hat{\mathbf{h}}$ employing

$$\hat{\mathbf{h}} = \hat{\mathbf{h}} + \mu \begin{bmatrix} \hat{\mathbf{z}}_1'^*(n) & -\hat{\mathbf{z}}_2'(n) \\ \hat{\mathbf{z}}_2'^*(n) & \hat{\mathbf{z}}_1'(n) \end{bmatrix} \mathbf{r}(n)$$

- **Step 5**. Normalize the estimated channel

$$\hat{\mathbf{h}} = \frac{\hat{\mathbf{h}}}{||\hat{\mathbf{h}}||}$$

- **Step 6**. Update the estimated source vector of instant n

$$\mathbf{z}(n) = \hat{\mathbf{H}}^{\mathrm{H}}\mathbf{x}(n)$$

- **Step 7**. Let $n = n+1$. If $n \leq L$ go back to step 2. Otherwise, output the vector $\hat{\mathbf{h}}$.

Table 3.4: The complex-valued adaptive SOS algorithm

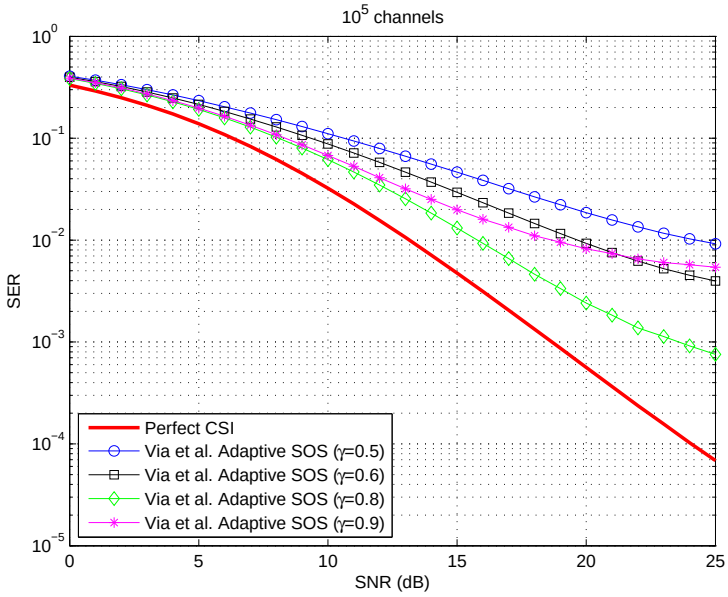


Figure 3.3: Performance of the Via et al. adaptive SOS algorithm employing different precoding matrices with Rayleigh channel distribution: SER versus SNR

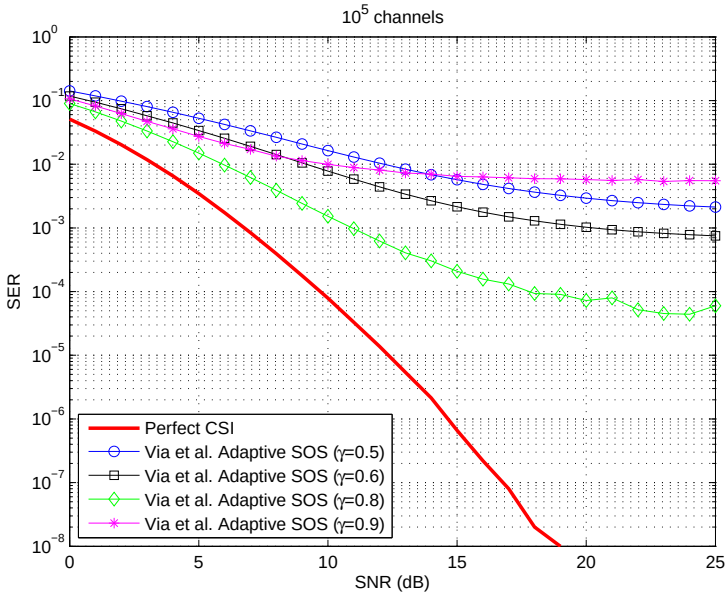


Figure 3.4: Performance of the Via et al. adaptive SOS algorithm employing different precoding matrices with Rice channel distribution: SER versus SNR

the best performance, as in the case of the batch Shahbazpanahi et al. algorithm shown in Figure 3.2. It also presents a high flooring effect for high SNRs.

3.4 Choosing the Optimal Precoding Matrix

The approaches presented in the previous sections make it possible to estimate OSTBC systems employing SOS through a batch or an adaptive algorithm. The key is to modify the encoder by adding a precoding matrix that produces an unbalancing of the sources variance. In this section, we will show how to choose the best precoding matrix from empirical results.

As we have explained above, the sources $\mathbf{s}$ have been unbalanced using a diagonal matrix $\mathbf{D}$, such as $\mathbf{z} = \mathbf{D}\mathbf{s}$. In order to keep the same variance as in the balanced case, this matrix takes the form

$$\mathbf{D} = \begin{bmatrix} \sigma_{z_1} & 0 \\ 0 & \sigma_{z_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{2}{1+\gamma^2}} & 0 \\ 0 & \gamma\sqrt{\frac{2}{1+\gamma^2}} \end{bmatrix} \tag{3.68}$$

where $\gamma^2 \in (0,1) \subset \mathbb{R}$ is the unbalancing parameter. The variance of the z_1 and z_2 are

$$\sigma_{z_1}^2 = \mathrm{E}[|z_1|^2] = \frac{2}{1+\gamma^2}\mathrm{E}[|s_1|^2] = \frac{2\sigma_s^2}{1+\gamma^2} \tag{3.69}$$

and

$$\sigma_{z_2}^2 = \mathrm{E}[|z_2|^2] = \gamma^2\frac{2}{1+\gamma^2}\mathrm{E}[|s_2|^2] = \gamma^2\mathrm{E}[|z_1|^2] = \gamma^2\frac{2\sigma_s^2}{1+\gamma^2} \tag{3.70}$$

Note that if $\gamma^2 = 0$, the source $\mathbf{s}_2$ is completely cancelled and it is totally balanced if $\gamma^2 = 1$, i.e. $\sigma_{z_1}^2 = \sigma_{z_2}^2$. The factor $\frac{2}{1+\gamma^2}$ is used to guarantee that the mean variance will be the same as in the balanced case. Performing the same reasoning as in the equation

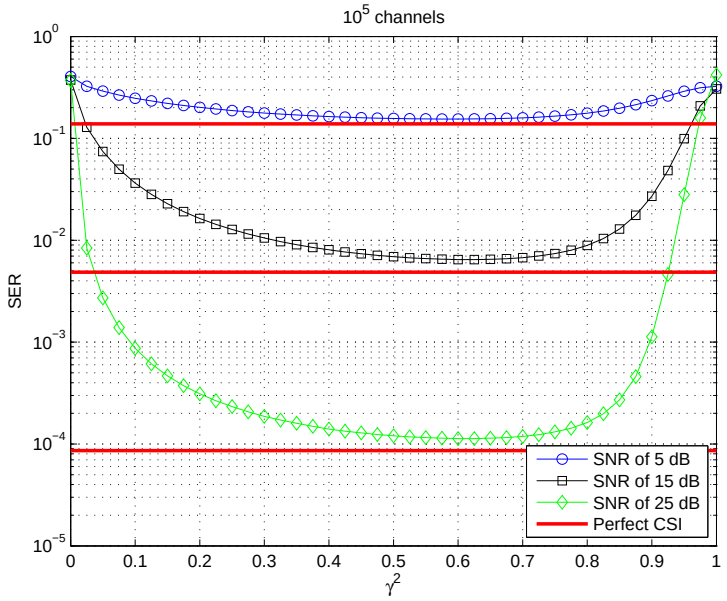


Figure 3.5: Finding the optimum γ^2 parameter for Rayleigh channel distribution: SER versus SNR

(3.2), it is straightforward to obtain the following expression

$$\mathbf{R_x} = \mathbf{H}\left(\mathbf{R_z} + \frac{\sigma_v^2}{||\mathbf{h}||^2}\mathbf{I}_2\right)\mathbf{H}^{\mathrm{H}} = \mathbf{H}\begin{bmatrix} \sigma_{z_1}^2 + \frac{\sigma_v^2}{||\mathbf{h}||^2} & 0 \\ 0 & \gamma^2\sigma_{z_1}^2 + \frac{\sigma_v^2}{||\mathbf{h}||^2} \end{bmatrix}\mathbf{H}^{\mathrm{H}} =$$
$$= \sigma_{z_1}^2\mathbf{H}\begin{bmatrix} 1 + \frac{\sigma_v^2}{\sigma_{z_1}^2||\mathbf{h}||^2} & 0 \\ 0 & \gamma^2 + \frac{\sigma_v^2}{\sigma_{z_1}^2||\mathbf{h}||^2} \end{bmatrix}\mathbf{H}^{\mathrm{H}} \quad (3.71)$$

As a result, the matrix $\mathbf{H}$ can be identified using an EVD.

In order to choose the best $\mathbf{D}$ precoding matrix, we will show the results of several computer simulations employing the general SOS algorithm shown at the beginning of this chapter in Table 3.1. Bearing this in mind, remember the scenario described in Section 2.5, where we have (2×1) Alamouti code, blocks of 1000 statistically independent symbols, identically distributed with a 4-QAM. Each block is rearranged as a block of 500 symbol vectors of dimension 2 and there is block fading. Also, the results have been averaged over 10^5 realizations.

Taking into account all these assumptions, Figure 3.5 and Figure 3.6 show, respectively,

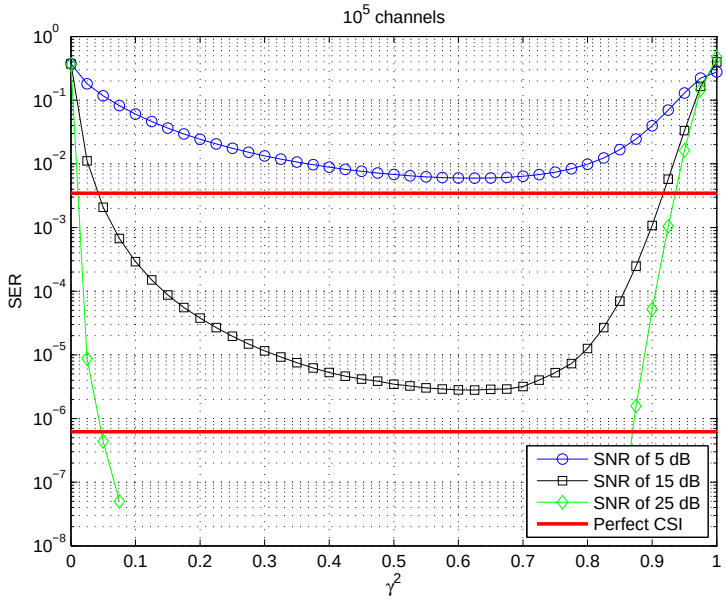


Figure 3.6: Finding the optimum γ^2 parameter for Rice channel distribution: SER versus SNR

the SER versus the γ^2 parameter for the Rayleigh and the Rice channel distributions for different values of SNR. We can empirically conclude that the optimum value for both the Rayleigh and the Rice distributions is the same, $\gamma^2 \simeq 0.64$. We can see that the performance is poor for the limits of the range of γ^2. This seems logical because in this case the unbalancing between the two sources is high. Performance is better for values of γ^2 close to the middle of the interval, which is the compromise between producing enough unbalancing to make channel identification possible and avoiding excessive degradation of one of the two sources. Note that these assertions are true for any value of SNR.

In Figure 3.7 and Figure 3.8 we can see, respectively, the performance in terms of SER versus SNR for Rayleigh and Rice channel distribution when employing the optimal value for $\gamma^2 = 0.64$. We show the performance for source z_1 and source z_2 separately, and the average performance. We can see that performance is close enough to the Perfect CSI: for source z_1 performance is better than the Perfect CSI because its variance is greater than 1, whilst for source z_2 performance does not quite reach the Perfect CSI because its variance is lower than 1.

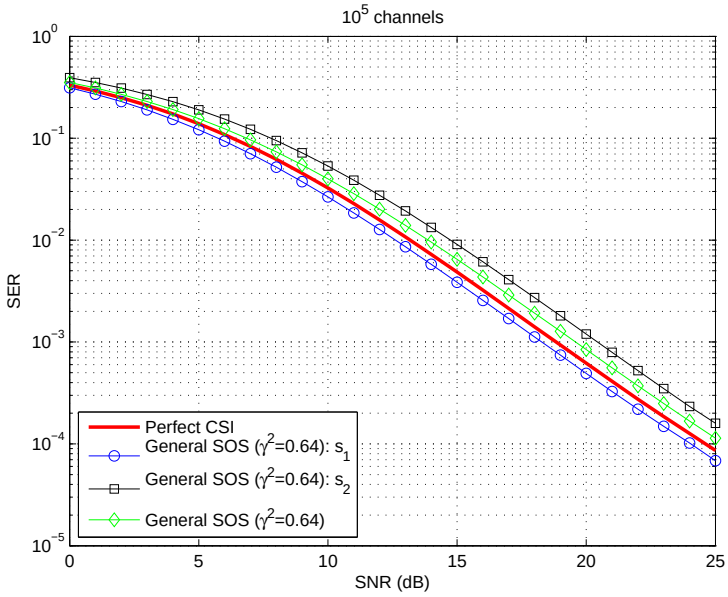


Figure 3.7: Performance of the general SOS algorithm employing the optimal encoding matrix with Rayleigh channel distribution: SER versus SNR

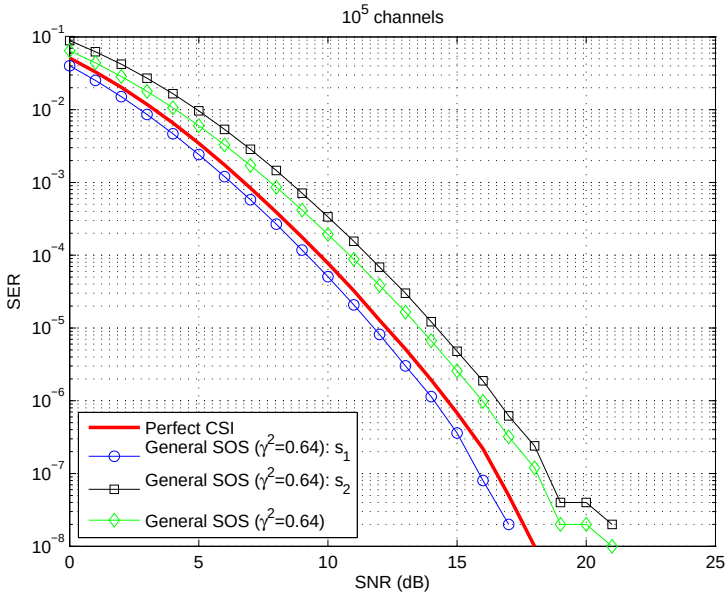


Figure 3.8: Performance of the general SOS algorithm employing the optimal encoding matrix with Rice channel distribution: SER versus SNR

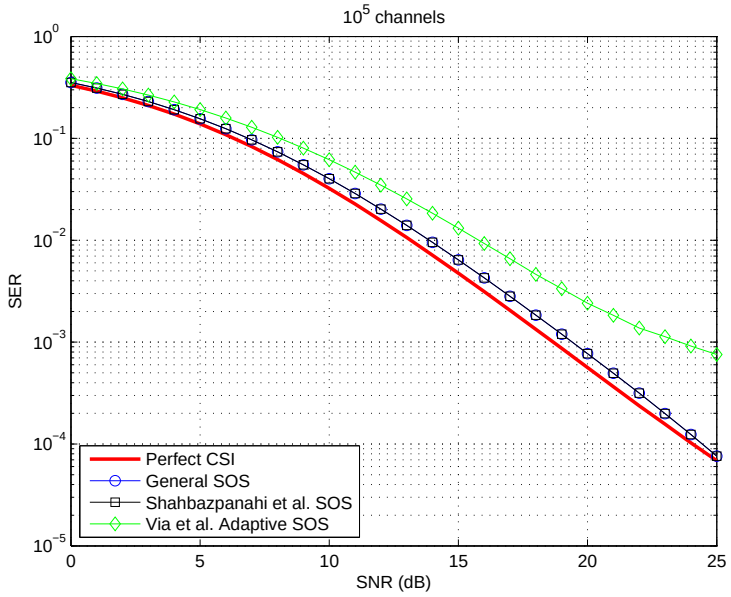


Figure 3.9: Comparison between SOS algorithms for Rayleigh channel distribution: SER versus SNR

Finally, Figures 3.9 and 3.10 compare the results obtained using the general SOS algorithm, the Shahbazpanahi et. al. approach and the Via et. al. method using the optimal encoding matrix. Note that the results of the general method match those obtained with the Shahbazpanahi algorithm, while the adaptive implementation presents a loss of performance.

The main advantage of the SOS based approaches presented above is the lower computation effort required to achieve the matrix $\mathbf{C}$ and the easy of implementation of this algorithm in DSPs or FPGAs. However, this approach needs to unbalance the variance between sources, which degrades the SER of one source.

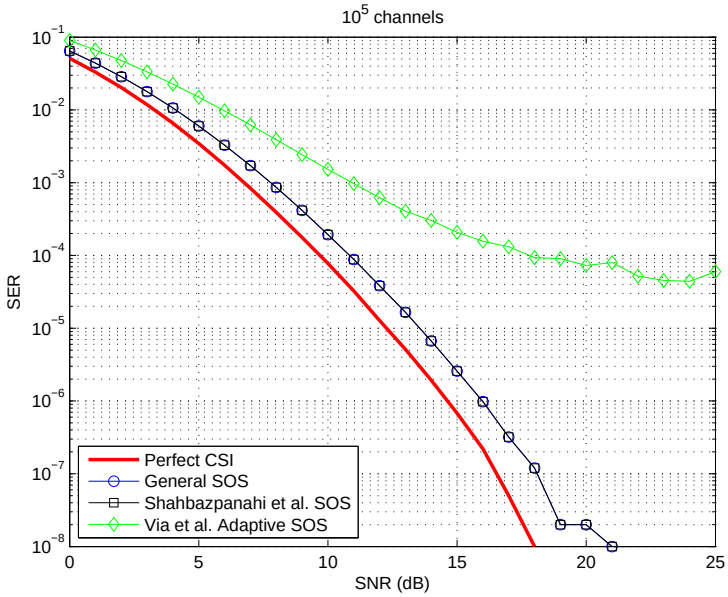


Figure 3.10: Comparison between SOS algorithms for Rice channel distribution: SER versus SNR

Chapter 4

HOS-based Approaches

In the previous chapter, we have shown that SOS can be used for the Alamouti coding system because the channel matrix is orthogonal but, however, it implies unbalancing the source variances. An undesirable consequence of this step is that the bit error for one source may be considerably increased.

In this chapter we will present blind channel identification approaches based on diagonalizing a linear combination or a sensible choice of 4^{th} order cumulant matrices, i.e. HOS-based approaches. For the particular case of a (2×1) Alamouti code, we have only 4 different matrices with the form

$$\begin{aligned} \mathbf{C}_{\text{HOS}}^{[k,l]} &= \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_k, x_l^*) \\ &= \begin{bmatrix} \text{cum}(x_1, x_1^*, x_k, x_l^*) & \text{cum}(x_1, x_2^*, x_k, x_l^*) \\ \text{cum}(x_2, x_1^*, x_k, x_l^*) & \text{cum}(x_2, x_2^*, x_k, x_l^*) \end{bmatrix}, \; k, l \in \{1, 2\} \end{aligned} \tag{4.1}$$

Due to the properties of cumulants, it is shown, in Appendix E.3, that there are only 6 different 4^{th} order cumulants.

The main idea of the techniques studied in this chapter is to consider that the cumulant matrices can be written as follows

$$\mathbf{C}_{\text{HOS}}^{[k,l]} = \mathbf{H} \breve{\mathbf{\Delta}}_{kl} \mathbf{H}^{\text{H}} \tag{4.2}$$

- **Step 1**. Obtain a matrix $\mathbf{C}$ depending on the chosen approach.
- **Step 2**. Make the EVD of the matrix $\mathbf{C}$. The result is $\mathbf{C} = \hat{\mathbf{H}}\boldsymbol{\Delta}_{\text{HOS}}\hat{\mathbf{H}}^{\text{H}}$, where $\hat{\mathbf{U}}$ is the matrix that contains the eigenvectors of $\mathbf{C}$, and $\boldsymbol{\Delta}_{\text{HOS}}$ is the diagonal matrix which contains the eigenvalues of $\mathbf{C}$.
- **Step 3**. Return an estimation of the channel matrix, $\hat{\mathbf{H}}$.

Table 4.1: General HOS Algorithm: ICA

where $\breve{\boldsymbol{\Delta}}_{kl}$ is a diagonal matrix. As a consequence, the channel matrix can be estimated by computing the eigenvectors of $\mathbf{C}_{\text{HOS}}^{[k,l]}$. Note that it is also possible to estimate $\mathbf{H}$ from the linear combination of the cumulant matrices because the matrix $\mathbf{H}$ in equation (4.2) is the same for all $\mathbf{C}_{\text{HOS}}^{[k,l]}$. Call any linear combination of these matrices $\mathbf{C}$. Thus, the eigenvalue decomposition will be $\mathbf{C} = \hat{\mathbf{H}}\boldsymbol{\Delta}_{\text{HOS}}\hat{\mathbf{H}}^{\text{H}}$ where $\boldsymbol{\Delta}_{\text{HOS}}$ is a diagonal matrix which contains the eigenvalues and $\hat{\mathbf{H}}$ is the matrix where each column is an eigenvector of $\mathbf{C}$ in any order, and note that they are normalized. Taking this into account, we can conclude that we will always obtain a normalized estimation $\hat{\mathbf{H}}$ of the matrix channel. Therefore, we have a permutation and scale indeterminacy. Table 4.1 describes the general algorithm for the approaches that employ HOS for estimating the channel matrix.

This chapter is structured as follows. Section 4.1 describes the approach proposed by Beres and Adve in [6]. Subsequently, Section 4.2 shows how to improve the performance of this approach by using a linear combination of the cumulant matrices used independently by Beres and Adve. Section 4.3 studies the performance when using a different selection of the cumulant matrix. Section 4.4 contains the main contribution of this chapter, which is to present an analytical way of determining the optimum cumulant matrix to be diagonalized. We will also propose a suboptimal technique which presents a good performance with a low computational cost. Finally, Section 4.5 particularizes the well-known *Joint Approximate Diagonalization of Eigenmatrices* (JADE) to the (2×1) Alamouti code.

4.1 The Beres and Adve Approach

Beres and Adve have presented in [6] a blind channel estimation method for OSTBC in MISO systems. These authors consider the diagonalization of 4^{th} order cumulants matrices obtained by considering $k = l$ in equation (4.1) and, hence, we have only two matrices: $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[2,2]}$. In Appendix E, we show that these matrices can be decomposed as follows

$$\mathbf{C}_{\text{HOS}}^{[k,k]} = \mathbf{H}\breve{\mathbf{\Delta}}_{kk}\mathbf{H}^{\text{H}} \tag{4.3}$$

where

$$\breve{\mathbf{\Delta}}_{11} = \rho \begin{bmatrix} |h_1|^2 & 0 \\ 0 & |h_2|^2 \end{bmatrix}, \ \breve{\mathbf{\Delta}}_{22} = \rho \begin{bmatrix} |h_2|^2 & 0 \\ 0 & |h_1|^2 \end{bmatrix} \tag{4.4}$$

Taking this into account, we can employ the common algorithm shown in Table 4.1 to obtain an estimation of the matrix $\mathbf{H}$, choosing as matrix $\mathbf{C}$ between the matrices $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[2,2]}$. Note that the diagonalization fails when the entries in $\breve{\mathbf{\Delta}}_{11}$ (or $\breve{\mathbf{\Delta}}_{22}$) are close, i.e. when the modules of the channel realization are similar $|h_1| \simeq |h_2|$.

It is interesting to note that this approach needs to compute a reduced number of 4^{th} order cumulant matrices. In fact, using the definitions and properties presented in Appendix E, these matrices can be computed as follows

$$\mathbf{C}_{\text{HOS}}^{[1,1]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_1, x_1^*) = \begin{bmatrix} \text{cum}(x_1, x_1^*, x_1, x_1^*) & \text{cum}(x_1, x_2^*, x_1, x_1^*) \\ \text{cum}^*(x_1, x_2^*, x_1, x_1^*) & \text{cum}(x_2, x_2^*, x_1, x_1^*) \end{bmatrix} \tag{4.5}$$

$$\mathbf{C}_{\text{HOS}}^{[2,2]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_2, x_2^*) = \begin{bmatrix} \text{cum}(x_1, x_1^*, x_2, x_2^*) & \text{cum}(x_1, x_2^*, x_2, x_2^*) \\ \text{cum}^*(x_1, x_2^*, x_2, x_2^*) & \text{cum}(x_2, x_2^*, x_2, x_2^*) \end{bmatrix} \tag{4.6}$$

We can see that each matrix needs to compute only 3 different cumulants. In practice, these 4^{th} order cumulants matrices are computed by sample averaging on the observations.

4.1.1 Simulation Results

In order to illustrate the performance of this approach, we will show the results of several computer simulations using the scenario described in Section 2.5. Remember that blocks

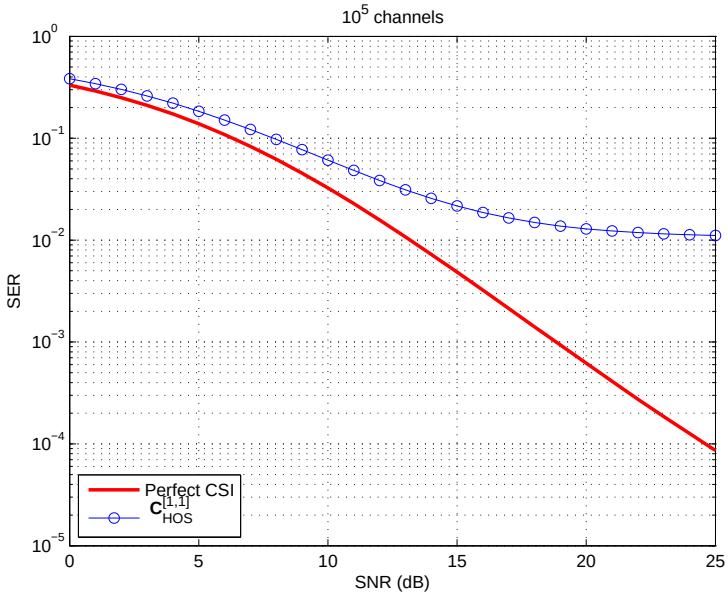


Figure 4.1: Performance of the Beres and Adve HOS algorithm employing $\mathbf{C}_{\mathrm{HOS}}^{[1,1]}$ (or $\mathbf{C}_{\mathrm{HOS}}^{[2,2]}$) with Rayleigh channel distribution: SER versus SNR

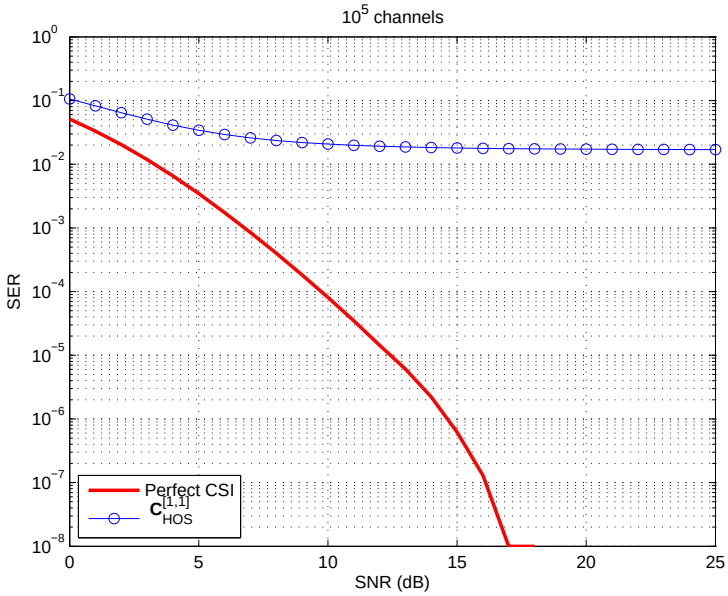


Figure 4.2: Performance of the Beres and Adve HOS algorithm employing $\mathbf{C}_{\mathrm{HOS}}^{[1,1]}$ (or $\mathbf{C}_{\mathrm{HOS}}^{[2,2]}$) with Rice channel distribution: SER versus SNR

of 1000 symbols have been generated from an equiprobable distribution whose symbols belong to a 4-QAM and employing the (2×1) Alamouti code. Hence, each block generated is rearranged as a block of 500 symbol vectors of dimension 2, where each component is a source, and thus we have 500 realizations for each source. The transmission of the symbols is carried out through block fading channels following a Rayleigh or a Rice distribution and the results have been averaged over 10^5 realizations.

Considering these parameters of simulation, Figures 4.1 and 4.2 show the performance in terms of SER versus SNR for Rayleigh and Rice distributions, respectively. We can see, for both distributions, that as SNR values increase, performance deteriorates and, in general, is very far from the Perfect CSI. The reason of this undesirable situation is that for some experiments the channel realizations have similar modules, $|h_1|^2 \simeq |h_2|^2$ and hence the estimation of the channel matrix (and of the source symbols) is inadequate.

4.2 The Beres and Adve Approach Improvement I: Linear Combination of Cumulants

A direct question arises from the approach proposed by Beres and Adve shown in Section 4.1: Why use only $\mathbf{C}_{\text{HOS}}^{[1,1]}$ or $\mathbf{C}_{\text{HOS}}^{[2,2]}$? We propose to employ a simple linear combination from the two matrices $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[2,2]}$,

$$\mathbf{C}_{\text{HOS}}^{\lambda} = \mathbf{C}_{\text{HOS}}^{[1,1]} + \lambda \mathbf{C}_{\text{HOS}}^{[2,2]} \tag{4.7}$$

where $\lambda \in (0,1) \subset \mathbb{R}$. Since $\mathbf{C}_{\text{HOS}}^{\lambda}$ is an addition of two 4^{th} order cumulant matrices, it admits the following decomposition

$$\mathbf{C}_{\text{HOS}}^{\lambda} = \mathbf{C}_{\text{HOS}}^{[1,1]} + \lambda \mathbf{C}_{\text{HOS}}^{[2,2]} = \mathbf{H}\breve{\mathbf{\Delta}}_{1,1}\mathbf{H}^{\text{H}} + \lambda \mathbf{H}\breve{\mathbf{\Delta}}_{2,2}\mathbf{H}^{\text{H}} = \mathbf{H}\breve{\mathbf{\Delta}}_{\lambda}\mathbf{H}^{\text{H}} \tag{4.8}$$

where $\breve{\mathbf{\Delta}}_{1,1}$ and $\breve{\mathbf{\Delta}}_{2,2}$ have been defined in equation (4.4) and $\breve{\mathbf{\Delta}}_{\lambda}$ is the diagonal matrix which contains scaled eigenvalues given by

$$\breve{\mathbf{\Delta}}_{\lambda} = \rho \begin{bmatrix} |h_1|^2 + \lambda |h_2|^2 & 0 \\ 0 & |h_2|^2 + \lambda |h_1|^2 \end{bmatrix} \tag{4.9}$$

It is important to note that the channel cannot be identified when $\lambda = 1$ since the two entries in $\breve{\mathbf{\Delta}}_{\lambda}$ have the same value and hence

$$\begin{aligned} \mathbf{C}_{\text{HOS}}^{\lambda=1} &= \mathbf{C}_{\text{HOS}}^{[1,1]} + \mathbf{C}_{\text{HOS}}^{[2,2]} \\ &= \rho\, \mathbf{H} \begin{bmatrix} |h_1|^2 + |h_2|^2 & 0 \\ 0 & |h_2|^2 + |h_1|^2 \end{bmatrix} \mathbf{H}^{\text{H}} \\ &= \rho\, \mathbf{H}\, (|h_1|^2 + |h_2|^2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{H}^{\text{H}} \\ &= \rho\, (|h_1|^2 + |h_2|^2)\, \mathbf{H}\, \mathbf{H}^{\text{H}} = \rho\, (|h_1|^2 + |h_2|^2)\, \mathbf{I}_2 \end{aligned} \tag{4.10}$$

Thus, the two eigenvalues of the matrix $\mathbf{C}_{\text{HOS}}^{\lambda=1}$ are the same and it is not possible to achieve them by means of a traditional EVD. On the contrary, when $\lambda = -1$, the eigenvalues will be opposite, as we can see in the development of the following equation

$$\begin{aligned} \mathbf{C}_{\text{HOS}}^{\lambda=-1} &= \mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]} \\ &= \rho\, \mathbf{H} \begin{bmatrix} |h_1|^2 - |h_2|^2 & 0 \\ 0 & |h_2|^2 - |h_1|^2 \end{bmatrix} \mathbf{H}^{\text{H}} \\ &= \rho\, \mathbf{H}\, (|h_1|^2 - |h_2|^2) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{H}^{\text{H}} \end{aligned} \tag{4.11}$$

This point having been made, the common diagonalization algorithm shown in Table 4.1 with $\mathbf{C} = \mathbf{C}_{\lambda}$ can be used to achieve the channel matrix $\mathbf{H}$. Referring to the computation of the matrix $\mathbf{C}_{\lambda}$, note that only 5 different cumulants need to be computed, as can be seen in equations (4.5) and (4.6).

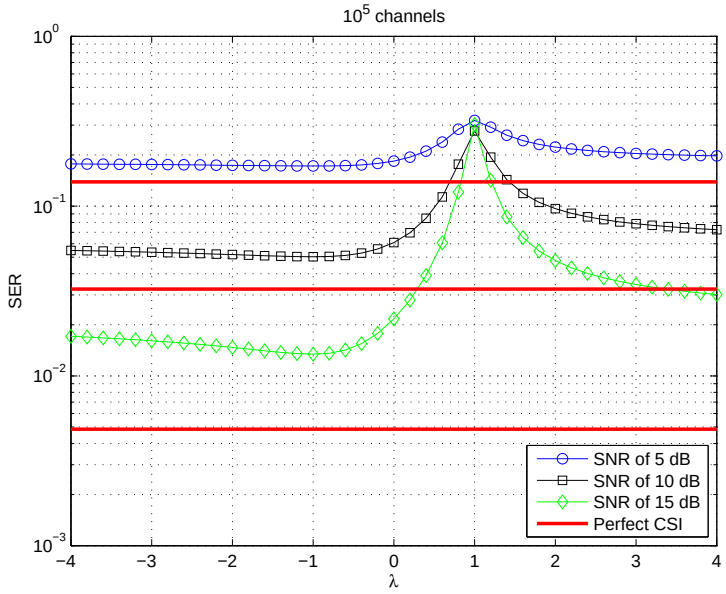


Figure 4.3: Finding the optimum λ parameter for Rayleigh channel distribution: SER versus SNR

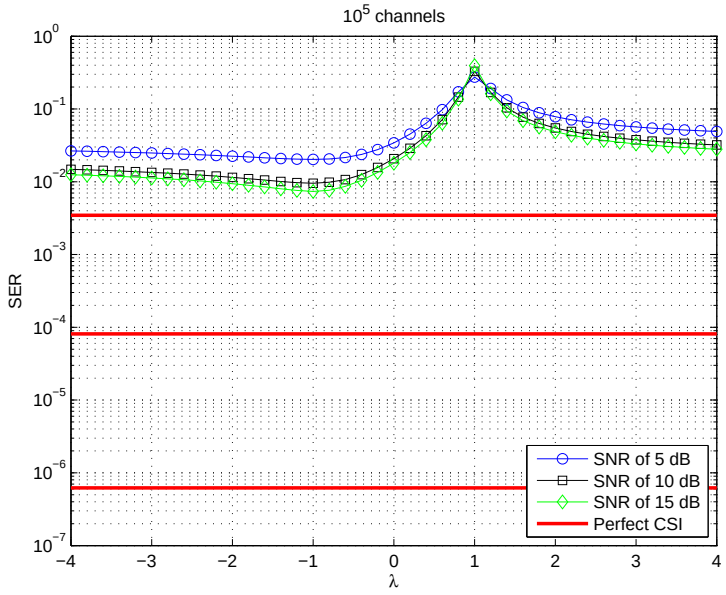


Figure 4.4: Finding the optimum λ parameter for Rice channel distribution: SER versus SNR

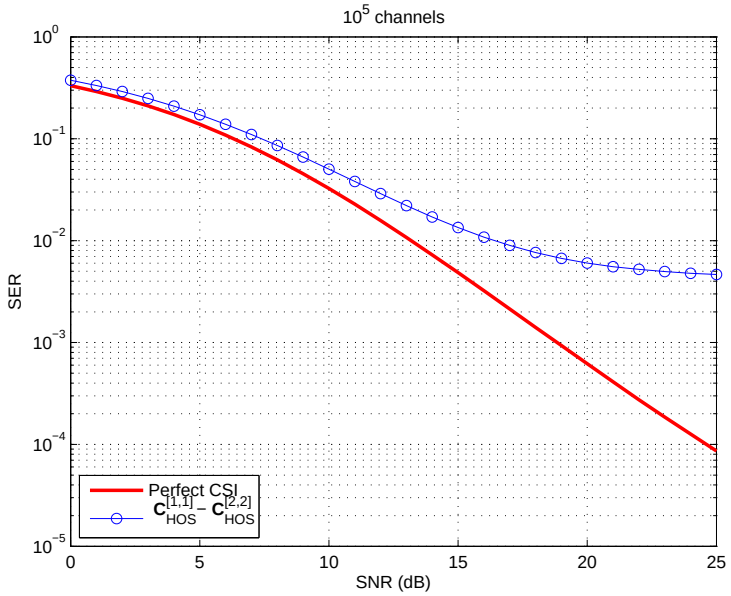


Figure 4.5: Performance of the Beres and Adve improvement I HOS algorithm with matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,1]} - \mathbf{C}_{\mathrm{HOS}}^{[2,2]}$ for Rayleigh channel distribution: SER versus SNR

4.2.1 Simulation Results

In order to show the importance of the parameter λ, employing the same scenario as in Section 4.1.1, Figures 4.3 and 4.4 show the SER versus the parameter λ for the Rayleigh and the Rice channel distributions, respectively. Each curve represents the performance for a different value of SNR. We can conclude that the optimum value for the Rayleigh and the Rice distribution is $\lambda = -1$ and that performance is better for values around the latter figure. This is because the difference is maximum between the eigenvalues of the matrix $\mathbf{C}_{\mathrm{HOS}}^{\lambda}$ for this value of λ. On the contrary, when λ is close to 1 the performance is poor because the eigenvalues of $\mathbf{C}_{\mathrm{HOS}}^{\lambda}$ are very similar and the EVD will therefore be inaccurate.

Figures 4.5 and 4.6 show the performance in terms of SER versus SNR for the optimum value of λ. We can see that, although the performance is better than for the original Beres and Adve approach, when the values of SNR are high the curve nevertheless continues to present a high flooring effect.

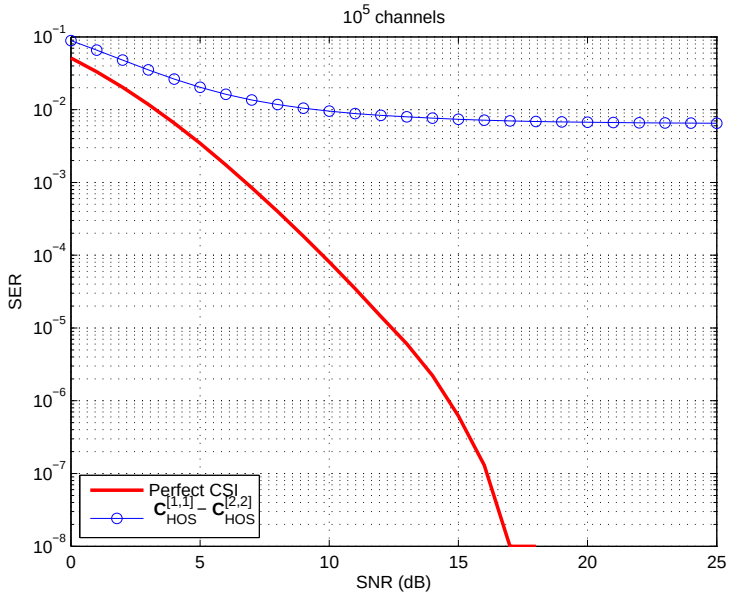


Figure 4.6: Performance of the Beres and Adve improvement I HOS algorithm with Matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,1]} - \mathbf{C}_{\mathrm{HOS}}^{[2,2]}$ for Rice channel distribution: SER versus SNR

4.3 The Beres and Adve Approach Improvement II: Cross-slot Cumulants

It is interesting to note that also it is possible to use the matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,2]}$ (or $\mathbf{C}_{\mathrm{HOS}}^{[2,1]}$) instead of $\mathbf{C}_{\mathrm{HOS}}^{[1,1]}$ (or $\mathbf{C}_{\mathrm{HOS}}^{[2,2]}$) in order to achieve a channel estimation. In Appendix E we show that $\mathbf{C}_{\mathrm{HOS}}^{[1,2]}$ admits the following decomposition

$$\begin{aligned}
\mathbf{C}_{\mathrm{HOS}}^{[1,2]} &= \mathrm{cum}(\mathbf{x}, \mathbf{x}^{\mathrm{H}}, x_1, x_2^*) \\
&= \begin{bmatrix} \mathrm{cum}(x_1, x_1^*, x_1, x_2^*) & \mathrm{cum}(x_1, x_2^*, x_1, x_2^*) \\ \mathrm{cum}(x_2, x_1^*, x_1, x_2^*) & \mathrm{cum}(x_2, x_2^*, x_1, x_2^*) \end{bmatrix} \\
&= \rho\, \mathbf{H} \begin{bmatrix} h_1 h_2 & 0 \\ 0 & -h_1 h_2 \end{bmatrix} \mathbf{H}^{\mathrm{H}}
\end{aligned} \tag{4.12}$$

and $\mathbf{C}_{\text{HOS}}^{[2,1]}$

$$\begin{aligned}\mathbf{C}_{\text{HOS}}^{[2,1]} &= \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_1, x_2^*) \\ &= \begin{bmatrix} \text{cum}(x_1, x_1^*, x_2, x_1^*) & \text{cum}(x_1, x_2^*, x_2, x_1^*) \\ \text{cum}(x_2, x_1^*, x_2, x_1^*) & \text{cum}(x_2, x_2^*, x_2, x_1^*) \end{bmatrix} \\ &= \rho\, \mathbf{H} \begin{bmatrix} h_1^* h_2^* & 0 \\ 0 & -h_1^* h_2^* \end{bmatrix} \mathbf{H}^{\text{H}}\end{aligned} \tag{4.13}$$

Unlike the matrices $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[2,2]}$, in this case the set of 4 cumulants in each matrix $\mathbf{C}_{\text{HOS}}^{[1,2]}$ and $\mathbf{C}_{\text{HOS}}^{[2,1]}$ cannot be reduced using the cumulant properties. Note that the EVD fails when $h_1 h_2 \simeq 0$, which is a less restrictive condition, in the context of a Rayleigh or a Rice channel distribution, than in the case $|h_1|^2 \simeq |h_2|^2$, which corresponds to the problem detected in the methods presented in Section 4.1 and Section 4.2.

Again, the algorithm is that described at the beginning of the chapter in Table 4.1, with $\mathbf{C}$ being the matrix $\mathbf{C}_{\text{HOS}}^{[1,2]}$ or $\mathbf{C}_{\text{HOS}}^{[2,1]}$.

4.3.1 Simulation Results

Figure 4.7 and Figure 4.8 show the SER versus SNR for improvement II. We can see that the SER is close to Perfect CSI for values lower than 15 dB in the Rayleigh distribution and is close to Perfect CSI for values lower than 11 dB in the Rice distribution. Unfortunately, this approach presents a floor effect at high values of SNR, although it achieves better performance than the methods shown previously.

4.4 Maximum Eigenvalue Spread Approach

The performance of EVD based methods depends on the distance between the eigenvalues of the matrix to be diagonalized because, as has already been mentioned, the eigenvectors associated with equal eigenvalues cannot be determined up to a unitary transformation

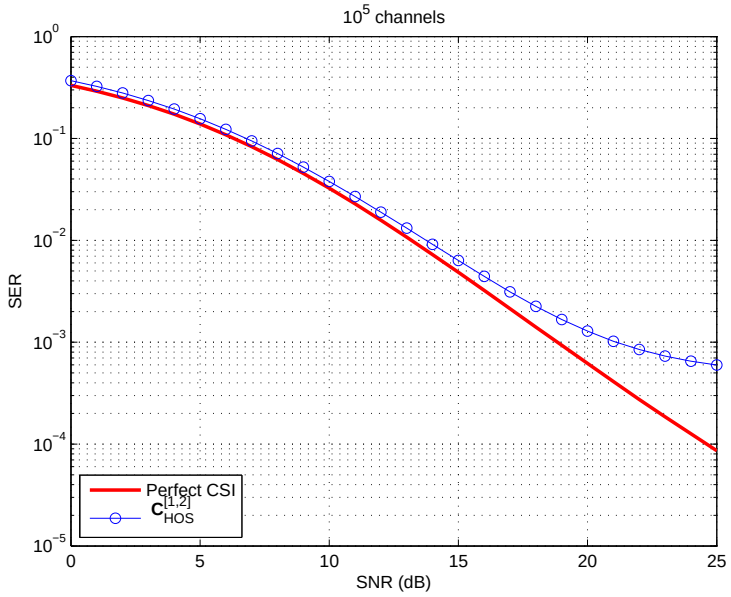


Figure 4.7: Performance of the Beres and Adve improvement II HOS algorithm with matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,2]}$ for Rayleigh channel distribution: SER versus SNR

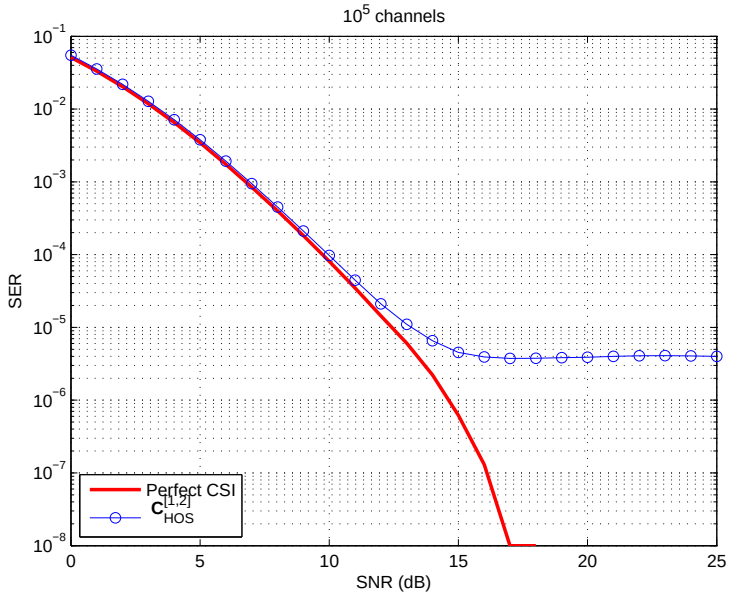


Figure 4.8: Performance of the Beres and Adve improvement II HOS algorithm with matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,2]}$ for Rice channel distribution: SER versus SNR

[29]. In this section, we focus on the problem of determining the optimum linear combination of cumulant matrices in order to guarantee a sound conditioning of the matrix to be diagonalized.

Any matrix $\mathbf{C}$ of dimension 2×2 can be decomposed as $\mathbf{C} = \mathbf{U}\mathbf{\Delta}\mathbf{U}^{-1}$ where $\mathbf{\Delta} = \text{diag}(\delta_1, \delta_2)$ contains the two eigenvalues of $\mathbf{C}$. Then, we can define the term **Eigenvalue Spread** as the absolute difference between the two eigenvalues: $L = |\delta_1 - \delta_2|$. In order to work with this term, we introduce a more convenient notation to define the cumulant matrices. Given a matrix $\mathbf{M} \in \mathbb{C}^{2\times 2}$, the 4^{th} order cumulant matrix $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ is defined in [5] as the 2×2 matrix with components

$$[\mathbf{Q}_{\mathbf{x}}(\mathbf{M})]_{ij} = \sum_{k,l=1}^{2} \text{cum}(x_i, x_j^*, x_k, x_l^*) m_{lk}, \ 1 \leq i, j \leq 2 \tag{4.14}$$

where $m_{lk} = [\mathbf{M}]_{lk}$. In Appendix E, it is shown that we can rewrite equation (4.14) as

$$\mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = m_{11}\mathbf{C}_{\text{HOS}}^{[1,1]} + m_{21}\mathbf{C}_{\text{HOS}}^{[1,2]} + m_{12}\mathbf{C}_{\text{HOS}}^{[2,1]} + m_{22}\mathbf{C}_{\text{HOS}}^{[2,2]} \tag{4.15}$$

Also, under a linear model such as the equation of the (2×1) Alamouti code (equation (2.5)) with statistically independent sources and orthogonal mixing matrix[1], in Appendix E we show that the cumulant matrix takes the form

$$\mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{H}\breve{\mathbf{\Delta}}_{\mathbf{M}}\mathbf{H}^{\text{H}} = \mathbf{U}\mathbf{\Delta}_{\mathbf{M}}\mathbf{U}^{\text{H}} \tag{4.16}$$

where the matrix $\breve{\mathbf{\Delta}}_{\mathbf{M}}$ is diagonal with

$$[\breve{\mathbf{\Delta}}_{\mathbf{M}}]_{ii} = \rho \mathbf{h}_i^{\text{H}} \mathbf{M} \mathbf{h}_i \tag{4.17}$$

This notation allows us to make any linear combination with the set of the 4 matrices shown in the equation (4.1). For this reason, by choosing the appropriate matrix $\mathbf{M}$ we can compute the methods shown in Sections 4.1, 4.2 and 4.3. Also, note that the EVD of equation (4.16) obtains an eigenvector orthogonal matrix $\mathbf{U}$ with the form

$$\mathbf{U} = \frac{\mathbf{H}}{||\mathbf{h}||} = \frac{1}{||\mathbf{h}||}[\mathbf{h}_1 \ \mathbf{h}_2] = \frac{1}{||\mathbf{h}||}\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \tag{4.18}$$

[1]In fact, expression (4.16) is valid for any mixing matrix $\mathbf{H}$, even if it is not orthogonal.

Thus, according to the equation (4.16), the diagonal eigenvalue matrix will be

$$\mathbf{\Delta_M} = ||\mathbf{h}||^2 \breve{\mathbf{\Delta}}_\mathbf{M} \tag{4.19}$$

Taking these assumptions into account, we will subsequently consider the following definition of the eigenvalue spread depending on $\mathbf{M}$ matrix

$$L(\mathbf{M}) = |[\mathbf{\Delta_M}]_{11} - [\mathbf{\Delta_M}]_{22}| = ||\mathbf{h}||^2 \left| [\breve{\mathbf{\Delta}}_\mathbf{M}]_{11} - [\breve{\mathbf{\Delta}}_\mathbf{M}]_{22} \right| \tag{4.20}$$

It is important to note that the eigendecomposition of the matrix from equation (4.16) allows the remaining orthogonal part of $\mathbf{H}$ to be identified if the eigenvalues of $\mathbf{Q_x}(\mathbf{M})$ are different, i.e. if the matrix $\mathbf{\Delta_M}$ contains different entries: $||\mathbf{h}||^2 \rho \mathbf{h}_1^{\mathrm{H}} \mathbf{M} \mathbf{h}_1 \neq ||\mathbf{h}||^2 \rho \mathbf{h}_2^{\mathrm{H}} \mathbf{M} \mathbf{h}_2$. In particular, substituting equation (4.17) in equation (4.20), the eigenvalue spread of the cumulant matrix $\mathbf{Q_x}(\mathbf{M})$ for the (2×1) Alamouti code is given by

$$L(\mathbf{M}) = ||\mathbf{h}||^2 |\rho| ||\mathbf{h}_1^{\mathrm{H}} \mathbf{M} \mathbf{h}_1 - \mathbf{h}_2^{\mathrm{H}} \mathbf{M} \mathbf{h}_2| = ||\mathbf{h}||^2 |\rho| |\tilde{\mathbf{h}}^{\mathrm{H}} \mathbf{m}| \tag{4.21}$$

where the vector $\tilde{\mathbf{h}}$ is

$$\tilde{\mathbf{h}} = \begin{bmatrix} |h_1|^2 - |h_2|^2 \\ 2h_1^* h_2^* \\ 2h_1 h_2 \\ |h_2|^2 - |h_1|^2 \end{bmatrix} \tag{4.22}$$

and $\mathbf{m}$ is a vector containing the components of matrix $\mathbf{M}$, i.e

$$\mathbf{m} = \begin{bmatrix} m_{11} \\ m_{21} \\ m_{12} \\ m_{22} \end{bmatrix} \tag{4.23}$$

Our aim is to find the vector $\mathbf{m}$ that maximizes the eigenvalue spread, which can be expressed as

$$\mathbf{m}_{\mathrm{opt}} = \underset{||\mathbf{m}||^2 = 1}{\arg \max} \; |\tilde{\mathbf{h}}^{\mathrm{H}} \mathbf{m}| \tag{4.24}$$

Since the scalar product between $\tilde{\mathbf{h}}$ and $\mathbf{m}$ will be maximum when $\mathbf{m}$ has the direction and sense of $\tilde{\mathbf{h}}$, the optimum value is

$$\mathbf{m}_{\text{opt}} = \frac{\tilde{\mathbf{h}}}{||\tilde{\mathbf{h}}||} \tag{4.25}$$

where

$$||\tilde{\mathbf{h}}|| = \sqrt{2(|h_1|^2 - |h_2|^2)^2 + 8|h_1|^2|h_2|^2} \tag{4.26}$$

and thus, multiplying and dividing in equation (4.26) by $|h_1|^2 - |h_2|^2$, the following expression is obtained

$$\begin{aligned} ||\tilde{\mathbf{h}}|| = \frac{|h_1|^2 - |h_2|^2}{|h_1|^2 - |h_2|^2}||\tilde{\mathbf{h}}|| &= (|h_1|^2 - |h_2|^2)\sqrt{2 + 2\frac{4|h_1|^2|h_2|^2}{(|h_1|^2 - |h_2|^2)^2}} \\ &= (|h_1|^2 - |h_2|^2)\sqrt{2 + 2\left|\frac{2h_1h_2}{|h_1|^2 - |h_2|^2}\right|^2} \end{aligned} \tag{4.27}$$

Defining the term β as

$$\beta = \frac{2h_1h_2}{|h_1|^2 - |h_2|^2} \tag{4.28}$$

we can obtain

$$||\tilde{\mathbf{h}}|| = (|h_1|^2 - |h_2|^2)\sqrt{2 + 2|\beta|^2} \tag{4.29}$$

Hence, replacing equations (4.22) and (4.29) into (4.25), the vector $\mathbf{m}_{\text{opt}}$ can be expressed as

$$\mathbf{m}_{\text{opt}} = \frac{1}{(|h_1|^2 - |h_2|^2)\sqrt{2 + 2|\beta|^2}} \begin{bmatrix} |h_1|^2 - |h_2|^2 \\ 2h_1^*h_2^* \\ 2h_1h_2 \\ |h_2|^2 - |h_1|^2 \end{bmatrix} = \frac{1}{\sqrt{2 + 2|\beta|^2}} \begin{bmatrix} 1 \\ \beta^* \\ \beta \\ -1 \end{bmatrix} \tag{4.30}$$

and rearranging these terms, the optimum matrix $\mathbf{M}_{\text{opt}}$ is given by

$$\mathbf{M}_{\text{opt}} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \frac{1}{\sqrt{2 + 2|\beta|^2}} \begin{bmatrix} 1 & \beta \\ \beta^* & -1 \end{bmatrix} \tag{4.31}$$

Note that this optimum matrix depends on the parameter β which requires the current values of channel coefficients h_1, h_2, which are unknown. In Section 4.4.3 we will show

a method of estimating this parameter. At this point, we can obtain a theoretical result, employing the common algorithm shown in Table 4.1 to achieve a channel matrix estimation $\hat{\mathbf{H}}$, making use of the matrix $\mathbf{C} = \mathbf{Q}_{\mathbf{x}}(\mathbf{M}_{\text{opt}})$.

4.4.1 Maximum Eigenvalue Spread Suboptimal Approach

In this section we will present a simplified and suboptimal version of the approach proposed above, which estimates the channel matrix by selecting the cumulant matrix with the highest eigenvalue spread. We will focus on the study of the following matrices

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{4.32}$$

Before continuing, note that it is shown in Appendix E that

$$\mathbf{Q}_{\mathbf{x}}(\mathbf{M}_1) = \mathbf{C}_{\text{HOS}}^{[1,1]}, \qquad \mathbf{Q}_{\mathbf{x}}(\mathbf{M}_2) = \mathbf{C}_{\text{HOS}}^{[1,2]} \tag{4.33}$$

Evaluating equation (4.21), we obtain that the associated eigenvalue spreads are

$$\begin{aligned} L(\mathbf{M}_1) &= ||\mathbf{h}||^2 \rho ||h_1|^2 - |h_2|^2| \\ L(\mathbf{M}_2) &= ||\mathbf{h}||^2 \rho 2|h_1 h_2| \end{aligned} \tag{4.34}$$

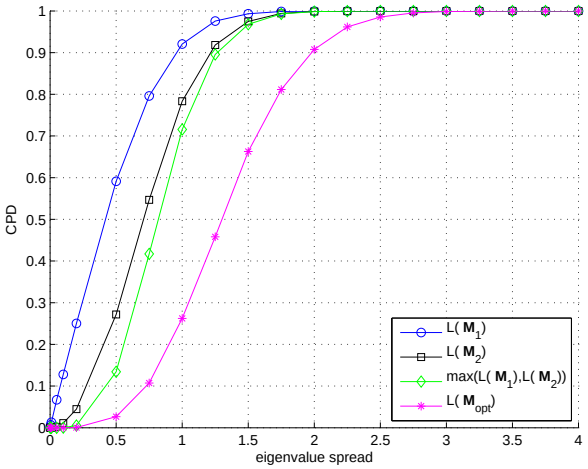
As a result, the matrix $\mathbf{M}$ that maximizes the eigenvalue spread can be selected using the following criterion

$$\frac{|L(\mathbf{M}_2)|}{|L(\mathbf{M}_1)|} = \frac{||\mathbf{h}||^2 2|h_1 h_2|}{||\mathbf{h}||^2 ||h_1|^2 - |h_2|^2|} = \frac{2|h_1||h_2|}{||h_1|^2 - |h_2|^2|} \underset{\mathbf{M}_2}{\overset{\mathbf{M}_1}{\lessgtr}} 1 \tag{4.35}$$

It is interesting to note that the decision criterion referred to above depends on the absolute value of parameter β defined in equation (4.28), i.e. the matrix $\mathbf{M}_1$ or $\mathbf{M}_2$ must be selected by using the rule

$$|\beta| \underset{\mathbf{M}_2}{\overset{\mathbf{M}_1}{\lessgtr}} 1. \tag{4.36}$$

(a) CPD



(b) Zoomed figure

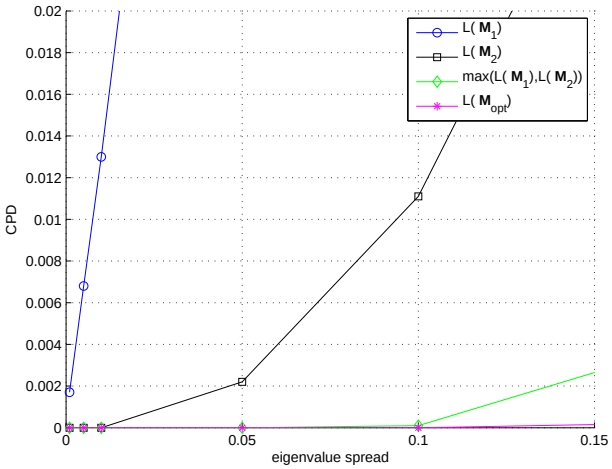


Figure 4.9: Cumulative probability distribution for different matrices $\mathbf{M}$: $L(\mathbf{M}_1)$, $L(\mathbf{M}_2)$, $\max(L(\mathbf{M}_1), L(\mathbf{M}_2))$ and $L(\mathbf{M}_{\text{opt}})$ with Rayleigh channel distribution

To summarize, in order to achieve an estimation of the channel matrix $\hat{\mathbf{H}}$ employing the common algorithm described at the beginning of this chapter, in Table 4.1, the choice of the $\mathbf{C}$ matrix between $\mathbf{Q}_{\mathbf{x}}(\mathbf{M}_1) = \mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{Q}_{\mathbf{x}}(\mathbf{M}_2) = \mathbf{C}_{\text{HOS}}^{[1,2]}$ is carried out by means of the parameter $|\beta|$.

4.4.2 Comparison Among Eigenvalue Spreads

One way of measuring the improvement obtained by using the optimal and suboptimal approaches consists in measuring the probability of the eigenvalue spread of the matrix to be diagonalized being close to zero. The best criterion has the lowest probability.

In order to compare the eigenvalue spread obtained with different matrices $\mathbf{M}$, we have evaluated equation (4.21) considering that the channel coefficients have a Rayleigh distribution and a Rice distribution. Subsequently, we have computed the *Cumulative Probability Distribution* (CPD) corresponding to each value of $L(\mathbf{M})$, employing 10^6 channel realizations. Figure 4.9 (a) plots the CPD following a Rayleigh channel distribution for $L(\mathbf{M}_1)$, $L(\mathbf{M}_2)$, the suboptimal approach which computes the maximum between $L(\mathbf{M}_1)$

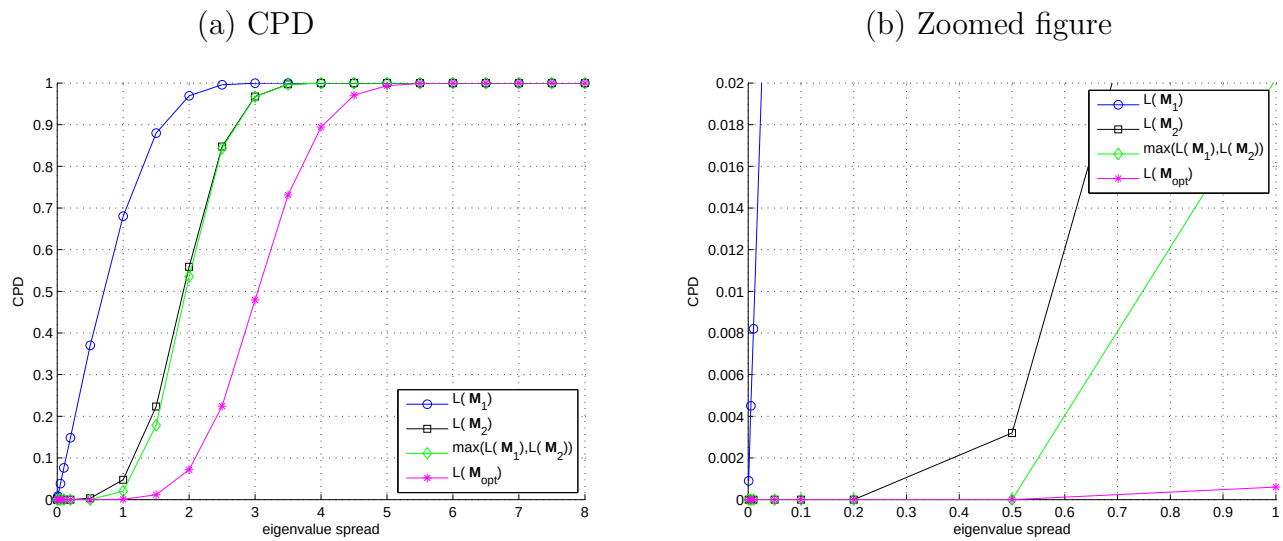


Figure 4.10: Cumulative probability distribution for different matrices $\mathbf{M}$: $L(\mathbf{M}_1)$, $L(\mathbf{M}_2)$, $\max(L(\mathbf{M}_1), L(\mathbf{M}_2))$ and $L(\mathbf{M}_{\text{opt}})$ with Rice channel distribution

and $L(\mathbf{M}_2)$, and the optimal approach which computes $L(\mathbf{M}_{\text{opt}})$. The channels have been normalized to avoid the influence of scale. It is apparent that $L(\mathbf{M}_1)$ has the highest probability of taking values close to zero. Figure 4.9 (b) zooms the part corresponding to an eigenvalue spread of less than 0.15, also for a Rayleigh distribution. Note that the CPD of the optimal approach coincides with the CPD of the suboptimal one when $L(\mathbf{M}) \leq 0.1$ and therefore it is reasonable to think that both approaches will achieve a similar performance when estimating the channel matrix.

Figure 4.10 shows the CPD, following a Rice channel distribution, for the same cases as in Figure 4.9. Note that Figure 4.10 (a) is very similar to Figure 4.9 (a), with a factor of scale in the axis corresponding to eigenvalue spread. Note also that the curves corresponding to $L(\mathbf{M}_2)$ and $\max(L(\mathbf{M}_1), L(\mathbf{M}_2))$ are closer than in the Rayleigh case. In Figure 4.10 (b), a zoomed version of Figure 4.10 (a), we can see that all the eigenvalue spreads, with the exception of $L(\mathbf{M}_1)$, are very low when $L(\mathbf{M}) \leq 0.5$.

To summarize, it is straightforward to conclude from these results that employing $\mathbf{M}_1$ we will obtain the worst result for Rayleigh and Rice channel distribution in terms of

eigenvalue spread, being the Beres and Adve Approach. Also, for both Rayleigh and Rice channel distributions the performance for the optimal and suboptimal approaches will be very similar. Finally, in the case of Rice channel distribution, for a lower eigenvalue spread, the probabilities are very similar for $\mathbf{M}_2$, suboptimal and optimal matrix $\mathbf{M}$.

4.4.3 Blind Estimation of Parameter β

The matrix $\mathbf{M}$, obtained in the approaches presented above, depends on parameter β given in equation (4.28), which is a function of the channel coefficients h_1 and h_2. Using the results in Appendix E, we can estimate this parameter from the 4^{th} order cumulants obtained from the observations. It is demonstrated that

$$\begin{aligned} \text{cum}(x_1, x_2^*, x_1, x_2^*) &= 2h_1^2 h_2^2 \rho & (4.37) \\ \text{cum}(x_1, x_1^*, x_1, x_2^*) &= (|h_1|^2 - |h_2|^2) h_1 h_2 \rho & (4.38) \end{aligned}$$

Using these definitions, it is straightforward to obtain

$$\beta = \frac{\text{cum}(x_1, x_2^*, x_1, x_2^*)}{\text{cum}(x_1, x_1^*, x_1, x_2^*)} = \frac{2h_1^2\ h_2^2\ \rho}{(|h_1|^2 - |h_2|^2) h_1\ h_2\ \rho} = \frac{2h_1 h_2}{|h_1|^2 - |h_2|^2} \tag{4.39}$$

Note that the suboptimal approach only needs to know the module of parameter β which can be estimated by computing the absolute value of equation (4.39). An alternative way of obtaining $|\beta|$ is to use the 4^{th} order cumulant of equation (4.38) and

$$\text{cum}(x_1, x_1^*, x_2, x_2^*) = 2|h_1|^2 |h_2|^2 \rho \tag{4.40}$$

In this case, we have

$$|\beta| = \frac{\text{cum}(x_1, x_1^*, x_2, x_2^*)}{|\text{cum}(x_1, x_1^*, x_1, x_2^*)|} = \frac{2|h_1|^2\ |h_2|^2 \rho}{||h_1|^2 - |h_2|^2| |h_1\ h_2| \rho} = \frac{2|h_1 h_2|}{||h_1|^2 - |h_2|^2|} \tag{4.41}$$

Simulation results presented in the following section show that this second way of estimating $|\beta|$ is more appropriate because the error in the estimation of $\text{cum}(x_1, x_2^*, x_1, x_2^*)$ is bigger than for $\text{cum}(x_1, x_1^*, x_2, x_2^*)$. Also, it is important to note that if $|h_1|$ is too close to $|h_2|$ then β cannot be properly computed.

- **Step 1**. Compute the 6 different cumulants, i.e $c_1 = \text{cum}(x_1, x_1^*, x_1, x_1^*)$, $c_2 = \text{cum}(x_1, x_1^*, x_1, x_2^*)$, $c_4 = \text{cum}(x_1, x_1^*, x_2, x_2^*)$, $c_6 = \text{cum}(x_1, x_2^*, x_1, x_2^*)$, $c_8 = \text{cum}(x_1, x_2^*, x_2, x_2^*)$ and $c_{16} = \text{cum}(x_2, x_2^*, x_2, x_2^*)$.
- **Step 2**. Compute $\beta = \frac{c_6}{c_2} = \frac{\text{cum}(x_1, x_2^*, x_1, x_2^*)}{\text{cum}(x_1, x_1^*, x_1, x_2^*)}$.
- **Step 3**. Let $\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \frac{1}{\sqrt{2+2|\beta|^2}} \begin{bmatrix} 1 & \beta \\ \beta^* & -1 \end{bmatrix}$.
- **Step 4**. Employing the cumulants which have been computed in step 1, form the following matrices

$$\mathbf{C}_{\text{HOS}}^{[1,1]} = \begin{bmatrix} c_1 & c_2 \\ c_2^* & c_4 \end{bmatrix} \quad \mathbf{C}_{\text{HOS}}^{[1,2]} = \begin{bmatrix} c_2 & c_6 \\ c_4 & c_8 \end{bmatrix}$$
$$\mathbf{C}_{\text{HOS}}^{[2,1]} = \begin{bmatrix} c_2^* & c_4 \\ c_6^* & c_8^* \end{bmatrix} \quad \mathbf{C}_{\text{HOS}}^{[2,2]} = \begin{bmatrix} c_4 & c_8 \\ c_8^* & c_{16} \end{bmatrix}$$

- **Step 5**. Obtain $\mathbf{C} = \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = m_{11}\mathbf{C}_{\text{HOS}}^{[1,1]} + m_{12}\mathbf{C}_{\text{HOS}}^{[2,1]} + m_{21}\mathbf{C}_{\text{HOS}}^{[1,2]} + m_{22}\mathbf{C}_{\text{HOS}}^{[2,2]}$
- **Step 6**. Employ the algorithm of Table 4.1 and obtain the estimation of the channel.

Table 4.2: *Maximum Eigenvalue Spread Optimal Approach* (MESOA)

Table 4.2 shows all the steps to obtain the matrix $\mathbf{Q}_{\mathbf{x}}(\mathbf{M}_{\text{opt}})$ displayed in Section 4.4, in order to employ it as matrix $\mathbf{C}$ in the General HOS Algorithm described in Table 4.1. Table 4.3 is devoted to explaining the steps of the suboptimal algorithm shown in Section 4.4.1.

4.4.4 Simulation Results

This section shows a set of experiments that has been performed with the scenario described in Section 4.1.1. In order to compare the two methods proposed above for estimat-

- **Step 1**. Compute the cumulants $c_4 = \text{cum}(x_1, x_1^*, x_2, x_2^*)$ and $c_2 = \text{cum}(x_1, x_1^*, x_1, x_2^*)$.

- **Step 2**. Obtain $|\beta|$ employing one of the following methods:

  Method 1 $|\beta| = \frac{|c_6|}{|c_2|} = \frac{|\text{cum}(x_1, x_2^*, x_1, x_2^*)|}{|\text{cum}(x_1, x_1^*, x_1, x_2^*)|}$.

  Method 2 $|\beta| = \frac{c_4}{|c_2|} = \frac{\text{cum}(x_1, x_1^*, x_2, x_2^*)}{|\text{cum}(x_1, x_1^*, x_1, x_2^*)|}$.

- **Step 3**. If $|\beta| < 1$ then

  compute the cumulant $c_1 = \text{cum}(x_1, x_1^*, x_1, x_1^*)$ and form the matrix

  $$\mathbf{C} = \mathbf{Q}_{\mathbf{x}}(\mathbf{M}_1) = \mathbf{C}_{\text{HOS}}^{[1,1]} = \begin{bmatrix} c_1 & c_2 \\ c_2^* & c_4 \end{bmatrix}$$

  else compute the cumulants $c_6 = \text{cum}(x_1, x_2^*, x_1, x_2^*)$ and $c_8 = \text{cum}(x_1, x_2^*, x_2, x_2^*)$.

  $$\mathbf{C} = \mathbf{Q}_{\mathbf{x}}(\mathbf{M}_2) = \mathbf{C}_{\text{HOS}}^{[1,2]} = \begin{bmatrix} c_2 & c_6 \\ c_4 & c_8 \end{bmatrix}$$

- **Step 4**. Employ the algorithm described in Table 4.1 and obtain the estimation of the channel.

Table 4.3: *Maximum Eigenvalue Spread Suboptimal Approach* (MESSA)

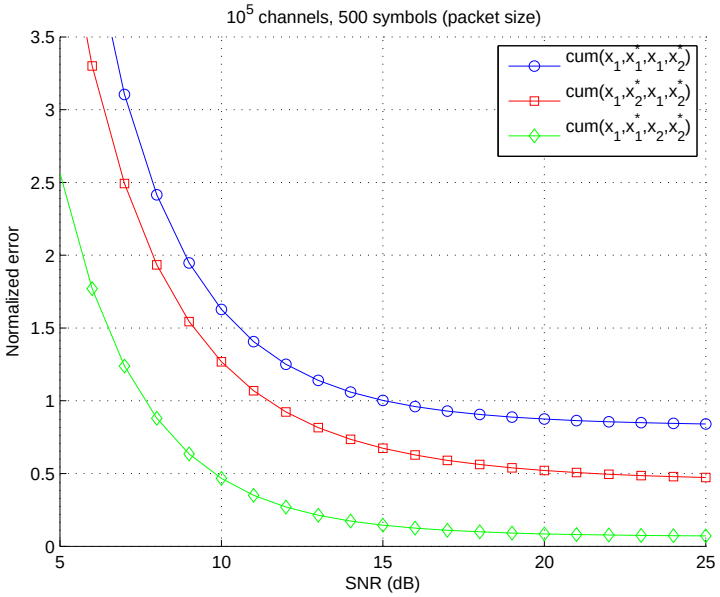


Figure 4.11: Normalized error between the theoretical and the estimated value of the 4^{th} order cumulants $\text{cum}(x_1, x_1^*, x_1, x_2^*)$, $\text{cum}(x_1, x_2^*, x_1, x_2^*)$ and $\text{cum}(x_1, x_1^*, x_2, x_2^*)$.

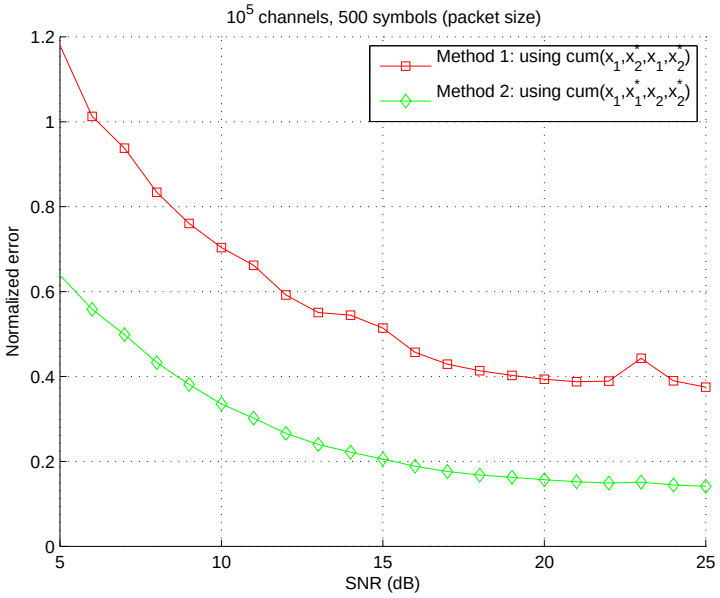


Figure 4.12: Normalized error in the estimation of $|\beta|$ using $|\beta| = \frac{|\text{cum}(x_1,x_2^*,x_1,x_2^*)|}{|\text{cum}(x_1,x_1^*,x_1,x_2^*)|}$ and $|\beta| = \frac{\text{cum}(x_1,x_1^*,x_2,x_2^*)}{|\text{cum}(x_1,x_1^*,x_1,x_2^*)|}$.

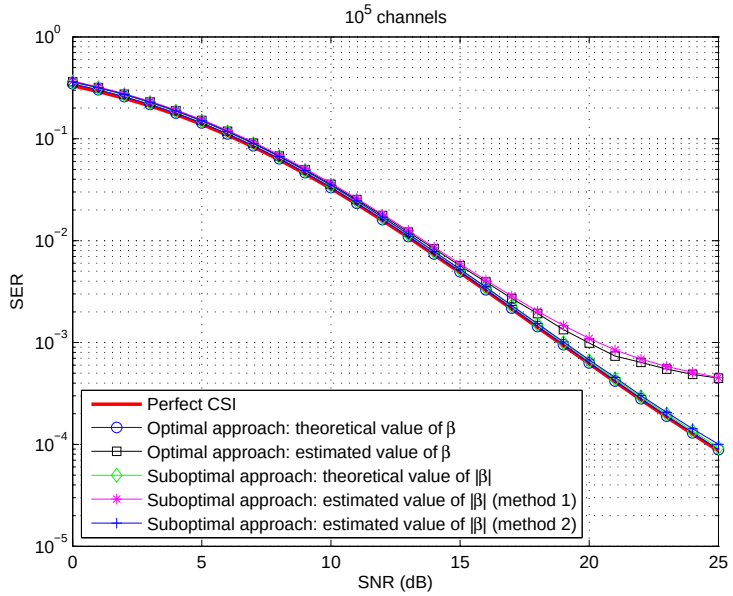


Figure 4.13: Performance of the eigenvalue spread HOS algorithms for Rayleigh channel distribution: SER versus SNR

ing the module of parameter β, Figure 4.11 shows the normalized error in the estimation of the 4^{th} order cumulants $\text{cum}(x_1, x_1^*, x_1, x_2^*)$, $\text{cum}(x_1, x_2^*, x_1, x_2^*)$ and $\text{cum}(x_1, x_1^*, x_2, x_2^*)$. The normalized error has been computed using the following expression

$$\text{error} \quad = \quad \frac{|\text{cum}(x_i, x_j^*, x_k, x_l^*) - \hat{\text{cum}}(x_i, x_j^*, x_k, x_l^*)|}{|\text{cum}(x_i, x_j^*, x_k, x_l^*)|} \tag{4.42}$$

where $\text{cum}(x_i, x_j^*, x_k, x_l^*)$ represents the theoretical value, computed directly from the channel realizations from a Rayleigh distribution, using equations (4.37), (4.38) and (4.40), and $\hat{\text{cum}}(x_i, x_j^*, x_k, x_l^*)$ is the estimated value obtained by sample averaging a sequence of $K = 500$ symbol vectors of 4-QAM following an equiprobable distribution. It is apparent that the error in the estimation of $\text{cum}(x_1, x_1^*, x_2, x_2^*)$ is considerably smaller than the error obtained for $\text{cum}(x_1, x_2^*, x_1, x_2^*)$. This means that the best way of estimating $|\beta|$ consists in estimating the 4^{th} order cumulants used in equation (4.41) instead of considering the module of equation (4.39), as can be observed in Figure 4.12.

Figures 4.13 and 4.14 plot the performance of the proposed approaches, respectively, for Rayleigh and Rice channel distributions. In both figures, both algorithms match the

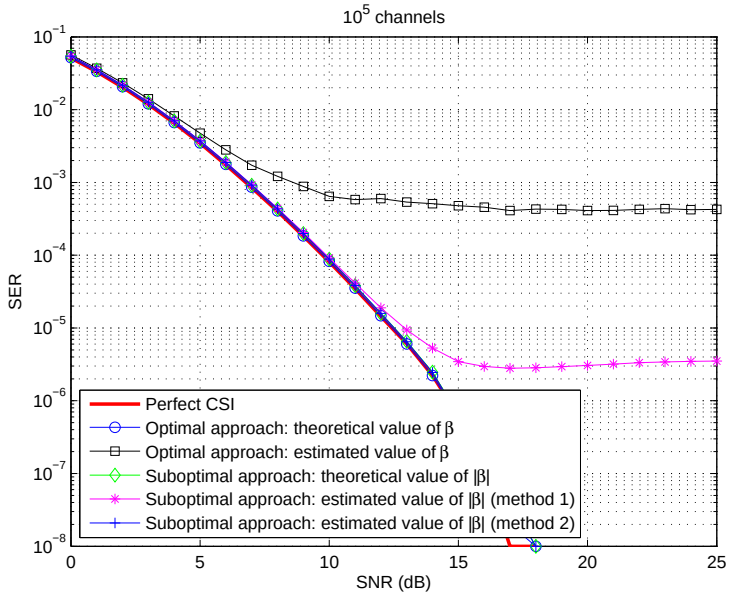


Figure 4.14: Performance of the eigenvalue spread HOS algorithms for Rice channel distribution: SER versus SNR

Perfect CSI when the theoretical value of β is used. Note also that both the optimal and the suboptimal approaches present a loss of performance for high SNRs when the parameter β is estimated using equation (4.39). This undesirable situation does not appear when the module of β is computed employing equation (4.41).

4.5 Joint-Diagonalization Approach

The JADE algorithm presented in Section 2.4.2 is another method for estimating the channel matrix by diagonalizing 4^{th} order cumulant matrices. However, it presents a high computational cost because, for the (2×1) Alamouti code, it computes 16 cumulants and diagonalizes a 3×3 matrix. Also, note that the first step consists of whitening the observations.

The first way to reduce the computational load is to eliminate the whitening step because the matrix $\mathbf{H}$ in the Alamouti scheme is orthogonal and the observations are

- **Step 1** Compute the cumulants $c_1 = \text{cum}(x_1, x_1^*, x_1, x_1^*)$, $c_2 = \text{cum}(x_1, x_1^*, x_1, x_2^*)$, $c_4 = \text{cum}(x_1, x_1^*, x_2, x_2^*)$, $c_6 = \text{cum}(x_1, x_2^*, x_1, x_2^*)$ and $c_8 = \text{cum}(x_1, x_2^*, x_2, x_2^*)$.
- **Step 2** Form the two matrices

$$\mathbf{C}_{\text{HOS}}^{[1,1]} = \begin{bmatrix} c_1 & c_2 \\ c_2^* & c_4 \end{bmatrix} \quad \mathbf{C}_{\text{HOS}}^{[1,2]} = \begin{bmatrix} c_2 & c_6 \\ c_4 & c_8 \end{bmatrix}$$

- **Step 3.** Jointly diagonalize the previous set formed by two matrices employing the method described in Appendix F. The $\hat{\mathbf{H}}$ matrix obtained is the estimation of the channel.

Table 4.4: Joint-Diagonalization HOS algorithm

therefore uncorrelated.

The second consideration is that there are only 6 different cumulants and 4 cumulant matrices. There is also a clear relation between $\mathbf{C}_{\text{HOS}}^{[1,2]}$ and $\mathbf{C}_{\text{HOS}}^{[2,1]}$, which can be observed in Appendix E: $\mathbf{C}_{\text{HOS}}^{[2,1]} = (\mathbf{C}_{\text{HOS}}^{[1,2]})^{\text{H}}$. In a situation where the channel behavior is similar in the two time slots, it is reasonable to think that it is equivalent to compute either $\mathbf{C}_{\text{HOS}}^{[1,1]}$ or $\mathbf{C}_{\text{HOS}}^{[2,2]}$. With these considerations, the computational load can be reduced by computing only the matrices $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[1,2]}$. Finally, in Appendix F, we propose a simplified method for diagonalizing these matrices using the Jacobi technique, which presents a low computational load because the eigendecomposition of the 3×3 matrix is directly obtained using the method described in the Appendix C.

We can see in Table 4.4 the steps required to implement this approach considering the simulatenous diagonalization of $\mathbf{C}_{\text{HOS}}^{[1,1]}$ and $\mathbf{C}_{\text{HOS}}^{[1,2]}$.

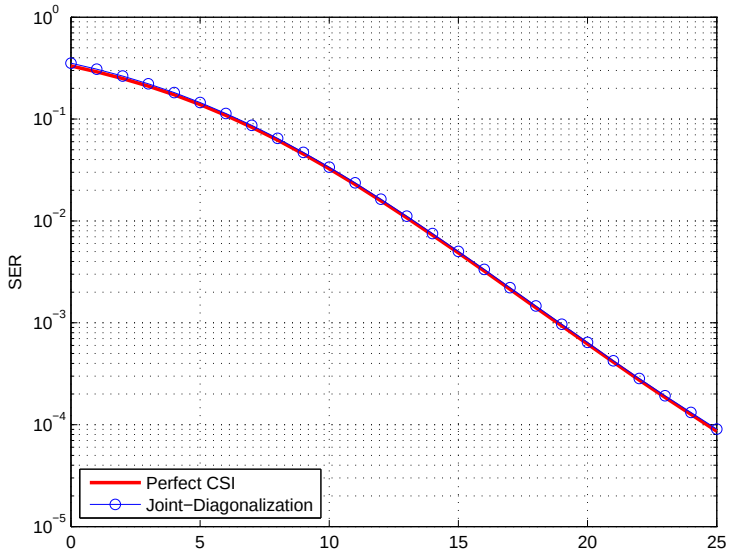


Figure 4.15: Performance of Joint-Diagonalization HOS algorithm for Rayleigh channel distribution: SER versus SNR

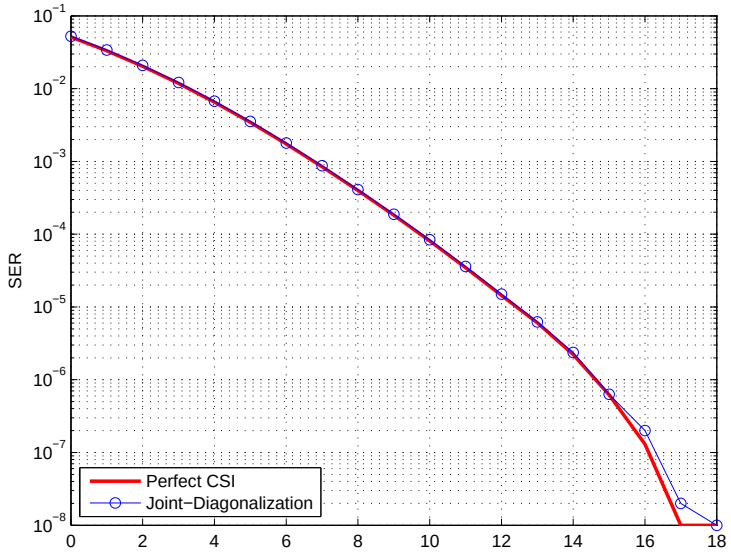


Figure 4.16: Performance of Joint-Diagonalization HOS algorithm for Rice channel distribution: SER versus SNR

4.5.1 Simulation Results

Figure 4.15 and Figure 4.16 show the SER versus SNR for a Rayleigh distribution and Rice distribution, respectively. We can see that the SER always matches the Perfect CSI for the two distributions. Comparing it with the experiment shown in Section 2.5, we can conclude that this algorithm works very well with a lower computational cost than the JADE algorithm.

Chapter 5

Performance Comparison

In this chapter, we compare the different types of blind channel estimation algorithms presented in Chapters 3 and 4. These algorithms have been proposed considering the specific properties of the (2×1) Alamouti code and are therefore more efficient than those introduced in Chapter 2 in the context of BSS. The approaches analyzed will be the following:

- The general SOS algorithm that uses a matrix $\mathbf{C}_{\text{SOS}}$ formed by 2^{nd} order statistics, which has been presented in Chapter 3.
- The general HOS algorithm with the matrix $\mathbf{C}_{\text{HOS}}^{[1,1]}$, i.e. the Beres and Adve approach presented in Section 4.1.
- The general HOS algorithm with the matrix $\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$, i.e. the Beres and Adve approach improvement I shown in Section 4.2.
- The general HOS algorithm with $\mathbf{C}_{\text{HOS}}^{[1,2]}$ matrix, i.e. the Beres and Adve approach improvement II presented in Section 4.3
- *Maximum Eigenvalue Spread Optimal Approach* (MESOA) presented in Section 4.4. The parameter β has been estimated from equation (4.39).

- *Maximum Eigenvalue Spread Suboptimal Approach* (MESSA) presented in Subsection 4.4.1, employing $|\beta|$ estimated using equation (4.41).
- The Joint-Diagonalization approach presented in Section 4.5.

In order to reduce the computational load, we have used the properties of the cumulants to compute only those that are different.

The comparison is done using computer simulations where the sources are generated with an equiprobable distribution belonging to a 4-QAM and transmitted in packets through block fading channels following a Rayleigh or a Rice distribution. We measure the performance considering the SER obtained for different SNR and different packet sizes (Section 5.1), and the associated computational load needed to compute the cumulants and to diagonalize the matrix, whose form depends on the specific approach (Section 5.2).

5.1 Symbol Error Rate Comparison

In the first experiment we have considered that each source is transmitted in blocks of 500 symbols that are used to estimate the statistics corresponding to each approach. The results have been averaged over 10^5 channel and source realizations. Figure 5.1 shows the performance of these algorithms in terms of SER versus SNR for a Rayleigh channel distribution. We can see that the Beres and Adve approach ($\mathbf{C}_{\text{HOS}}^{[1,1]}$) presents a high flooring effect, occurring at about 8 dBs, and thus this approach has the worst performance. The Beres and Adve approach improvement I ($\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$) obtains a slightly better performance. The Beres and Adve approach improvement II ($\mathbf{C}_{\text{HOS}}^{[1,2]}$) achieves a significantly better behavior because the flooring effect appears at about 18 dBs. The performance for MESOA with estimated β is very similar to the preceding one. The SOS approach is close to Perfect CSI, and MESSA and the Joint-Diagonalization approach overlap with Perfect CSI.

Figure 5.2 shows the performance of these algorithms in terms of SER versus SNR when

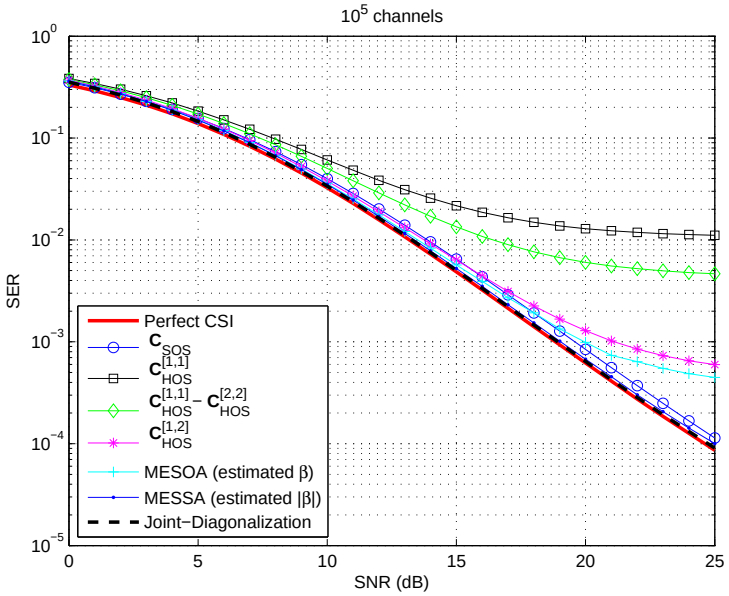


Figure 5.1: Performance comparison of algorithms for Rayleigh channel distribution: SER versus SNR

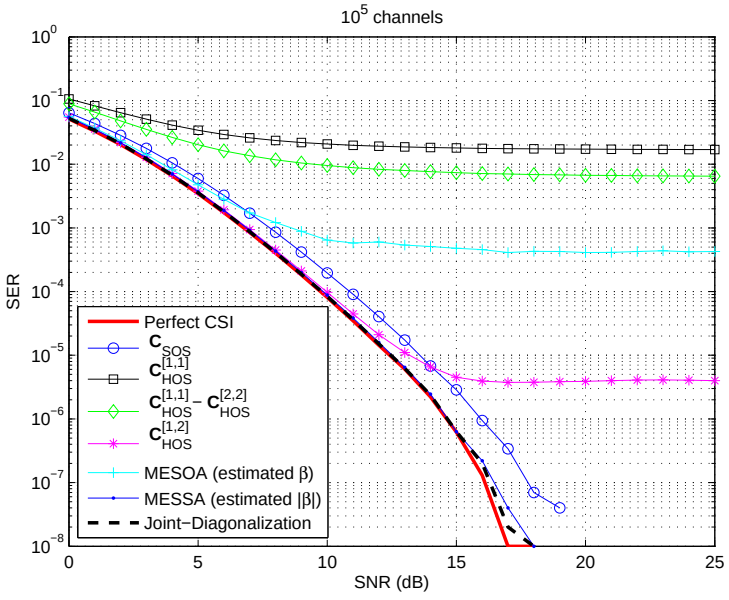


Figure 5.2: Performance comparison of algorithms for Rice channel distribution: SER versus SNR

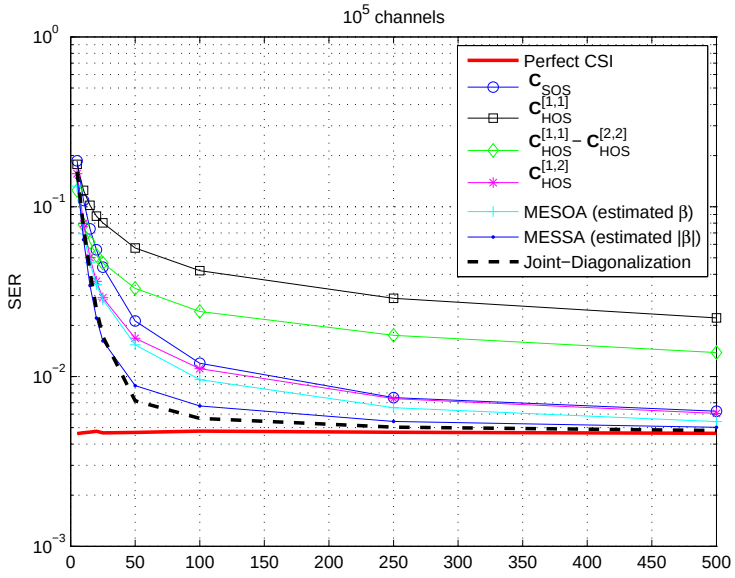


Figure 5.3: Performance comparison of algorithms for Rayleigh channels: SER versus block size

the channel follows a Rice channel distribution. We can see that the difference between the algorithms is more significant than in the case of a Rayleigh channel distribution, but again the worst performance is obtained by the Beres and Adve approach. The SOS approach differs more with regard to the Perfect CSI than in the previous experiment. Note also that MESSA and the Joint-Diagonalization approach again match the Perfect CSI performace.

The second experiment tests the dependence of the SER with respect to the packet size. In this case, Figure 5.3 shows the performance for a Rayleigh channel distribution for an SNR of 15 dB, varying the block size. We can see that the Beres and Adve approach shows the worst performance, followed by the Beres and Adve Improvement I. The Beres and Adve Improvement II works significantly better, overlapping with the SOS approach and MESOA. MESSA and the Joint-Diagonalization approach are very similar in terms of performance and almost achieve Perfect CSI when the block size is greater than 350.

Figure 5.4 shows the performance with Rice channel distribution. Again, the results

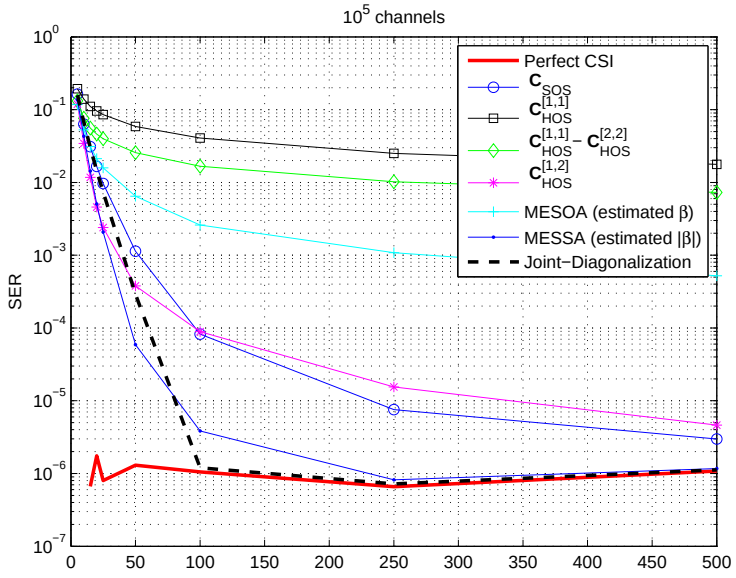


Figure 5.4: Performance comparison of algorithms for Rice channels: SER versus block size

are very similar to the Rayleigh channel distribution case. The most significant difference is that the Joint-Diagonalization method matches the Perfect CSI with only 100 symbols while MESSA needs 250 symbols. Note, however, that the achieved SER for MESSA is better when the number of symbols is lower than 50.

5.2 Computational Cost

The decoding complexity of methods based on 2^{nd} or 4^{th} order statistics matrix diagonalization depends on two parameters: the number of cumulants to be computed and the size of the matrix to be diagonalized. Table 5.1 shows the number of operations (summations and multiplications) and the *FLoating-point OPerations* (flop)[1] associated with computing 2^{nd} and 4^{th} order cumulants for a block of K points. We consider that all the operations are performed with complex-valued numbers: a summation corresponds to 2

[1]We will use the term flop for both singular and plural to avoid confusion with Floating point Operations Per Second (FLOPS).

	Sums (2 flop)	Multiplications (6 flop)	Squared root (8 flop)	flop
2^{nd} order statistic (SOS)	$K-1$	$K+1$	0	$8K+4$
4^{th} order statistic (FOS)	$7K-4$	$9K+4$	0	$68K+16$
Compute eigenvectors	6	11	3	102
Joint-Diagonalization	21	24	1	$194K+\text{flop(EVD)}$

Table 5.1: Computational load corresponding to the computation of 2^{nd} and 4^{th} order cumulants and matrix diagonalization

Approach	Number of cumulants	Size of the matrix to be diagonalized	flop
$\mathbf{C}_{\text{SOS}}$	3 SOS	2×2	$24K+114$
$\mathbf{C}_{\text{HOS}}^{[1,1]}$	3 FOS	2×2	$204K+150$
$\mathbf{C}_{\text{HOS}}^{[1,1]}-\mathbf{C}_{\text{HOS}}^{[2,2]}$	5 FOS	2×2	$340K+182$
$\mathbf{C}_{\text{HOS}}^{[1,2]}$	4 FOS	2×2	$272K+166$
MESOA	6 FOS	2×2	$408K+198$
MESSA	3,7 FOS	2×2	$251.2K+161.2$
Joint-Diagonalization Approach	5 FOS	3×3	$340K+\text{flop(EVD)}+274$

Table 5.2: Computational load of the approaches studied

flop, a multiplication corresponds to 6 flop and a square root to 8 flop. This table also shows the operations (and flop) related with the procedure used to diagonalize the cumulant matrices. It should be remembered that all the approaches described in Chapter 3 and Chapter 4, except the Joint-Diagonalization approach, compute eigenvectors of a 2×2 matrix using expressions shown in Appendix C. The term flop(EVD) denotes the number of flop needed to compute the eigenvalues of a 3×3 matrix, which is a significant quantity, as we can also see in Appendix C.

We can see that the general SOS algorithm which uses $\mathbf{C}_{\text{SOS}}$ matrix, is by far the least expensive in terms of computation. The approaches which employ the general HOS algorithm presented in Chapter 4 with different selection of matrix $\mathbf{C}$ are very similar in

Approach	Performance	Computational Cost
Joint-Diagonalization Approach	1^{st} place	7^{th} place
MESSA	2^{nd} place	3^{rd} place
$\mathbf{C}_{\text{SOS}}$	3^{rd} place	1^{st} place
$\mathbf{C}_{\text{HOS}}^{[1,2]}$	4^{th} place	4^{th} place
MESOA	5^{th} place	6^{th} place
$\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$	6^{th} place	5^{th} place
$\mathbf{C}_{\text{HOS}}^{[1,1]}$	7^{th} place	2^{nd} place

Table 5.3: Ranking classification in terms of performance and computational cost

terms of computation, the difference lying in the number of 4^{th} order statistics, which is 3, 5, 4 and 6 for $\mathbf{C}_{\text{HOS}}^{[1,1]}$, $\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$, $\mathbf{C}_{\text{HOS}}^{[1,2]}$ and MESOA, respectively. In the case of MESSA, the number of 4^{th} order cumulants to be computed depends on the decision criterion given in equation (4.36). Referring back to the equation (4.32), in 10^5 independent simulations with a Rayleigh channel distribution, we have obtained that $\mathbf{M}_1$ (3 different 4^{th} order cumulants) is used 30% of the time and $\mathbf{M}_2$ (4 different 4^{th} order cumulants) is computed 70% of the time, which corresponds to computing an average number of 3.7 when MESSA is used. The Joint-Diagonalization approach is the most expensive because the term flop(EVD) requires a significant amount of computation. Taking into account these considerations, Table 5.2 shows the number of 2^{nd} and 4^{th} order cumulants (denoted by SOS and FOS) and the size of the matrix to be diagonalized. The last column shows the flop of each approach computed by considering Table 5.1.

Finally, Table 5.3 shows the classification of the approaches following the terms of performance (SER versus SNR) and computational cost. Hence, we can conclude that MESSA and $\mathbf{C}_{\text{SOS}}$ are the most interesting approaches. Altough the choice will be dependent on the most important parameter (SER or computational load) of each specific application, it must be taken into account that the unbalance step introduced in SOS produces a degradation in the SER obtained for one source.

Chapter 6

MIMO Testbed

This chapter is devoted to evaluating the performance of the channel estimation algorithms described in the previous chapters using a MIMO testbed developed at Universidade da Coruña, using hardware from Sundance Multiprocessor Ltd. This MIMO testbed, configured as a 2×1 system, has been designed to operate at the 2.4 GHz Industrial, Scientific and Medical (ISM) band, and is intended for the testing and rapid prototyping of MIMO baseband modules. The testbed operation consists of performing the signal processing off-line at both the transmitter and the receiver while the data are sent and acquired in real time. This property enables, at the signal generation stage, modulation and space-time coding operations to be carried out off-line using `MATLAB`. At the receiver, the acquired data stream is also processed in `MATLAB`: time and frequency synchronization, channel estimation, space-time decoding and, finally, symbol-by-symbol detection are the fundamental operational blocks.

This chapter is structured as follows. Section 6.1 is devoted to describing the elements that make up the MIMO testbed. Section 6.2 describes the scenarios employed to make the transmissions. Section 6.3 shows and discusses the experimental results.

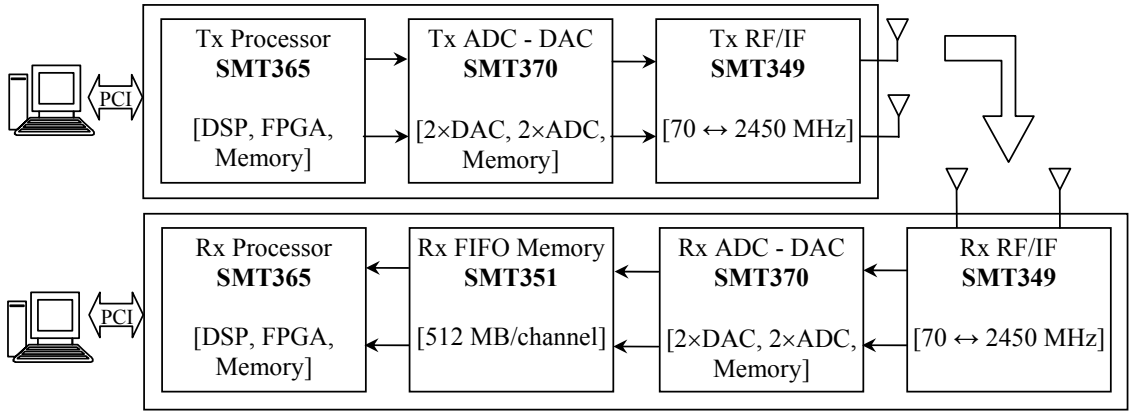


Figure 6.1: MIMO testbed

6.1 Description of the MIMO Testbed

A first release of the MIMO testbed, developed by the Universidade da Coruña, was presented in [49]. Figure 6.1 shows its general organization and Figure 6.2 shows a picture of the current equipment.

The testbed is mainly constructed by using two conventional PCs, each equipped with a `Sundance` SMT310Q PCI carrier board composed of four different kinds of module: the processing module (SMT365), the storage module (SMT351), the A/D and D/A conversion module (SMT370) and, finally, the RF front-end module (SMT349).

The transmit PC host consists of a Sundance SMT310Q PCI carrier board, an SMT365 processing module that acts as the master module, an A/D and D/A SMT370 converter module and, finally, an RF/IF SMT349 converter module (see Figure 6.1). The processing module contains a DSP TMS320C6416 at 600 MHz from Texas Instruments, a Xilinx Virtex-II FPGA, 16 MB of ZBTRAM and two SHB ports. The SMT370 is equipped with a dual D/A converter AD9777 from Analog Devices. It has 16 bits of resolution and is capable of transmitting up to 400 MS/s. The SMT370 also has two AD6645-105 A/D converters from Analog Devices with 14 bits of resolution and maximum speed of 105 MS/s. The transfer speed between processing and conversion modules achieves 200 MB/s per channel. The SMT370 also has a 1 MSample/channel memory that is used as a buffer

Figure 6.2: MIMO testbed picture

for each frame to be sent.

The transmitter is configured to generate the IF signal at the carrier frequency of 10 MHz with a sampling frequency of 80 MHz in order to generate a 70 MHz replica at the input of the RF/IF converter module. The SMT349 upconverts the IF signals to a carrier RF frequency of 2.45 GHz in two stages.

The receiver host contains another SMT310Q carrier board with the same modules as the transmitter but a SMT351 1 GB FIFO storage module is additionally placed between the processing and A/D converter modules. This memory module is used to save the data acquired by the A/D converters of the SMT370 in real time. The SMT349 captures signals at 2.45 GHz and downconverts them to 70 MHz. The sampling frequency of the A/D converters is also set at 80 MHz in order to obtain a 10 MHz replica.

We have developed a complete multilevel protocol in order to synchronize both transmitter and receiver either at physical, transport and application levels. The protocol implementation is based on 3L Diamond, Texas Instruments Code Composer and the Sundance SMT6025 software development kit. The first level manages module configuration and data transfer between host and DSP through the PCI bus. The second level uses

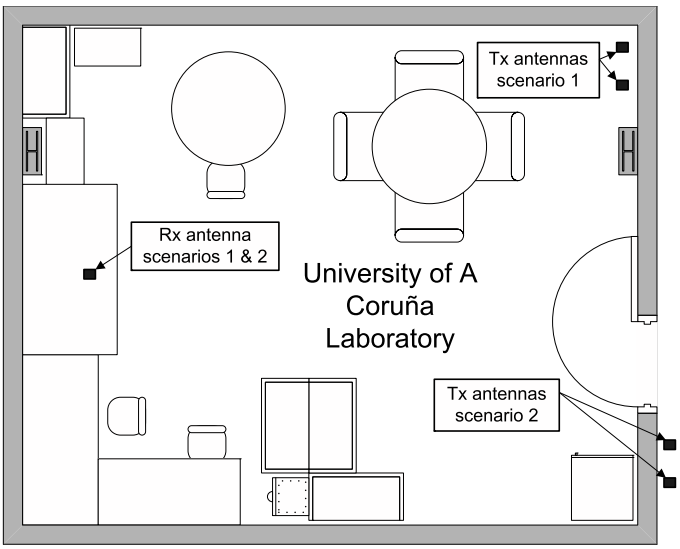


Figure 6.3: Experiment locations

a common network link to synchronize the transmitting and receiving hosts and the processing transmitter and receiver. Finally, the third level can be used to platform remote application access.

6.2 Experiment Description

We have carried out several experiments over the two scenarios shown in Figure 6.3 to compare the performance of the different blind channel estimation methods presented in Chapter 3 and Chapter 4. We have configured our testbed with 2 transmit antennas and 1 receive antenna. The experiments took place in the laboratory of the *Grupo de Tecnología Electrónica y Comunicaciones* (GTEC) at Universidade da Coruña. In scenario 1 the transmitter and the receiver were approximately 5 meters apart with a clear *Line Of Sight* (LOS) between them. In the scenario 2 configuration the transmitter was also placed about 5 meters away from the receiver, but without LOS (NLOS scenario). Note that the radio transmission frequency is 2.45 GHz, so, according to the formula $\lambda = \frac{c}{f}$, the wave length is $\lambda = \frac{3 \cdot 10^8 \text{ m/s}}{2.45 \cdot 10^9 \text{ s}^{-1}} = 12.2445$ cm. Hence, in both cases the transmit antennas were placed about 30 cm apart, i.e. more than two times λ, in order to provide a good

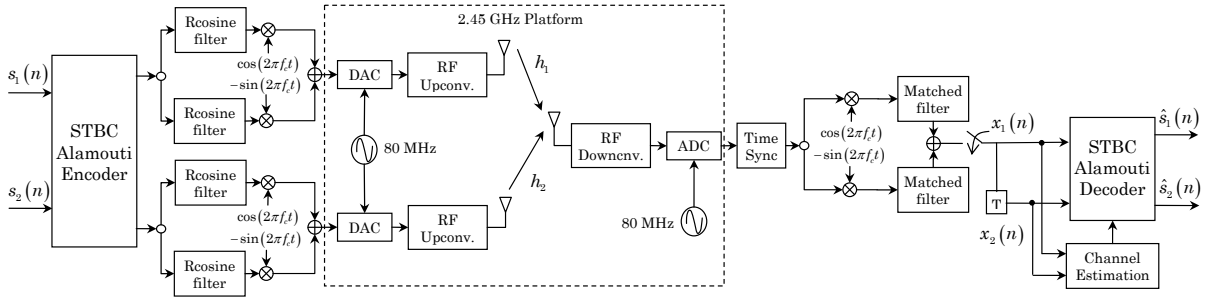


Figure 6.4: (2×1) Alamouti complete scheme

spatial diversity.

Figure 6.4 shows the block diagram of a 4-QAM modulator combined with a (2×1) Alamouti encoder that was implemented on the testbed. In order to test the different estimation methods, frames of $1,000$ 4-QAM vector symbols are transmitted, i.e. $2,000$ symbols ($1,000$ symbols for each source). Accordingly, the duration of the frame is 200 μs, lower than the channel time coherence, which ensures that the channel remains constant for the whole frame. The first subframe of 2×500 symbols is used to test the HOS-based Approaches. The second subframe, also composed of 2×500 symbols, is employed to test the SOS-based approach. The power of the second subframe is unbalanced before the Alamouti coder according to

$$\frac{1}{\sqrt{\frac{1+\gamma^2}{2}}} \cdot \begin{bmatrix} s_1 \\ \gamma s_2 \end{bmatrix}, \; 0 < \gamma < 1. \tag{6.1}$$

This equation guarantees that the total mean power of the sources in both subframes is the same with a correction factor. As a performance bound, we have also evaluated the results with *Least Squares* (LS) [50] channel estimation considering that all symbols of the first subframe are used for training. Subsequently, LS channel estimation has been used to decode only this subframe. Therefore, its performance is very close to the situation where perfect CSI is available at reception.

After the (2×1) Alamouti encoder, the following operations are carried out over the discrete complex symbols for each antenna:

- The real and imaginary parts of each symbol are split into two different sets.
- Each set is firstly up-sampled (16 samples per symbol) and filtered using a square root raised cosine pulse shaping with a roll-off factor of 40%. As a result, each set is a 2-PAM baseband signal.
- The two previous binary modulated signals are IQ modulated in order to obtain the 4-QAM signal with carrier and sampling frequency of 10 and 80 MHz respectively. The resulting signal has 7 MHz of bandwidth and 5 MBauds symbol rate.
- The next step is to pass the signal through the D/A converter configured with a sampling frequency of 80 MHz, so that a replica at 70 MHz is filtered out and up converted to a carrier RF frequency of 2.45 GHz.

With the aim of achieving a correct time synchronization for each frame, a 50 pseudo-random symbol sequence is added at the beginning of the frame obtained after the Alamouti encoder. The preamble sequence is only transmitted by one of the two antennas while the other is idle. The resulting real frame is composed of a 50 symbol preamble and $2,000$ data symbols. After the Alamouti encoder the frame has 50 preamble symbols and $4,000$ data symbols. Since we are using 16 samples per symbol, the frame contains $65,600$ 16-bit signal samples which results in a frame size equal to $128,125$ KBytes. At the receiver, the known preamble is correlated with the acquired signal to determine the first frame sample. A true carrier recovery task must also be incorporated after the time synchronization to fix signal frequency impairment due to reference oscillator errors. After IQ demodulation, a root raised cosine matched filter is used in each demodulator branch followed by a down sampler to produce the I and the Q components of the baseband signal.

In a real transmission, the SNR of the channel cannot be established a priori and in order to test the real performance for each estimation method every frame is sent several times, varying the transmitting power. Different channel realizations and distinct signal strength values are obtained in this manner. Hence, the SNR is estimated for each received frame jointly with the SER obtained by the estimation method. Finally, the pairs formed

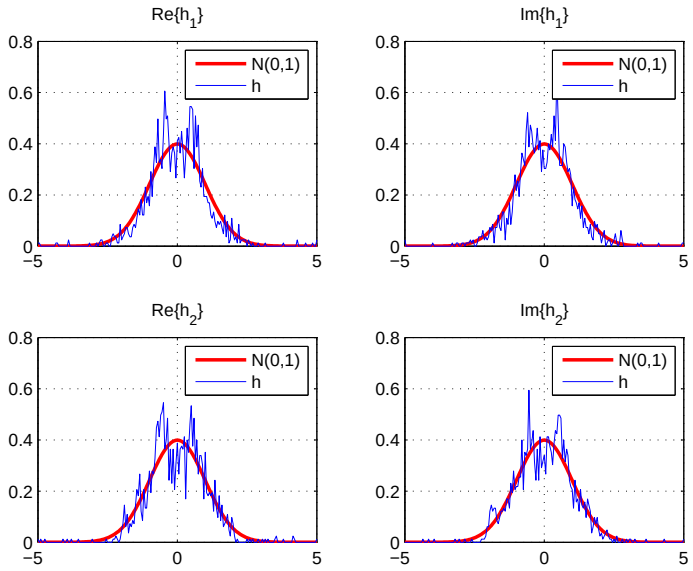


Figure 6.5: Scenario 1: LOS - probability density function

by the SNR and SER are clustered, averaged, sorted by SNR value and plotted to obtain a performance curve.

6.2.1 Channel Characterization

In order to establish a comparison with the simulation experiments presented in the previous chapter, in this section we will compare the *Probability Density Function* (PDF) and the *Cumulative Probability Distribution* (CPD) of the channel measures with a theoretical Gaussian distribution of zero mean and unit variance. Since there are two complex paths h_1 and h_2 for each transmission, we will consider 4 different channel distributions. The comparison is done before centering the channel realizations and normalizing their variance.

Figure 6.5 shows the PDF obtained in scenario 1 (LOS). It seems that the channel realization is similar to the Gaussian distribution but it appears to peak at values close to zero. In Figure 6.6 we can see that the CPD fits perfectly with the Gaussian of zero

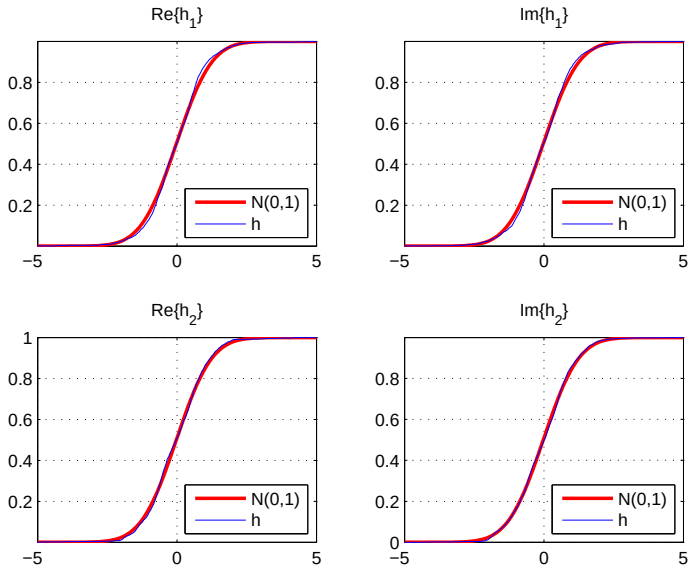


Figure 6.6: Scenario 1: LOS - cumulative distribution function

mean and unit variance.

For scenario 2 (NLOS), Figure 6.7 shows the PDF and Figure 6.8 represents the CPD. We can see that the PDF is more concentrated around zero than in the LOS scenario. Although there are also two peaks in the center, they are more significant than in LOS. It can be concluded that the channel distribution is not Gaussian. The same conclusion can be obtained from Figure 6.8 where we can see that the CPD of the real/imaginary part of two channels differs from a Gaussian of zero mean unit variance.

In order to obtain more theoretical results, we will use the chi-square (χ^2) Pearson goodness-of-fit test [51], which is considered as a non-parametric test that measures the discrepancy between an observed distribution and another, theoretical, one. From this discrepancy, the test obtains a confidence value. The procedure is simple: the observed and the theoretical distributions are divided into the same number of cells, n, and, subsequently, the corresponding number of realizations of each distribution at each interval is obtained. Let O_i and E_i be, respectively, the observed frequency and the expected (theoretical) frequency in the i-th cell.

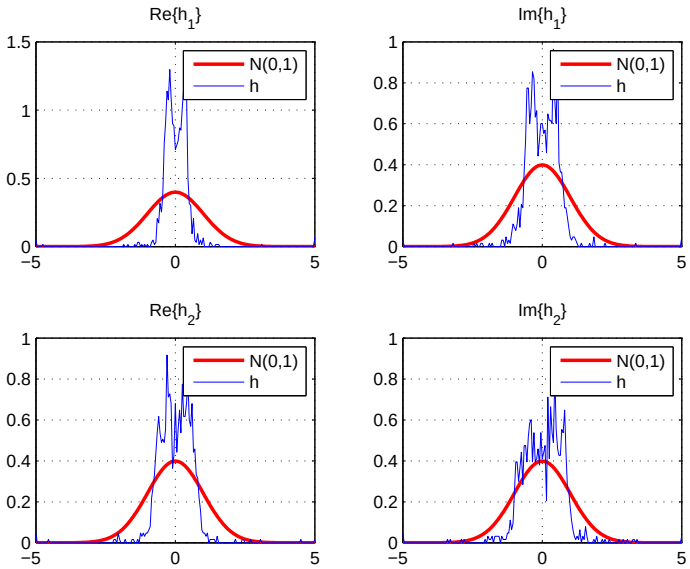


Figure 6.7: Scenario 2: NLOS - probability density function

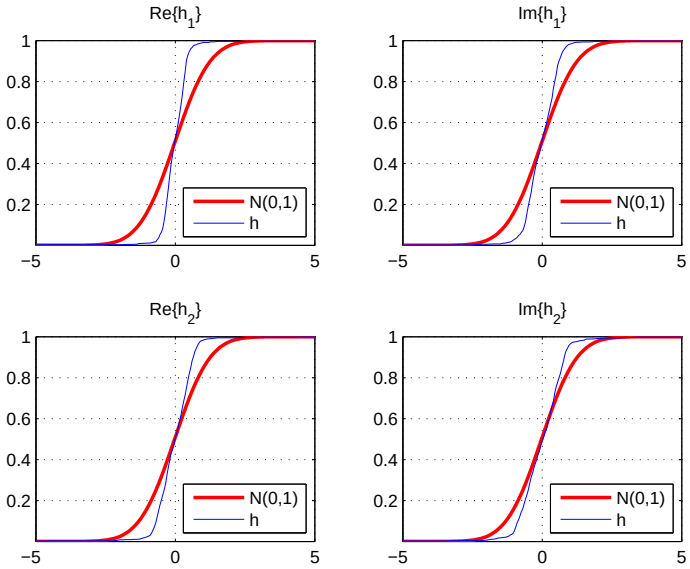


Figure 6.8: Scenario 2: NLOS - cumulative distribution function

Channel	Statistic X^2	Test	Significance
$\Re\{h_1\}$	$66,95$	Rejected	100%
$\Im\{h_1\}$	$53,13$	Rejected	100%
$\Re\{h_1\}$	$58,07$	Rejected	100%
$\Im\{h_2\}$	$31,66$	Rejected	99.91%

Table 6.1: χ^2 Pearson goodness-of-fit test - scenario 1: LOS

Channel	Statistic X^2	Test	Significance
$\Re\{h_1\}$	$1497,77$	Rejected	100%
$\Im\{h_1\}$	$671,27$	Rejected	100%
$\Re\{h_1\}$	$578,13$	Rejected	100%
$\Im\{h_2\}$	$338,35$	Rejected	100%

Table 6.2: χ^2 Pearson goodness-of-fit test - scenario 2: NLOS

The discrepancy between the two distributions is computed using the following expression

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6.2}$$

where X^2 is the test statistic that asymptotically approaches a χ^2 distribution. The X^2 statistic can then be used to calculate a p-value by comparing the value of the statistic with respect to a χ^2 distribution. The number of degrees of freedom is equal to the number of cells, n, minus the reduction in degrees of freedom, $p = 1$. In this particular case, we have chosen to divide the distributions into 12 intervals and, hence, we have used the distribution χ^2 with 11 degrees of freedom.

In Table 6.1, we can see the results of applying the χ^2 Pearson goodness-of-fit test for scenario 1 (LOS). We can see that statistics obtained have a high value, and looking at this value in the table of a χ^2 with 11 degrees of freedom, we can reject, with a significance very close to 100%, the hypothesis that these channels follow a Gaussian distribution.

In Table 6.2 we show the results of applying the Goodness-of-Fit Pearson χ^2 test to the channels of the scenario 2 (NLOS). From these values, the hypothesis that the channels

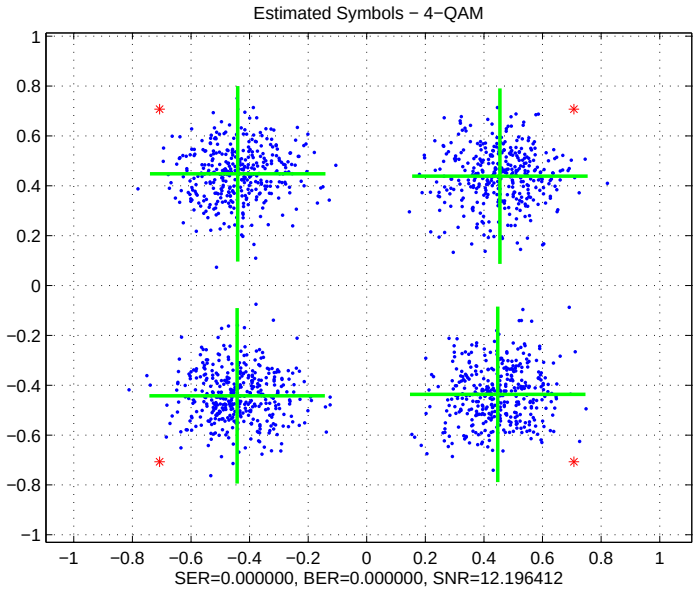


Figure 6.9: SNR estimation

follow a Gaussian distribution is also rejected with a significance very close to 100%.

6.2.2 SNR Estimation

In order to measure the performance of the channel estimation algorithms, we must determine a way of estimating the SNR at reception. Figure 6.9 plots an example of the symbols estimated using LS criterion in a transmission with the testbed using a 4-QAM. The red stars represent the symbols of the constellation $\frac{\pm 1 \pm j}{\sqrt{2}}$. We divide the estimated symbols taking into account the constellation. Thus, we obtain four clouds termed C_{1+j}, C_{1-j}, C_{-1+j} and C_{-1-j}. Subsequently, the center point of each cloud, $c_{\pm 1 \pm j}$, is computed by sample averaging of their symbols which are represented with green crosses in the center of the cloud.

The energy in reception is then estimated by computing the average of the squared modulus of the center points, i.e.

$$E_C = \frac{1}{4}\left(|c_{1+j}|^2 + |c_{1-j}|^2 + |c_{-1+j}|^2 + |c_{-1-j}|^2\right) \tag{6.3}$$

the next step is to estimate the noise power. To do so, we compute the scattering of the points in a cloud with respect to the center point,

$$\begin{aligned} E_v &= \frac{1}{4}\left(\mathrm{E}[|C_{1+j}|^2] - |c_{1+j}|^2 + \mathrm{E}[|C_{1-j}|^2] - |c_{1-j}|^2\right. \\ &+ \left.\mathrm{E}[|C_{-1+j}|^2] - |c_{-1+j}|^2 + \mathrm{E}[|C_{-1-j}|^2] - |c_{-1-j}|^2\right) \end{aligned}$$

Finally, the SNR is estimated using

$$SNR = \frac{E_C}{E_v}$$

6.3 Experimental Results

In this section we compare the same methods as in Chapter 5, but in this case over the real scenarios described above. The methods tested were:

- *Least Squares* (LS) estimator assuming knowledge of all transmitted symbols. In the case of a real transmission it is not possible to achieve Perfect CSI, and for this reason LS with the whole frame as pilot symbols is a good approximation.
- The general SOS algorithm that uses a matrix $\mathbf{C}_{\mathrm{SOS}}$ formed by 2^{nd} order statistics, which has been presented in Chapter 3.
- The general HOS algorithm with the matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,1]}$, i.e. the Beres and Adve approach presented in Section 4.1.
- The general HOS algorithm with the matrix $\mathbf{C}_{\mathrm{HOS}}^{[1,1]} - \mathbf{C}_{\mathrm{HOS}}^{[2,2]}$, i.e. the Beres and Adve approach improvement I showed in Section 4.2.
- The general HOS algorithm with $\mathbf{C}_{\mathrm{HOS}}^{[1,2]}$ matrix, i.e. Beres and Adve approach improvement II presented in Section 4.3
- *Maximum Eigenvalue Spread Optimal Approach* (MESOA), presented in Section 4.4. The parameter β has been estimated from Equation (4.39).

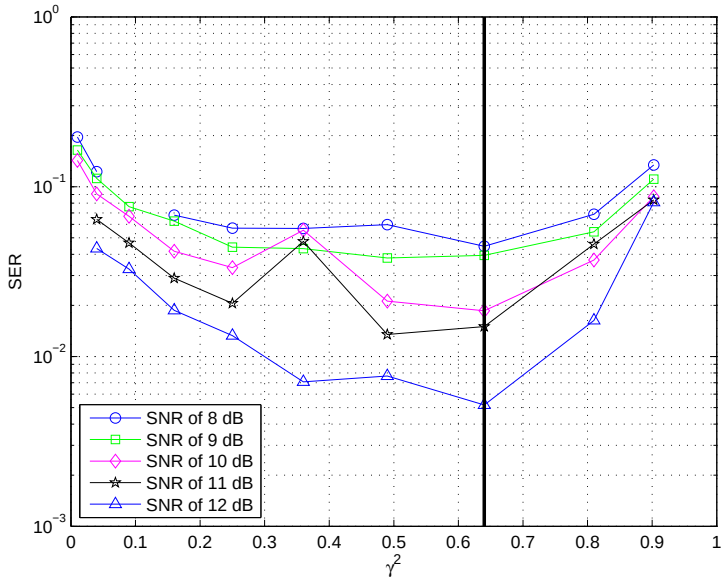


Figure 6.10: Scenario LOS: performance for the SOS-based approach according to the unbalance power factor γ^2.

- *Maximum Eigenvalue Spread Suboptimal Approach* (MESSA), presented in Subsection 4.4.1, employing $|\beta|$ estimated using Equation (4.41).

- The Joint-Diagonalization approach presented in Section 4.5.

6.3.1 Scenario 1: LOS

The configuration of scenario 1 is shown in Figure 6.3, where the transmitter and the receiver are approximately 5 meters apart with a clear *Line Of Sight* (LOS) between them.

In order to optimize the General SOS algorithm, we have performed an experiment that consists of evaluating the SER for different values of γ^2. The results plotted in Figure 6.10 shows that the optimum value is between $\gamma^2 = 0.5$ and $\gamma^2 = 0.64$. This value is in accordance with the optimal value obtained by simulations for a Rayleigh and a Rice channel in Section 3.4, which is plotted with a vertical black line ($\gamma^2 = 0.64$).

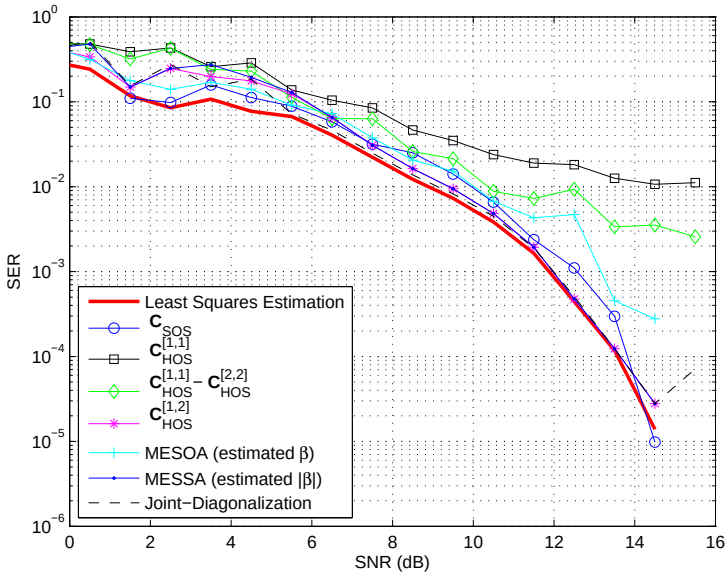


Figure 6.11: Scenario 1: LOS - performance results measuring SER according to the SNR

Figure 6.11 shows the performance results for SER versus SNR. As expected, the performance with LS estimation is the best, since it assumes that all transmitted symbols are known at the receiver. In general, the Beres and Adve Approach, which uses matrix $\mathbf{C}_{\text{HOS}}^{[1,1]}$, obtains the worst performance with a high flooring effect. The Beres and Adve Approach Improvement I, which uses matrix $\mathbf{C}_{\text{HOS}}^{[1,1]} - \mathbf{C}_{\text{HOS}}^{[2,2]}$, works a little better, and MESOA improves on this result. The SOS approach is close to LS estimation. The remaining approaches ($\mathbf{C}_{\text{HOS}}^{[1,2]}$, MESSA and Joint-Diagonalization) overlap with LS estimation for an SNR greater than 6 dBs. Note that, in general, these results are very similar to those obtained for the simulations shown in Chapter 5.

We can see, in Figure 6.12, the performance for an SNR of 10 dB in terms of SER versus block size. The red line is the LS estimation employed as optimal performance. The approaches that achieve the best performance with the least number of symbols are the Beres and Adve Improvement II ($\mathbf{C}_{\text{HOS}}^{[1,2]}$), MESSA and Joint-Diagonalization. MESOA works in a similar fashion to the Joint-Diagonalization until 250 symbols, although there is no practical improvement in performance when more symbols are employed.

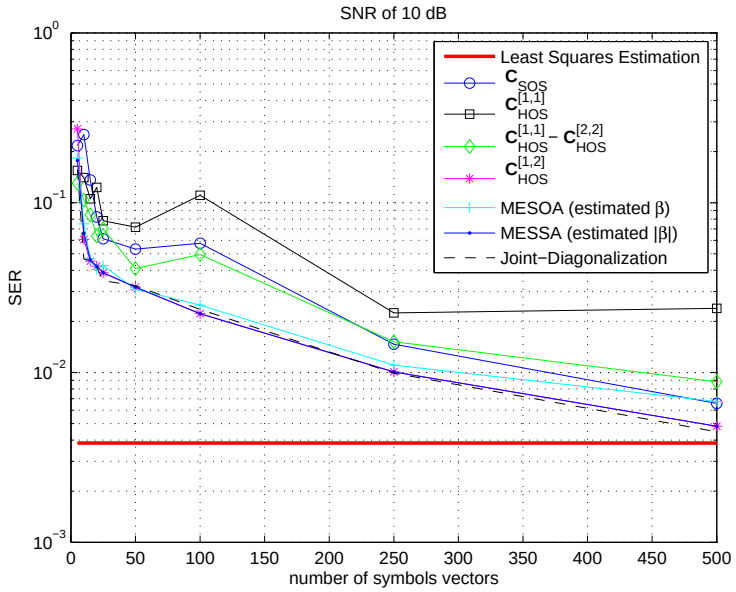


Figure 6.12: Scenario 1: LOS - performance results measuring SER according to block size for an SNR of 10 dB

6.3.2 Scenario 2: NLOS

The configuration of scenario 2 is showed in the Figure 6.3, where the transmitter and the receiver are approximately 5 meters apart, and the *Line Of Sight* LOS is avoided.

As in the LOS scenario, in order to optimize the General SOS algorithm we have performed an experiment evaluating the SER for different values of γ^2. The results obtained are shown in Figure 6.13, revealing that the optimum value is around $\gamma^2 = 0.5$. Note that this value is not in accordance with the optimal value obtained by simulations for a Rayleigh and a Rice channel in Section 3.4, which is plotted with a vertical black line ($\gamma^2 = 0.64$). Figure 6.14 plots the SER versus SNR for this scenario using $\gamma^2 = 0.5$ and 500 symbols by frame. The performance of all algorithms can be considered as acceptable and is better than in the LOS scenario. Note that the Beres and Adve Approach Improvement II ($\mathbf{C}_{\text{HOS}}^{[1,2]}$), MESOA, MESSA and Joint-Diagonalization are completely overlapped with the LS estimation.

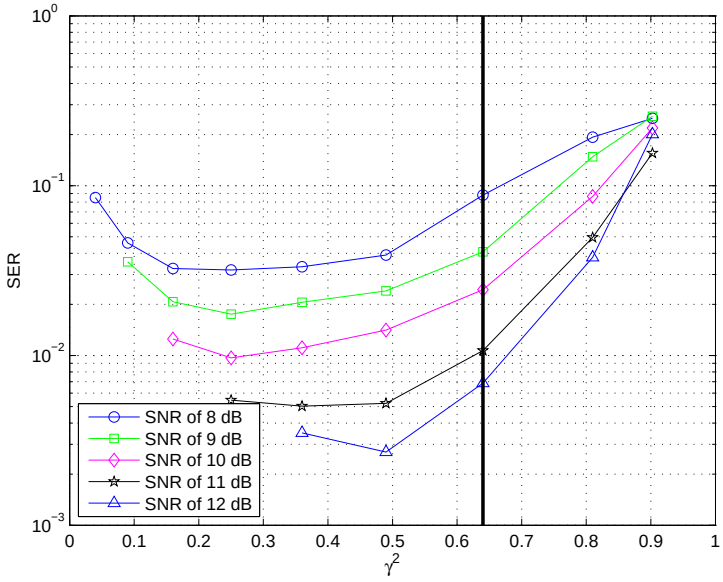


Figure 6.13: Scenario NLOS: performance for the SOS-based approach according to the unbalance power factor γ^2.

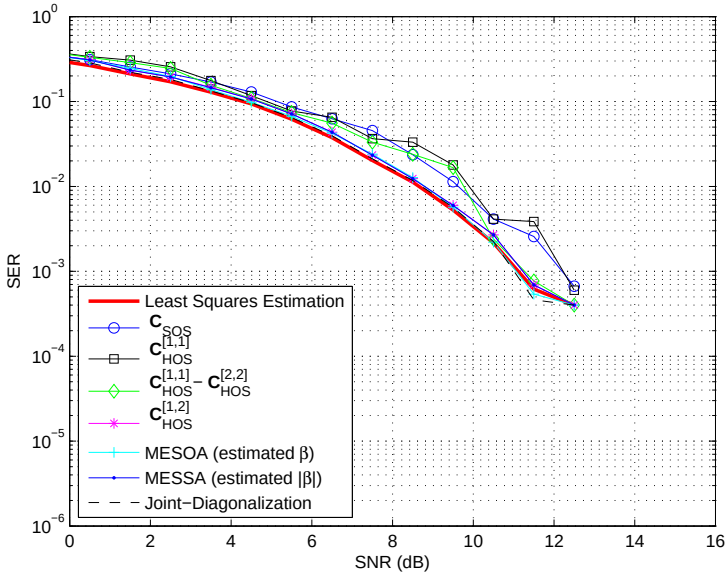


Figure 6.14: Scenario 2: NLOS - performance results measuring SER according to the SNR

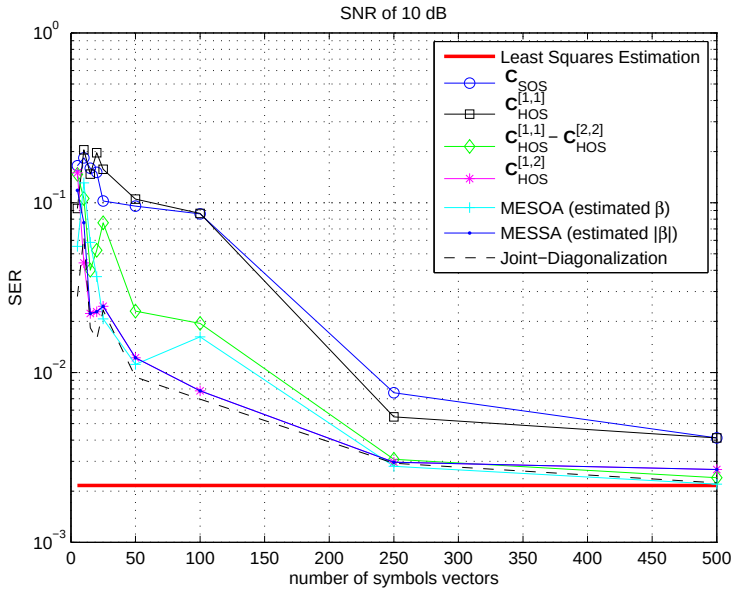


Figure 6.15: Scenario 2: NLOS - performance results measuring SER according to block size for an SNR of 10 dB

Figure 6.15 shows the SER versus block size for a transmission with an SNR of 10 dB. The worst results were obtained by the General SOS Approach and the Beres and Adve Approach ($\mathbf{C}_{\mathrm{HOS}}^{[1,1]}$), neither of which achieves the optimum performance. The other methods present a similar performance when the number of symbols is greater than 250. For fewer symbols, MESSA and Joint-Diagonalization are the best alternatives.

Chapter 7

Conclusions and Future Work

The main objective of this work is to examine and compare different types of EVD-based techniques for estimating the channel matrix (and recovering the transmitted signals) in communication systems that make use of the (2×1) Alamouti code. We started by introducing the signal model and the channel characteristic of the space-time block coding proposed by Alamouti by considering 2 transmit antenna and only 1 receive antenna. The basic premise of this code is to encode the transmitted symbols into a 2×2 unitary matrix so as to spatially decouple their ML detection, which can be seen as a matched filter followed by a symbol-by-symbol detector.

The orthogonality property imposed by OSTBC has been exploited to design specific algorithms that estimate the channel parameter by computing the eigenvalues and the eigenvectors of a matrix formed by statistics of the observations, i.e. performing an EVD. We have classified these algorithms in two classes taking into account the statistics used to estimate the matrix: SOS-based and HOS-based methods. A general condition to guarantee that the channel matrix can be identified using these methods is that the eigenvalues of the matrix to be diagonalized be different.

SOS-based methods are based on diagonalizing the correlation matrix of the observations. We have shown that the channel matrix can be identified only when the transmitted

signals have different variance. In this work, we have studied different approaches to solving this limitation. The first of these has been proposed by Shahbazpanahi et al. in [3]. These authors showed that the SOS-based channel estimation in the (2×1) Alamouti coding system can be performed by using a simple linear precoding matrix selected in order to guarantee that the eigenvalues of the correlation matrix have an order equal to one. In this sense, Via et al. [4] have proposed implementing this SOS-based approach considering the real-valued representation of the Alamouti coding scheme. In Chapter 3, we have derived a simpler implementation of this algorithm by assuming the complex-valued notation.

Another important contribution of Chapter 3 is to present simulation results obtained in different environments that allow us to empirically determine the optimal precoding matrix that must be used in SOS-based approaches. Our simulation results show that the SOS-based approach using the optimal unbalancing parameter provides an adequate mean performance, but an undesirable consequence of using a precoding step to unbalance the signal power is that the bit error for one transmitted signal may be considerable increased.

The HOS-based approaches studied in Chapter 4 consider the diagonalization of a matrix formed by 4^{th} order cumulants. This idea was initially proposed by Beres et. al. in [6]. These authors propose to using a matrix which involves a reduced set of 4^{th} order cumulants. We show that the poor performance of this method can be improved by making a better selection of the 4^{th} order cumulants. In fact, a considerable improvement is obtained when the matrix to be diagonalized is computed as a linear combination of the 4^{th} order cumulants used by Beres et. al.

The most important contribution of Chapter 4 is to obtain the closed form of the optimum matrix 4^{th} order cumulant matrix that maximizes the eigenvalue spread of cumulant matrices. Our analysis shows that the optimum matrix depends on a parameter β whose value must be computed taking into account the correct value of the channel coefficients. Simulation results verify that the performance of this approach matches the performance obtained when the receiver perfectly knows the channel parameters. We have also de-

termined a simple form of estimating the parameter β by computing the 4^{th} order cross cumulants of the observations but, unfortunately, a loss of performance for high SNR has been observed due to finite-sample estimation errors.

Another contribution of Chapter 4 is to propose a suboptimal approach which selects the matrix of highest eigenvalue spread from a set of only two cumulant matrices. This approach presents a satisfactory performance compared with the optimal approach and with other SOS- and HOS-based techniques. Furthermore, the suboptimal approach has important advantages for real implementation in FPGAs and DSPs: it computes fewer cross-cumulant matrices and diagonalizes a single 2×2 matrix. Moreover, parameter β is only used as a threshold in the suboptimal approach, making it less sensitive to estimation errors.

The performance of SOS- and HOS-based approaches has been evaluated both in computer simulations carried out assuming Rayleigh and Rice distributed channels (Chapter 5) and realistic indoor scenarios using a *Multiple Input Multiple Output* (MIMO) hardware testbed (Chapter 6). This MIMO testbed, configured as a *Two Input Single Output* (TISO) system, has been designed to operate at the 2.4 GHz *Industrial, Scientific and Medical* (ISM) band and is intended for the testing and rapid prototyping of MIMO baseband modules. The testbed operation consists of performing the signal processing off-line at both the transmitter and the receiver while the data are sent and acquired in real-time. This property enables, at the signal generation stage, modulation and space-time coding operations to be carried out off-line using `MATLAB`. At the receiver, the acquired data stream is also processed in `MATLAB`: time and frequency synchronization, channel estimation, space-time decoding and, finally, symbol-by-symbol detection are the fundamental operational blocks.

7.1 Related Publications

The results presented in this thesis have been published in different journals and international conferences:

7.1.1 Journals

1. D. Ramírez, I. Santamaría, J. Pérez, J. Vía (Dpt. de Ingeniería de Comunicaciones, Universidad de Cantabria), J. A. García-Naya, T. M. Fernández-Caramés, H. J. Pérez-Iglesias, M. González-López, L. Castedo (Dpt. de Electrónica y Sistemas, Universidade da Coruña), J. M. Torres-Royo (IMS System Engineering, Motorola Inc.), "A Comparative Study of STBC Transmissions at 2.4 GHz over Indoor Channels Using a 2 x 2 MIMO Testbed", *Wireless Communications and Mobile Computing*, John Wiley and Sons, vol. 8, no. 9, pp. 1449–1164, November 2008.
   In this paper we employ a 2×2 MIMO hardware platform to evaluate, in realistic indoor scenarios, the performance of different STBC transmissions at 2.4 GHz. In particular, we focus on the Alamouti orthogonal scheme considering two types of channel state information estimation: a conventional pilot-aided supervised technique and SOS-based approaches. For comparison purposes, we also evaluate the performance of a Differential (non-coherent) space-time block coding (DSTBC).

2. H. J. Pérez-Iglesias, J. A. García-Naya, A. Dapena, L. Castedo (Dpt. de Electrónica y Sistemas, Universidade da Coruña), V. Zarzoso (Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis), "Blind Channel Identification in Alamouti Coded Systems: A Comparative Study of Eigendecomposition Methods in Indoor Transmissions at 2.4 GHz", in *European Transactions on Telecomunications. Special Issue: European Wireless 2007*, vol. 19, Issue 7, pp. 751–759, November 2008.
   A comparative study of several channel estimation techniques in both simulated and realistic scenarios is presented. The experimental evaluation is carried out on the testbed developed at the Universidade da Coruña. The results show the superior

performance of the SOS-based blind channel estimation respect to the Beres et al. HOS approach in both LOS and NLOS channels.

3. A. Dapena, H. J. Pérez-Iglesias (Dpt. de Electrónica y Sistemas, Universidade da Coruña), V. Zarzoso (Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis), "Blind Channel Estimation Based on Maximizing the Eigenvalue Spread of Cumulant Matrices in (2 x 1) Alamouti's Coding Schemes", accepted in *Wireless Communications and Mobile Computing*, John Wiley and Sons.
   This paper addresses the problem of blind channel identification in (2×1) Alamouti coded systems employing the eigendecomposition of matrices composed of 2^{nd} and 4^{th} order statistics of the received signal. The main contribution is to propose the MESOA and the MESSA tecnhiques which perform the diagonalization of a single cumulant matrix, which is judiciously chosen by maximizing its expected eigenvalue spread. Simulation results show that the resulting technique improves existing blind (2×1) Alamouti channel estimation methods and achieves a performance close to JADE's at a fraction of the computational cost.

7.1.2 International Conferences

1. D. Ramírez, I. Santamaría, J. Pérez, J. Vía, A. Tazón, J. A. García-Naya, T. M. Fernández-Caramés, M. González López, H. J. Pérez-Iglesias, L. Castedo, "A Flexible Testbed for the Rapid Prototyping of MIMO Baseband Modules", in *ISWCS 2006*, Valencia, Spain, September 2006.

2. H. J. Pérez-Iglesias, A. Dapena, J. A. García-Naya, "Using BSS Algorithms in Alamouti Space-Time Block Coding Schemes", *2006 ICA Research Network International Workshop*, Liverpool, United Kingdom, September 2006.

3. H. J. Pérez-Iglesias, A. Dapena, L. Castedo, V. Zarzoso, "Blind Channel Identification for Alamouti's Coding System Based on Eigenvector Decomposition", *13th European Wireless Conference*, Paris, France, April 2007.

4. J. A. García-Naya, T. Fernández-Caramés, H. Pérez-Iglesias, M. González-López, L. Castedo, D. Ramírez, I. Santamaría, J. Pérez-Arriaga, J. Vía, J. M. Torres-Royo, "Performance of STBC Transmissions with Real Data", *16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, July 2007.

5. H. J. Pérez-Iglesias, A. Dapena, "Channel Estimation for O-STBC MISO Systems Using Fourth-Order Cross-Cumulants", *ICA 2007*, London, England, September 2007.

6. J. A. García Naya, T. M. Fernández Caramés, H. J. Pérez Iglesias, M. González López, L. Castedo, "A Flexible MIMO Testbed Developed at the University of A Coruña", *1st Workshop on CMCS 2007*, Duisburg, Germany, September 2007.

7. H. J. Pérez-Iglesias, J. A. García-Naya, A. Dapena, "A Blind Channel Estimation Strategy for the 2x1 Alamouti System Based on Diagonalising 4th-order Cumulant Matrices", *ICASSP 2008*, Las Vegas (Nevada), U.S.A. March 2008.

8. J. A. García-Naya, H. J. Pérez-Iglesias, A. Dapena, L. Castedo, "A Comparative Study of Blind Channel Identification Methods for Alamouti Coded Systems over Indoor Transmissions at 2.4 GHz", *SAM 2008*, Darmstadt, Germany, July 2008.

9. J. A. García-Naya, H. J. Pérez-Iglesias, T. M. Fernández-Caramés, M. González-López, L. Castedo, "A Distributed Multilayer Architecture Enabling End-User Access to MIMO Testbeds", *PIMRC 2008*, Cannes, France, September 2008.

10. H. J. Pérez-Iglesias, D. Iglesia, A. Dapena, V. Zarzoso, "Blind Channel Identification in (2x1) Alamouti Coded Systems Based on Maximizing the Eigenvalue Spread of Cumulant Matrices", *ICA 2009*, Paraty, Brazil, March 2009.

7.2 Further work

The results presented in this thesis indicate that blind techniques provide an important benefit in wireless digital communications systems because they avoid the transmission of pilot symbols that reduce the rate performance. Further work deals with the following three research lines:

Obtaining the analytical expression of SOS

The SOS-based methods presented in Chapter 3 are based on obtaining a correlation matrix whose eigenvalues have a multiplicity equal to 1. A way of satisfying this condition is to unbalance the power of the sources at the transmitter by using an additional stage before the Alamouti encode. The computer experiments presented in Chapter 3 for Rayleigh and Rice channels show that the optimum unbalancing value is $\gamma^2 = 0.64$. This value is in agreement with the results obtained in Chapter 6 using the MIMO testbed considering a scenario where there is a direct line of sight from the transmitter to the receiver. However, the results obtained in a non-line of sight environment show that other values provide better performance. Our future work deals with determining the analytical expression of the probability error of the system as a function of the value of γ taking into account the channel distribution and the type of modulation. The objective is then to obtain the parameter γ that minimizes the error probability.

Generalizing the proposed strategies

We are working on generalizing the methods studied for other applications where the mixing matrix is not orthogonal. The idea is to transform the general problem

$$\mathbf{z} = \mathbf{A}\mathbf{s} \tag{7.1}$$

where $\mathbf{A}$ is non-orthogonal into another problem

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{7.2}$$

where $\mathbf{H}$ is an orthogonal matrix. To this end, we need to compute the EVD of $\mathbf{R}_\mathbf{z} = E[\mathbf{z}\mathbf{z}^{\mathrm{H}}] = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$ and white the observations by using $\mathbf{z}_W = \mathbf{D}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{z}$. Substituting equation (7.1), we obtain that

$$\mathbf{z}_W = \mathbf{D}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{s} = \mathbf{H}\mathbf{s} \tag{7.3}$$

where $\mathbf{H} = \mathbf{D}^{-1/2}\mathbf{U}^{\mathrm{H}}\mathbf{A}$. Observe that the correlation matrix of $\mathbf{z}_W$ is given by

$$\mathbf{R}_{\mathbf{z}_W} = \mathrm{E}[\mathbf{z}_W\mathbf{z}_W^{\mathrm{H}}] = \mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^{\mathrm{H}} \tag{7.4}$$

Now, since $\mathbf{z}_W$ is a white signal ($\mathbf{R}_{\mathbf{z}_W} = \mathbf{I}$), the matrix $\mathbf{H}$ must be orthogonal. As a result, all the algorithms presented in this thesis can be applied to other typical applications of BSS [2], e.g. separation of artifacts encephalography data, reduction of noise in natural images, etc.

Another interesting reseach line is to generalize the MESSA and the MESOA algorithms to situations where the mixing matrix has a dimension greater than two. We can think, for example, of other kinds of OSTBC [15] or of any BSS problem with more than two observations and sources. For this situation, it is not clear which is the best definition of the term eigenvalue spread used in the above mentioned strategies. For this reason, it is necessary to determine the influence of different kinds of eigenvalue discrepancy measures on the estimation error. We can consider, for example, the separation between the closest eigenvalues of the separation between the biggest and the smallest eigenvalues.

Performance in OFDM systems

The combination of the Alamouti scheme with Orthogonal Frequency Division Multiplexing (OFDM) has been recently addressed in several works [52, 53].

OFDM is a frequency-division multiplexing scheme where a large number of closely-spaced orthogonal sub-carriers are used to carry data [54]. The data is divided into several parallel data streams or channels, one for each sub-carrier. Each sub-carrier is modulated with a conventional modulation scheme (such as QAM or PSK) at a low symbol rate, maintaining total data rates similar to conventional single-carrier modulation schemes in the same bandwidth. One advantage of OFDM is that frequency-selective channels are turned into a set of parallel frequency-flat channels. This allows us to apply the Alamouti scheme on each subcarrier separately so as to enable spatial diversity.

Our future work deals with testing the methods proposed in this thesis by configuring the testbed of a MIMO-OFDM. It is reasonable to think that this will present several difficulties. First, the permutation indeterminacy inherent to blind algorithms can lead to the sources being recovered in a different order in some sub-carriers. In [55], an MSE-based criterion has been proposed that allows this indeterminacy to be avoided in an OFDM system and it is predictable that a similar strategy can be also used when OFDM is combined with OSTBC. Second, we need new strategies to be developed to reduce the number of symbols required by the proposed approaches and focus on the problem of time-selective and time-varying channels.

Appendix A

Capacity of MIMO Channels

This Appendix is devoted to characterizing the capacity of *Multiple Input Multiple Output* (MIMO) channels in systems with the (2×1) and (2×2) Alamouti code. With this aim, firstly, we define the following channel model

$$\mathcal{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n_T} \\ h_{21} & h_{22} & \cdots & h_{2n_T} \\ \cdots & \cdots & \ddots & \cdots \\ h_{n_R1} & h_{n_R2} & \cdots & h_{n_Rn_T} \end{bmatrix} \tag{A.1}$$

where h_{ij} is the fading coefficient path between antennas i and j, n_T is the number of transmitting antennas and n_R is the number of receiving antennas. With this channel model and assuming that the channel is unknown at the transmitter, it can be shown in [56] that the capacity of a given MIMO channel $\mathcal{H}$ is

$$C(\mathcal{H}) = \log_2 \det \left(\mathbf{I}_{n_R} + \frac{\sigma_s^2}{\sigma_v^2} \mathcal{H}\mathcal{H}^{\mathrm{H}} \right) \tag{A.2}$$

where σ_s^2 is the variance of the sources and σ_v^2 is the variance of the noise.

A.1 Capacity of (2×1) Alamouti Code

In order to determine the capacity of (2×1) Alamouti code we define a different, but equivalent model, from the one shown in Section 2.1. The OSTBC (2×1) Alamouti code employs 2 transmit antennas and 1 receive antenna, thus, the according channel matrix in equation (A.1) is a row vector

$$\mathcal{H} = \left[\begin{array}{cc} h_{11} & h_{12} \end{array} \right] = \left[\begin{array}{cc} h_1 & h_2 \end{array} \right] = \mathbf{h}^{\mathrm{T}} \tag{A.3}$$

The capacity of the equivalent unconstrained MIMO channel is obtained by using the equation (A.2),

$$C(\mathcal{H}) = \log_2 \det \left(1 + \frac{\sigma_s^2}{\sigma_v^2} \mathcal{H}\mathcal{H}^{\mathrm{H}} \right) = \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2} \mathbf{h}^{\mathrm{T}}\mathbf{h} \right) = \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2} ||\mathbf{h}||^2 \right) \tag{A.4}$$

In the (2×1) Alamouti code, two slots are needed to transmit the sources s_1 and s_2. Then, we can express this redundancy as the following source matrix

$$\mathbf{S} = \left[\begin{array}{cc} s_1 & -s_2^* \\ s_2 & s_1^* \end{array} \right] \tag{A.5}$$

and we can obtain the following relationship between the transmit and the receive signals

$$\left[\begin{array}{cc} r_1 & r_2 \end{array} \right] = \left[\begin{array}{cc} h_1 & h_2 \end{array} \right] \left[\begin{array}{cc} s_1 & -s_2^* \\ s_2 & s_1^* \end{array} \right] + \left[\begin{array}{cc} v_1 & v_2^* \end{array} \right] \tag{A.6}$$

where v_1 and v_2^* are the noise, respectively, in the first and second time slots. In this equation, the terms r_1 and r_2 denote the observations in the first and second time slots. Note that this equation can be rewritten as follows

$$\underbrace{\left[\begin{array}{c} x_1 \\ x_2 \end{array} \right]}_{\mathbf{x}} = \left[\begin{array}{c} r_1 \\ r_2^* \end{array} \right] = \underbrace{\left[\begin{array}{cc} h_1 & h_2 \\ h_2^* & -h_1^* \end{array} \right]}_{\mathbf{H}} \underbrace{\left[\begin{array}{c} s_1 \\ s_2 \end{array} \right]}_{\mathbf{s}} + \underbrace{\left[\begin{array}{c} v_1 \\ v_2 \end{array} \right]}_{\mathbf{v}} \tag{A.7}$$

where the equivalent channel matrix $\mathbf{H}$ is a scaled orthogonal matrix which implies

$$\mathbf{H}\mathbf{H}^{\mathrm{H}} = \mathbf{H}^{\mathrm{H}}\mathbf{H} = (|h_1|^2 + |h_2|^2)\mathbf{I}_2 = ||\mathbf{h}||^2\mathbf{I}_2 \tag{A.8}$$

Hence, we can attain the unconstrained capacity of the equivalent channel **H** employing equation (A.2). Since we are using 2 time slots, the resulting capacity must be divided by 2

$$\begin{aligned} C(\mathbf{H}) &= \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}\mathbf{H}\mathbf{H}^{\mathrm{H}}\right) \\ &= \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}(|h_1|^2 + |h_2|^2)\mathbf{I}_2\right) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\sigma_s^2}{\sigma_v^2}(|h_1|^2 + |h_2|^2)\right)^2 \\ &= \log_2\left(1 + \frac{\sigma_s^2}{\sigma_v^2}(|h_1|^2 + |h_2|^2)\right) \\ &= \log_2\left(1 + \frac{\sigma_s^2}{\sigma_v^2}||\mathbf{h}||^2\right) \end{aligned} \tag{A.9}$$

which is the same capacity of a 2×1 unconstrained MIMO channel showed in equation (A.4).

A.2 Capacity of (2×2) Alamouti Code

First, we will determine the capacity of an unconstrained MIMO channel with 2 antennas in transmission and 2 antennas in reception whose channel matrix is

$$\mathcal{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \tag{A.10}$$

From equation (A.2), the capacity is given by

$$\begin{aligned}
C(\mathcal{H}) &= \log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}\mathcal{H}\mathcal{H}^{\mathrm{H}}\right) \\
&= \log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}\begin{bmatrix} (|h_{11}|^2+|h_{12}|^2) & (h_{11}h_{21}^*+h_{12}h_{22}^*) \\ (h_{11}^*h_{21}+h_{12}^*h_{22}) & (|h_{21}|^2+|h_{22}|^2) \end{bmatrix}\right) \\
&= \log_2 \det\left(\begin{bmatrix} 1+\frac{\sigma_s^2}{\sigma_v^2}(|h_{11}|^2+|h_{12}|^2) & \frac{\sigma_s^2}{\sigma_v^2}(h_{11}h_{21}^*+h_{12}h_{22}^*) \\ \frac{\sigma_s^2}{\sigma_v^2}(h_{11}^*h_{21}+h_{12}^*h_{22}) & 1+\frac{\sigma_s^2}{\sigma_v^2}(|h_{21}|^2+|h_{22}|^2) \end{bmatrix}\right) \\
&= \log_2\left(1+\frac{\sigma_s^2}{\sigma_v^2}(|h_{11}|^2+|h_{12}|^2+|h_{21}|^2+|h_{22}|^2)+\frac{\sigma_s^4}{\sigma_v^4}(|h_{11}|^2+|h_{12}|^2)(|h_{21}|^2+|h_{22}|^2)-\right. \\
&\quad \left.\frac{\sigma_s^4}{\sigma_v^4}(|h_{11}|^2|h_{21}|^2+h_{11}h_{21}^*h_{12}^*h_{22}+h_{11}^*h_{21}h_{12}h_{22}^*+|h_{12}|^2|h_{22}|^2)\right) \\
&= \log_2\left(1+\frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2+\frac{\sigma_s^4}{\sigma_v^4}(|h_{11}|^2|h_{22}|^2+|h_{12}|^2|h_{21}|^2-h_{11}h_{21}^*h_{12}^*h_{22}-h_{11}^*h_{21}h_{12}h_{22}^*)\right) \\
&= \log_2\left(1+\frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2+\frac{\sigma_s^4}{\sigma_v^4}\left(h_{11}h_{22}(h_{11}^*h_{22}^*-h_{21}^*h_{12}^*)-h_{12}h_{21}(h_{11}^*h_{22}^*-h_{21}^*h_{12}^*)\right)\right) \\
&= \log_2\left(1+\frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2+\frac{\sigma_s^4}{\sigma_v^4}|h_{11}h_{22}-h_{12}h_{21}|^2\right) \qquad \text{(A.11)}
\end{aligned}$$

Considering that the source matrix continues being the same as in equation (A.5), we define the relationship between the transmission and the reception for (2×2) Alamouti code as

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}\begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12}^* \\ v_{21} & v_{22}^* \end{bmatrix} \qquad \text{(A.12)}$$

This equation can be rewritten as

$$\underbrace{\begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} r_{11} \\ r_{12}^* \\ r_{21} \\ r_{22}^* \end{bmatrix} = \underbrace{\begin{bmatrix} h_{11} & h_{12} \\ h_{12}^* & -h_{11}^* \\ h_{21} & h_{22} \\ h_{22}^* & -h_{21}^* \end{bmatrix}}_{\mathbf{H}^{\mathrm{H}}}\underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} v_{11} \\ v_{12} \\ v_{21} \\ v_{22} \end{bmatrix}}_{\mathbf{v}} \qquad \text{(A.13)}$$

By means of the equation (A.1), we obtain the desired capacity of an (2×2) Alamouti code where again a factor $1/2$ must be included because 2 time slots are used to transmit

the same symbol. The resulting capacity is

$$\begin{aligned} C(\mathbf{H}) &= \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}\mathbf{H}\mathbf{H}^{\mathrm{H}}\right) \\ &= \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \frac{\sigma_s^2}{\sigma_v^2}(|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2)\mathbf{I}_2\right) \\ &= \frac{1}{2}\log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2}(|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2)\right)^2 \\ &= \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2}(|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2)\right) \\ &= \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2\right) \end{aligned} \tag{A.14}$$

which is lower or equal to the capacity of a (2×2) MIMO system in equation (A.11), i.e.

$$\log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2\right) \leq \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_v^2}||\mathcal{H}||_F^2 + \frac{\sigma_s^4}{\sigma_v^4}|h_{11}h_{22} - h_{12}h_{21}|^2\right) \tag{A.15}$$

Note that both capacities are equal only when $h_{11}h_{22} = h_{12}h_{21}$.

Appendix B

Computation of a Whitening Matrix

Whitening procedures are used in many BSS algorithms, like FastICA and JADE, to reduce the complexity of the problem to solve. This appendix explains two different methods to obtain white observations.

We define a non-white vector of M variables $\mathbf{x}$, such as its correlation is the following $M \times M$ matrix

$$\mathbf{R}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}] \tag{B.1}$$

The aim is to achieve a new white vector $\tilde{\mathbf{x}}$ of dimmension N from $\mathbf{x}$ which satisfies the correlation matrix

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathbf{I}_N \tag{B.2}$$

Note that $N \leq M$. In order to obtain this vector $\tilde{\mathbf{x}}$, we compute the EVD of $\mathbf{R}_{\mathbf{x}}$ given by

$$\mathbf{R}_{\mathbf{x}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}} \tag{B.3}$$

This matrix is a hermitian matrix and its eigenvalues in matrix $\mathbf{D}$ are real numbers. Instead of employing the totality of the eigenvectors, it is possible to choose only the N eigenvectors that have the corresponding major eigenvalues. In this case, $\mathbf{U}$ will have dimmension $M \times N$. From equation B.3 we define the whitening matrix

$$\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}} \tag{B.4}$$

and multiplying $\mathbf{x}$ by the $\mathbf{W}$ whitening matrix, the white vector $\tilde{\mathbf{x}}$ is obtained, i.e

$$\tilde{\mathbf{x}} = \mathbf{W}\mathbf{x} \tag{B.5}$$

Substituting the definition of $\tilde{\mathbf{x}}$ in the correlation matrix of the vector $\tilde{\mathbf{x}}$, we can demonstrated that the new set of observations are white

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{x}}} &= \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathrm{E}[\mathbf{W}\mathbf{x}(\mathbf{W}\mathbf{x})^{\mathrm{H}}] = \mathrm{E}[\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathbf{x}\mathbf{x}^{\mathrm{H}}(\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}})^{\mathrm{H}}] \\ &= \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]\mathbf{U}\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I}_N \end{aligned} \tag{B.6}$$

Instead of using the whitening matrix in equation (B.4), it is possible to employ the alternative $\mathbf{W}$ matrix

$$\mathbf{W} = \mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}} \tag{B.7}$$

In this case, substituting equation (B.7) in equation (B.2), the following relationship is obtained

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{x}}} &= \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = \mathrm{E}[\mathbf{W}\mathbf{x}(\mathbf{W}\mathbf{x})^{\mathrm{H}}] = \mathrm{E}[\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathbf{x}\mathbf{x}^{\mathrm{H}}(\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}})^{\mathrm{H}}] \\ &= \mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}} = \mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}} \\ &= \mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}}\mathbf{U} = \mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}_N \end{aligned} \tag{B.8}$$

Appendix C

EVD Computation

The *Eigenvalue Decomposition* (EVD) is important in the context of blind source separation because it allows to extract the mixing matrix by means of the decomposition of the cumulant matrices.

Bearing this in mind, the eigenvectors and eigenvalues of any square matrix $\mathbf{A}$ are defined as the pairs $(\mathbf{x}, \lambda)$ that satisfy the following relationship

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \ ||\mathbf{x}|| = 1 \tag{C.1}$$

From above equation, it is straightforward to obtain

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = 0 \tag{C.2}$$

and thus the λ values can be achieved by equating the determinant of equation (C.2) to zero, i.e.

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{C.3}$$

once the eigenvalues λ have been obtained, it is possible to calculate the eigenvectors $\mathbf{x}$, substituting the different λ values in equation (C.2) and solving the system.

We will describe the concrete algorithms to compute eigenvalues and eigenvectors for the cases of matrices of dimension 2×2 and 3×3.

C.1 EVD of a 2×2 Matrix

For any matrix $\mathbf{A}$ of dimmension 2×2

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{C.4}$$

from equations (C.3) and (C.4) we obtain the following expression

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{I}_2) = & \quad \det\left(\begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}\right) \\ = & \quad (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ = & \quad \lambda^2 - \mathrm{Tr}(\mathbf{A})\lambda + \det(\mathbf{A}) \\ = & \quad \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} \quad = 0 \end{aligned} \tag{C.5}$$

Thus, solving the 2^{nd} degree equation we obtain

$$\lambda_{1,2} = \frac{a_{11} + a_{22} \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}}{2} \tag{C.6}$$

Substituting the values of $\lambda_i,\ i = 1, 2$ in equation (C.2) and solving the system, we obtain the corresponding eigenvectors

$$\begin{bmatrix} a_{11} - \lambda_i & a_{12} \\ a_{21} & a_{22} - \lambda_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{C.7}$$

This system is homogeneous and, thus if any solution $\mathbf{x}$ different from trivial solution ($x_1 = x_2 = 0$) exists, then for any $k \in \mathbb{Z}^*,\ k\mathbf{x}$ will also be a solution of this system. Thus, we establish the value of one of the components of $\mathbf{x}$, for example, we let $x_2 = 1$, so

$$\begin{bmatrix} a_{11} - \lambda_i & a_{12} \\ a_{21} & a_{22} - \lambda_i \end{bmatrix} \begin{bmatrix} x_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{C.8}$$

It is possible to achieve the value of x_1 via the following two ways

$$(a_{11} - \lambda_i)x_1 + a_{12} = 0 \Rightarrow x_1 = \frac{a_{12}}{\lambda_i - a_{11}} \tag{C.9}$$

$$a_{21}x_1 + a_{22} - \lambda_i = 0 \Rightarrow x_1 = \frac{\lambda_i - a_{22}}{a_{21}} \tag{C.10}$$

Algorithm I

The following algorithm is a good choice when the requirements are to obtain eigenvalues and eigenvectors.

- **Step 1**. Define the variables b_1 and b_0 from the coefficients of the 2^{nd} degree polynomial equation (C.5)

$$b_1 = -(a_{11} + a_{22}),\ b_0 = a_{11}a_{22} - a_{21}a_{12} \tag{C.11}$$

- **Step 2**. Solve the 2^{nd} degree polynomial equation (C.5) and obtain the eigenvalues of $\mathbf{A}$ with the following expression

$$\lambda_{1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_0}}{2} \tag{C.12}$$

- **Step 3**. Achieve the eigenvectors

$$\mathbf{U}' = \begin{bmatrix} \mathbf{u}_1' & \mathbf{u}_2' \end{bmatrix} = \begin{bmatrix} \frac{a_{12}}{\lambda_1 - a_{11}} & \frac{a_{12}}{\lambda_2 - a_{11}} \\ 1 & 1 \end{bmatrix} \tag{C.13}$$

- **Step 4**. Obtain the norm of the eigenvectors

$$||\mathbf{u}_1'|| = \sqrt{\left(\frac{a_{12}}{\lambda_1 - a_{11}}\right)^2 + 1},\ ||\mathbf{u}_2'|| = \sqrt{\left(\frac{a_{12}}{\lambda_2 - a_{11}}\right)^2 + 1} \tag{C.14}$$

- **Step 5**. Finally, normalize the eigenvectors, i.e

$$\mathbf{U} = \begin{bmatrix} \frac{\mathbf{u}_1'}{||\mathbf{u}_1'||} & \frac{\mathbf{u}_2'}{||\mathbf{u}_2'||} \end{bmatrix} \tag{C.15}$$

Algorithm II

In the case that we need to obtain only the eigenvectors, the following algorithm is a better choice.

- **Step 1**. Define the new variables p and q according with the following expression

$$p = \frac{a_{11} - a_{22}}{2},\ q = \sqrt{p^2 + a_{12}a_{21}} \tag{C.16}$$

- **Step 2**. Achieve the non-normalized eigenvectors $\mathbf{u}'_1$ and $\mathbf{u}'_2$ from the above equation (C.16) employing the following definition

$$\mathbf{U}' = \begin{bmatrix} \mathbf{u}'_1 & \mathbf{u}'_2 \end{bmatrix} = \begin{bmatrix} \frac{a_{12}}{-p+q} & \frac{a_{12}}{-p-q} \\ 1 & 1 \end{bmatrix} \tag{C.17}$$

- **Step 3**. Obtain the norm of the eigenvectors

$$||\mathbf{u}'_1|| = \sqrt{\left(\frac{a_{12}}{-p+q}\right)^2 + 1},\ ||\mathbf{u}'_2|| = \sqrt{\left(\frac{a_{12}}{-p-q}\right)^2 + 1} \tag{C.18}$$

- **Step 4**. Finally, normalize the eigenvectors, i.e

$$\mathbf{U} = \begin{bmatrix} \frac{\mathbf{u}'_1}{||\mathbf{u}'_1||} & \frac{\mathbf{u}'_2}{||\mathbf{u}'_2||} \end{bmatrix} \tag{C.19}$$

C.2 EVD of a 3×3 Matrix

For any matrix $\mathbf{A}$ of dimension 3×3

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{C.20}$$

from equations (C.3) and (C.4) we obtain the following expression

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}_3) = & \det\left(\begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{bmatrix}\right) \\ = & -\lambda_3 + (a_{11} + a_{22} + a_{33})\lambda^2 \\ & (a_{12}a_{21} + a_{13}a_{31} + a_{23}a_{32} - a_{11}a_{22} - a_{11}a_{33} - a_{22}a_{33})\lambda \\ & +a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & -a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ = & -\lambda^3 + \mathrm{Tr}(\mathbf{A})\lambda^2 + \tfrac{1}{2}(\mathrm{Tr}(\mathbf{A}^2) - \mathrm{Tr}^2(\mathbf{A}))\lambda + \det(\mathbf{A}) = 0 \end{aligned} \tag{C.21}$$

So, at this point, we need to solve a cubic function and Cardano's method [57] is a good choice to achieve this aim. Its steps are the following

- **Step 1**. Define the variables b_2, b_1 and b_0 from the coefficients of the 3^{rd} degree polynomial equation (C.21)

$$b_2 = -\mathrm{Tr}(\mathbf{A}),\ b_1 = \frac{1}{2}\left(b_2^2 - \mathrm{Tr}(\mathbf{A}^2)\right),\ b_0 = -\det(\mathbf{A}) \tag{C.22}$$

- **Step 2**. Define the new variables p and q according to the following expression

$$p = b_1 - \frac{b_2^2}{3},\ q = b_0 + \frac{2b_2^3 - 9b_2b_1}{27} \tag{C.23}$$

- **Step 3**. From the above variables, compute the variables v_1, v_2 and v_3 obeying the following equation

$$v_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},\ v_2 = \frac{-1+\sqrt{3}j}{2}v_1,\ v_3 = \frac{-1-\sqrt{3}j}{2}v_1 \tag{C.24}$$

- **Step 4**. We can achieve the eigenvalues by means of

$$\lambda_i = u_i - \frac{p}{3v_i} - \frac{b_2}{3},\ i = 1, 2, 3 \tag{C.25}$$

- **Step 5**. Substituting every value of $\lambda_i,\ i = 1, 2, 3$ in equation (C.2) and solving the system, we obtain the corresponding eigenvector

$$\begin{bmatrix} a_{11}-\lambda_i & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda_i & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{C.26}$$

As this is a homogeneous system, we let $x_3 = 1$ and we have to solve the following system of 2 variables

$$\begin{aligned} (a_{11}-\lambda_i)x_1 + a_{12}x_2 + a_{13} = 0 \\ a_{21}x_1 + (a_{22}-\lambda_i)x_2 + a_{23} = 0 \end{aligned} \tag{C.27}$$

Hence, we conclude that the solution is

$$x_1 = \frac{-a_{12}a_{23} + a_{13}(a_{22} - \lambda_i)}{a_{12}a_{21} - (a_{11} - \lambda_i)(a_{22} - \lambda_i)},\ x_2 = \frac{-a_{13}a_{21} + a_{23}(a_{11} - \lambda_i)}{a_{12}a_{21} - (a_{11} - \lambda_i)(a_{22} - \lambda_i)},\ x_3 = 1 \tag{C.28}$$

Note that for each value of $\lambda_i,\ i = 1, 2, 3$ we have a different solution. Taking this into account, we can obtain the 3 eigenvectors

$$\mathbf{U}' = \begin{bmatrix} \mathbf{u}'_1 & \mathbf{u}'_2 & \mathbf{u}'_3 \end{bmatrix} = \begin{bmatrix} x_1(\lambda_1) & x_1(\lambda_2) & x_1(\lambda_3) \\ x_2(\lambda_1) & x_2(\lambda_2) & x_2(\lambda_3) \\ 1 & 1 & 1 \end{bmatrix} \tag{C.29}$$

Finally, we divide by their norm in order to achieve the normalized eigenvectors

$$\mathbf{U} = \begin{bmatrix} \frac{\mathbf{u}'_1}{||\mathbf{u}'_1||} & \frac{\mathbf{u}'_2}{||\mathbf{u}'_2||} & \frac{\mathbf{u}'_3}{||\mathbf{u}'_3||} \end{bmatrix} \tag{C.30}$$

Appendix D

Demonstration: Equations of Shahbazpanahi et al. Approach

This appendix is devoted to explain how to rewrite some concrete equations shown in Section 3.2 which make use of the operator $(\underline{\cdot})$, the operator $\text{vec}(\cdot)$ and the Kronecker product. We will start by presenting the properties of this operator.

D.1 The Kronecker Product

The Kronecker product is an special case of the tensor product. The Kronecker product operator between two matrices $\mathbf{A}$ of dimension $M \times N$ and $\mathbf{B}$ of dimension $P \times Q$ is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix} \tag{D.1}$$

Properties

- **Property 1: Bilinearity and Associativity**. Since the Kronecker product is a special case of the tensor product, so it is bilinear and associative. Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be three matrices and k be a scalar. Taking this into account, the Kronecker product obeys the following relationships

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} \tag{D.2}$$

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C} \tag{D.3}$$

$$(k\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (k\mathbf{B}) = k(\mathbf{A} \otimes \mathbf{B}) \tag{D.4}$$

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \tag{D.5}$$

  In general, the Kronecker product is not commutative

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A} \tag{D.6}$$

  But it is permutation equivalent, this means that there exist two permutation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ such that

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{P}_1 (\mathbf{B} \otimes \mathbf{A}) \mathbf{P}_2 \tag{D.7}$$

- **Property 2: Mixed-Product**. There is a relation between the ordinary product of matrices and the Kronecker product, it can be summarized in the following equation

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}) \tag{D.8}$$

  As a consequence of this property, if the matrices $\mathbf{A}$ and $\mathbf{B}$ have an inverse, then it is possible to establish the following relationship

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{D.9}$$

- **Property 3: Transpose**. The operation of transposition is distributive over the Kronecker product, i.e.

$$(\mathbf{A} \otimes \mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}} \tag{D.10}$$

- **Property 4: Spectrum**. If $\mathbf{A}$ and $\mathbf{B}$ are square matrices of rank M and N respectively, let $\delta_1, \delta_2, \ldots, \delta_M$ be the eigenvalues of $\mathbf{A}$ and let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{B}$. Hence, the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are $\delta_i \lambda_j$ where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. Consequently, the trace and determinant of a Kronecker product are given by

$$\mathrm{Tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{B}) \tag{D.11}$$

and

$$\det(\mathbf{A} \otimes \mathbf{B}) = \det(\mathbf{A})^N \det(\mathbf{B})^M \tag{D.12}$$

- **Property 5: Singular Values**. Assuming that $\mathbf{A}$ and $\mathbf{B}$ are rectangular (non-square) matrices and supposing that $\mathbf{A}$ has M non-zero singular values $\delta_1, \delta_2, \ldots, \delta_M$ and $\mathbf{B}$ has N non-zero singular values, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ has $M \cdot N$ non-zero singular values. Since the rank of a matrix is equal to the number of non-zero singular values, we have that

$$\mathrm{rank}(\mathbf{AB}) = \mathrm{rank}(\mathbf{A})\mathrm{rank}(\mathbf{B}) \tag{D.13}$$

- **Property 6: Vector**. Considering that matrix $\mathbf{A}$ has dimension $M \times N$, matrix $\mathbf{B}$ has dimension $N \times P$ and matrix $\mathbf{C}$ has dimension $P \times M$, it is interesting to remark the following property

$$\mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{BC}) = \mathrm{vec}^{\mathrm{T}}(\mathbf{A})(\mathbf{I}_N \otimes \mathbf{B})\mathrm{vec}(\mathbf{C}) \tag{D.14}$$

D.2 Demonstration of Equation (3.29)

We will demonstrate equation (3.29),

$$\mathrm{Tr}(\underline{\mathbf{W}}^{\mathrm{T}}(\hat{\mathbf{h}})\mathbf{R}_{\underline{\mathbf{r}}}\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}) = \mathrm{vec}^{\mathrm{T}}(\underline{\mathbf{W}}(\hat{\mathbf{h}}))(\mathbf{I}_4 \otimes \mathbf{R}_{\underline{\mathbf{r}}})\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}) \tag{D.15}$$

Letting $\mathbf{A} = \underline{\mathbf{W}}(\hat{\mathbf{h}})$, $\mathbf{B} = \mathbf{R}_{\underline{\mathbf{r}}}$ and $\mathbf{C} = \underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}$, equation (D.15) obeys property (D.14), and thus it has been demonstrated.

D.3 Demonstration of Equation (3.35)

We will show how to obtain the expression given in equation (3.35),

$$\text{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})) = \mathbf{\Phi}\underline{\hat{\mathbf{h}}} \tag{D.16}$$

Remember the equation (3.10). We will employ $\hat{\mathbf{h}}$ instead of $\mathbf{h}$, and thus applying the operator $\underline{(\cdot)}$ we obtain

$$\underline{\hat{\mathbf{h}}} = \begin{bmatrix} \Re\{\hat{h}_1\} \\ \Re\{\hat{h}_2\} \\ \Im\{\hat{h}_1\} \\ \Im\{\hat{h}_2\} \end{bmatrix} \tag{D.17}$$

Obeying the definition of the $\underline{\mathcal{D}}_i$ matrices in equation (3.31), we can rewrite them in an expanded version as

$$\underline{\mathcal{D}}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_4, \quad \underline{\mathcal{D}}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

$$\underline{\mathcal{D}}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \qquad \underline{\mathcal{D}}_4 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{D.18}$$

Note that all of these matrices have an element distinct from 0 for every row and every element distinct from 0 is 1 or -1. Thus, from equations (D.17) and (D.18), it is straightforward to obtain

$$\underline{\mathcal{D}}_1\underline{\hat{\mathbf{h}}} = \begin{bmatrix} \Re\{\hat{h}_1\} \\ \Re\{\hat{h}_2\} \\ \Im\{\hat{h}_1\} \\ \Im\{\hat{h}_2\} \end{bmatrix} = \underline{\hat{\mathbf{h}}}, \quad \underline{\mathcal{D}}_2\underline{\hat{\mathbf{h}}} = \begin{bmatrix} \Re\{\hat{h}_2\} \\ \Re\{-\hat{h}_1\} \\ \Im\{\hat{h}_2\} \\ \Im\{-\hat{h}_1\} \end{bmatrix}$$

$$\underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} = \begin{bmatrix} \Im\{-\hat{h}_1\} \\ \Im\{\hat{h}_2\} \\ \Re\{\hat{h}_1\} \\ \Re\{-\hat{h}_2\} \end{bmatrix}, \quad \underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} = \begin{bmatrix} \Im\{-\hat{h}_2\} \\ \Im\{-\hat{h}_1\} \\ \Re\{\hat{h}_2\} \\ \Re\{\hat{h}_1\} \end{bmatrix} \tag{D.19}$$

Replacing $\mathbf{h}$ by $\hat{\mathbf{h}}$ in the expression of equation (3.16), we obtain

$$\underline{\mathbf{W}}(\hat{\mathbf{h}}) = \begin{bmatrix} \Re\{\hat{h}_1\} & \Re\{\hat{h}_2\} & -\Im\{\hat{h}_1\} & -\Im\{\hat{h}_2\} \\ \Re\{\hat{h}_2\} & -\Re\{\hat{h}_1\} & \Im\{\hat{h}_2\} & -\Im\{\hat{h}_1\} \\ \Im\{\hat{h}_1\} & \Im\{\hat{h}_2\} & \Re\{\hat{h}_1\} & \Re\{\hat{h}_2\} \\ \Im\{\hat{h}_2\} & -\Im\{\hat{h}_1\} & -\Re\{\hat{h}_2\} & \Re\{\hat{h}_1\} \end{bmatrix} \tag{D.20}$$

Substituting equations (D.19) in (D.20), it is easy to establish that

$$\underline{\mathbf{W}}(\hat{\mathbf{h}}) = \begin{bmatrix} \underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} \tag{D.21}$$

Thus, if we employ $\mathrm{vec}(\cdot)$ in the equation (D.21), the following expression is obtained

$$\mathrm{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})) = \begin{bmatrix} \underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} \tag{D.22}$$

Defining a new matrix $\mathbf{\Phi}$ from the set of matrices in equation (D.18)

$$\mathbf{\Phi} = \begin{bmatrix} \underline{\mathcal{D}}_1 \\ \underline{\mathcal{D}}_2 \\ \underline{\mathcal{D}}_3 \\ \underline{\mathcal{D}}_4 \end{bmatrix} \tag{D.23}$$

Hence, multiplying on the right side by $\hat{\underline{\mathbf{h}}}$, we obtain

$$\mathbf{\Phi}\hat{\underline{\mathbf{h}}} = \begin{bmatrix} \underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} \\ \underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} \tag{D.24}$$

which is the same one shown in equation (D.22), i.e.

$$\text{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})) = \mathbf{\Phi}\hat{\underline{\mathbf{h}}} \tag{D.25}$$

Thus it has been demonstrated the relationship shown in equation (3.35).

D.4 Demonstration of Equation (3.38)

At this point, the aim is to demonstrate the relationship shown in (3.38), i.e

$$\text{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') = \mathbf{\Psi}\underline{\mathbf{h}} \tag{D.26}$$

With this objective, remember equation (3.22), i.e

$$\mathbf{D}' = \begin{bmatrix} \sigma_{z_1} & 0 & 0 & 0 \\ 0 & \sigma_{z_2} & 0 & 0 \\ 0 & 0 & \sigma_{z_1} & 0 \\ 0 & 0 & 0 & \sigma_{z_2} \end{bmatrix} \tag{D.27}$$

Multiplying on the right side by equation (3.22), the following relationship is obtained

$$\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}' = \begin{bmatrix} \underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} & \underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix}\mathbf{D}' = \begin{bmatrix} \sigma_{z_1}\underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} & \sigma_{z_2}\underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} & \sigma_{z_1}\underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} & \sigma_{z_2}\underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} \tag{D.28}$$

And, thus, applying the operator vec$(\cdot)$ over equation (D.28), it is straightforward to obtain

$$\text{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') = \begin{bmatrix} \sigma_{z_1}\underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} \\ \sigma_{z_2}\underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} \\ \sigma_{z_1}\underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} \\ \sigma_{z_2}\underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} \tag{D.29}$$

Remembering the definition of $\mathbf{\Psi} = (\mathbf{D}' \otimes \mathbf{I}_4)\mathbf{\Phi}$ given in equation (3.37), we can rewrite

equation (D.29) as

$$\begin{bmatrix} \sigma_{z_1}\underline{\mathcal{D}}_1\hat{\underline{\mathbf{h}}} \\ \sigma_{z_2}\underline{\mathcal{D}}_2\hat{\underline{\mathbf{h}}} \\ \sigma_{z_1}\underline{\mathcal{D}}_3\hat{\underline{\mathbf{h}}} \\ \sigma_{z_2}\underline{\mathcal{D}}_4\hat{\underline{\mathbf{h}}} \end{bmatrix} = (\mathbf{D}' \otimes \mathbf{I_4})\mathbf{\Phi}\underline{\mathbf{h}} = \mathbf{\Psi}\underline{\mathbf{h}} \tag{D.30}$$

And finally, from equations (D.29) and (D.30) we obtain

$$\text{vec}(\underline{\mathbf{W}}(\hat{\mathbf{h}})\mathbf{D}') = \mathbf{\Psi}\underline{\mathbf{h}} \tag{D.31}$$

which veryfies the equation (3.38).

Appendix E

HOS Matrices for (2×1) Alamouti Code

In this appendix, we present the formal definition of HOS and its more important properties. We also show the specific form of HOS matrices for the concrete case of the Alamouti code, which have been employed in Chapter 4.

E.1 Higher Order Statistics (HOS) definition

Let $\mathbf{x} = [x_1\ x_2\ \ldots\ x_M]^{\mathrm{T}}$ and $\mathbf{w} = [w_1\ w_2\ \ldots\ w_N]^{\mathrm{T}}$ be two sets of random complex-valued variables. The cumulant of k-order is defined as the coefficients associated to the k-order polynom of the Taylor Series, with respect to $\mathbf{w}$, of the generator function

$$\boldsymbol{\Theta}(\mathbf{w}) = \ln(\boldsymbol{\Phi}(\mathbf{w})) \tag{E.1}$$

where $\boldsymbol{\Phi}(\cdot)$ is the jointly characteristic function of the random variables $\mathbf{x}$

$$\boldsymbol{\Phi}(\mathbf{w}) = \mathrm{E}[\exp(j\mathbf{w}^{\mathrm{T}}\mathbf{x})] \tag{E.2}$$

and the Taylor series expansion provides the elements of $\mathbf{x}$. Thus, it is possible to establish the next relationship for the 2^{nd}, 3^{rd} and 4^{th} order cumulants of a variable $\mathbf{x}$ (of zero mean)

and its moments

$$\begin{aligned} \text{cum}(x_i, x_j) &= \text{E}[x_i x_j] && \text{(E.3)} \\ \text{cum}(x_i, x_j, x_k) &= \text{E}[x_i x_j x_k] && \text{(E.4)} \\ \text{cum}(x_i, x_j, x_k, x_l) &= \text{E}[x_i x_j x_k x_l] - \text{E}[x_i x_j]\text{E}[x_k x_l] - \text{E}[x_i x_l]\text{E}[x_j x_k] - \text{E}[x_i x_k]\text{E}[x_j x_l] && \text{(E.5)} \end{aligned}$$

If the variables do not have zero mean, it is necessary to replace every variable $x_m,\ m \in \{i, j, k, l\}$ by $x_m - \text{E}[x_m]$. Note that equation (E.3) corresponds to the correlation between the variables x_i and x_j. Also, it is important to remark that the kurtosis of x_i is

$$\text{kurt}(x_i) = \text{cum}(x_i, x_i^*, x_i, x_i^*) = \text{E}[|x_i|^4] - 2\text{E}^2[|x_i|^2] - |\text{E}[x_i^2]|^2 \qquad \text{(E.6)}$$

And in the case of a real-valued variable, $x_i = x_i^*$, thus, from equation (E.6) we obtain

$$\text{kurt}(x_i) = \text{cum}(x_i, x_i, x_i, x_i) = \text{E}[|x_i|^4] - 3\text{E}^2[|x_i|^2] \qquad \text{(E.7)}$$

E.2 HOS Properties

- **Property 1.** The cumulants of weighted quantities are equal to the product of all scale factors by the cumulant of the non-scaling quantities

$$\text{cum}(a_1 x_1, \ldots, a_N x_N) = \left(\prod_{i=1}^{N} a_i \right) \text{cum}(x_1, \ldots, x_N) \qquad \text{(E.8)}$$

  where $a_i,\ i =, 1 \ldots, N$ are constants and $x_i,\ i = 1, \ldots, N$ are random variables.

- **Property 2.** The cumulants are symmetric with respect to their arguments

$$\text{cum}(x_1, \ldots, x_N) = \text{cum}(x_{i1}, \ldots, x_{iN}) \qquad \text{(E.9)}$$

  where $i1, \ldots, iN$ is any permutation of the indexes $i = 1, \ldots, N$.

- **Property 3.** The cumulants are unaltered by additive constants, i.e., if K is a constant then

$$\text{cum}(K + x_1, \ldots, x_N) = \text{cum}(x_1, \ldots, x_N) \qquad \text{(E.10)}$$

- **Property 4.** The cumulant of the addition between independent pairs of variables $(x_i, y_i),\ i = 1, \ldots, N$ is equal to the addition of the cumulants of every variable, i.e

$$\text{cum}(x_1 + y_1, \ldots, x_N + y_N) = \text{cum}(x_1, \ldots, x_N) + \text{cum}(y_1, \ldots, y_N) \tag{E.11}$$

- **Property 5.** The cumulant of a set formed by the variables $x_1, \ldots, x_N$ is zero if some variable x_i is independent, i.e.

$$\text{cum}(x_1, \ldots, x_N) = 0 \tag{E.12}$$

- **Property 6**. The conjugate of the cumulant is equal to the cumulant obtained conjugating all the variables, i.e.

$$\text{cum}(x_1, \ldots, x_N) = \text{cum}^*(x_1^*, \ldots, x_N^*) \tag{E.13}$$

E.3 4^{th} Order Cumulants for 2 Variables

We will consider that we have 2 complex-valued variables x_1,x_2 and, therefore, they are 16 different 4^{th} order cumulants

$$\begin{aligned}
c_1 &= \text{cum}(x_1, x_1^*, x_1, x_1^*) & c_9 &= \text{cum}(x_2, x_1^*, x_1, x_1^*) \\
c_2 &= \text{cum}(x_1, x_1^*, x_1, x_2^*) & c_{10} &= \text{cum}(x_2, x_1^*, x_1, x_2^*) \\
c_3 &= \text{cum}(x_1, x_1^*, x_2, x_1^*) & c_{11} &= \text{cum}(x_2, x_1^*, x_2, x_1^*) \\
c_4 &= \text{cum}(x_1, x_1^*, x_2, x_2^*) & c_{12} &= \text{cum}(x_2, x_1^*, x_2, x_2^*) \\
c_5 &= \text{cum}(x_1, x_2^*, x_1, x_1^*) & c_{13} &= \text{cum}(x_2, x_2^*, x_1, x_1^*) \\
c_6 &= \text{cum}(x_1, x_2^*, x_1, x_2^*) & c_{14} &= \text{cum}(x_2, x_2^*, x_1, x_2^*) \\
c_7 &= \text{cum}(x_1, x_2^*, x_2, x_1^*) & c_{15} &= \text{cum}(x_2, x_2^*, x_2, x_1^*) \\
c_8 &= \text{cum}(x_1, x_2^*, x_2, x_2^*) & c_{16} &= \text{cum}(x_2, x_2^*, x_2, x_2^*)
\end{aligned} \tag{E.14}$$

Note that employing the cumulant properties, we can conclude that there are actually only 6 different cumulants and they correspond to

$$\begin{aligned}
c_1 &= \mathrm{cum}(x_1, x_1^*, x_1, x_1^*) \in \mathbb{R} \\
c_2 &= \mathrm{cum}(x_1, x_1^*, x_1, x_2^*) = c_3^* = c_5 = c_9^* \in \mathbb{C} \\
c_4 &= \mathrm{cum}(x_1, x_1^*, x_2, x_2^*) = c_7 = c_{10} = c_{13} \in \mathbb{R} \\
c_6 &= \mathrm{cum}(x_1, x_2^*, x_1, x_2^*) = c_{11}^* \in \mathbb{C} \\
c_8 &= \mathrm{cum}(x_1, x_2^*, x_2, x_2^*) = c_{12}^* = c_{14} = c_{15}^* \in \mathbb{C} \\
c_{16} &= \mathrm{cum}(x_2, x_2^*, x_2, x_2^*) \in \mathbb{R}
\end{aligned} \tag{E.15}$$

Taking this into account, we define the vector $\mathbf{x}$ as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{E.16}$$

and, thus, we can form the following 4 different matrices $\mathbf{C}_{\mathrm{HOS}}^{[k,l]}$ where $k, l = 1, 2$

$$\mathbf{C}_{\mathrm{HOS}}^{[k,l]} = \mathrm{cum}(\mathbf{x}, \mathbf{x}^{\mathrm{H}}, x_k, x_l^*) = \begin{bmatrix} \mathrm{cum}(x_1, x_1^*, x_k, x_l^*) & \mathrm{cum}(x_1, x_2^*, x_k, x_l^*) \\ \mathrm{cum}(x_2, x_1^*, x_k, x_l^*) & \mathrm{cum}(x_2, x_2^*, x_k, x_l^*) \end{bmatrix} \tag{E.17}$$

In order to rewrite the matrix in equation (E.17), we define the matrix $\mathbf{M}$ as

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \tag{E.18}$$

And, we define the matrix $\mathbf{Q}$ for a vector $\mathbf{x}$ and a matrix $\mathbf{M}$ by means of the following expression

$$\begin{aligned}
\mathbf{Q}_{\mathbf{x}}(\mathbf{M}) &= \sum_{k,l=1}^{2} \mathrm{cum}(x_i, x_j^*, x_k, x_l^*) m_{lk} \\
&= m_{11}\mathbf{C}_{\mathrm{HOS}}^{[1,1]} + m_{21}\mathbf{C}_{\mathrm{HOS}}^{[1,2]} + m_{12}\mathbf{C}_{\mathrm{HOS}}^{[2,1]} + m_{22}\mathbf{C}_{\mathrm{HOS}}^{[2,2]}
\end{aligned} \tag{E.19}$$

If we let $\mathbf{M}$ be a matrix of dimmension 2×2 with only one component distinct from 0 and equal to 1, we obtain the following $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ matrices

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{C}_{\text{HOS}}^{[1,1]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_1, x_1) = \begin{bmatrix} c_1 & c_2 \\ c_2^* & c_4 \end{bmatrix} \quad \text{(E.20)}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \Rightarrow \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{C}_{\text{HOS}}^{[1,2]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_1, x_2) = \begin{bmatrix} c_2 & c_6 \\ c_4 & c_8 \end{bmatrix} \quad \text{(E.21)}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{C}_{\text{HOS}}^{[2,1]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_2, x_1) = \begin{bmatrix} c_2^* & c_4 \\ c_6^* & c_8^* \end{bmatrix} \quad \text{(E.22)}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{C}_{\text{HOS}}^{[2,2]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_2, x_2) = \begin{bmatrix} c_4 & c_8 \\ c_8^* & c_{16} \end{bmatrix} \quad \text{(E.23)}$$

Note that this set of 4 matrices contains the 16 cumulants shown in equation (E.14).

E.4 HOS Matrices for (2×1) Alamouti Code

In order to obtain the expression of the 4^{th} order cumulants for the (2×1) Alamouti code, we will define the channel matrix as follows

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & h_1 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 \end{bmatrix} \quad \text{(E.24)}$$

Remember that the observations vector $\mathbf{x}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_1 s_1 + h_2 s_2 + v_1 \\ h_2^* s_1 - h_1 s_2^* + v_2 \end{bmatrix} = \mathbf{H}\mathbf{s} + \mathbf{v} \quad \text{(E.25)}$$

Heretoafter, we will not consider the noise because its 4^{th} order cumulants are equal to zero. Thus, we simplify the equation (E.25), obtaining

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_1 s_1 + h_2 s_2 \\ h_2^* s_1 - h_1 s_2^* \end{bmatrix} = \mathbf{H}\mathbf{s} \quad \text{(E.26)}$$

Hence, substituting the above equation (E.26) in (E.14), we can see that the cumulant c_1 take the following value

$$\begin{aligned} c_1 = & \quad \text{cum}(x_1, x_1^*, x_1, x_1^*) \\ = & \quad \text{cum}(h_1 s_1 + h_2 s_2, h_1^* s_1^* + h_2^* s_2^*, h_1 s_1 + h_2 s_2, h_1^* s_1^* + h_2^* s_2^*) \\ = & \quad \text{cum}(h_1 s_1, h_1^* s_1^*, h_1 s_1, h_1^* s_1^*) + \text{cum}(h_2 s_2, h_2^* s_2^*, h_2 s_2, h_2^* s_2^*) \\ = & \quad h_1 h_1^* h_1 h_1^* \text{cum}(s_1, s_1^*, s_1, s_1^*) + h_2 h_2^* h_2 h_2^* \text{cum}(s_2, s_2^*, s_2, s_2^*) \\ = & \quad |h_1|^4 \text{cum}(s_1, s_1^*, s_1, s_1^*) + |h_2|^4 \text{cum}(s_2, s_2^*, s_2, s_2^*) \end{aligned} \tag{E.27}$$

Considering that ρ_i is the corresponding kurtosis of the variable s_i, $\rho_i = \text{cum}(s_i, s_i^*, s_i, s_i^*)$, and that the variables s_1 and s_2 have the same kurtosis, we can establish that $\rho_1 = \rho_2 = \rho$, and it is possible to simplify the equation (E.27) in

$$c_1 = (|h_1|^4 + |h_2|^4)\rho \tag{E.28}$$

In analogous form, we can obtain the values for the remaining cumulants

$$\begin{aligned} c_1 &= (|h_1|^4 + |h_2|^4)\rho & c_2 &= (|h_1|^2 - |h_2|^2)h_1 h_2 \rho \\ c_4 &= 2|h_1|^2 |h_2|^2 \rho & c_6 &= 2h_1^2 h_2^2 \rho \\ c_8 &= (|h_2|^2 - |h_1|^2)h_1 h_2 \rho & c_{16} &= (|h_1|^4 + |h_2|^4)\rho \end{aligned} \tag{E.29}$$

In order to study the eigenvalue decomposition of the matrix (E.19) in the (2 × 1) Alamouti code context, we rewrite the matrix of equation (E.17) employing the equation (E.26). We obtain

$$\mathbf{C}_{\text{HOS}}^{[k,l]} = \text{cum}(\mathbf{x}, \mathbf{x}^{\text{H}}, x_k, x_l^*) = \text{cum}(\mathbf{H}\mathbf{s}, \mathbf{s}^{\text{H}}\mathbf{H}^{\text{H}}, x_k, x_l^*) = \mathbf{H}\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_k, x_l^*)\mathbf{H}^{\text{H}} \tag{E.30}$$

It is straightforward and interesting to acquire a new version of equation (E.19) employing the equation (E.30) and introducing the term $\breve{\mathbf{\Delta}}_{\mathbf{M}}$. Hence, we obtain

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{H}(m_{11}\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_1, x_1^*) + m_{21}\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_1, x_2^*) + \\ m_{12}\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_2, x_1^*) + m_{22}\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_2, x_2^*))\mathbf{H}^{\text{H}} = \mathbf{H}\breve{\mathbf{\Delta}}_{\mathbf{M}}\mathbf{H}^{\text{H}} \end{aligned} \tag{E.31}$$

Also, replacing x_1, x_1^*, x_2 and x_2^* in equation (E.31) by their corresponding values from equation (E.26), it is possible to isolate the kurtosis of s_1 and s_2 as

$$\begin{aligned}
\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_1, x_1^*) &= \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, h_1 s_1 + h_2 s_2, h_1^* s_1^* + h_2^* s_2^*) \\
&= |h_1|^2 \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, s_1, s_1^*) + |h_2|^2 \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, s_2, s_2^*) \\
&= |h_1|^2 \begin{bmatrix} \rho_1 & 0 \\ 0 & 0 \end{bmatrix} + |h_2|^2 \begin{bmatrix} 0 & 0 \\ 0 & \rho_2 \end{bmatrix} \\
&= |h_1|^2 \mathbf{P_1} + |h_2|^2 \mathbf{P_2}
\end{aligned} \tag{E.32}$$

Remark that the terms $\mathbf{P}_1$ and $\mathbf{P}_2$ are introduced in order to reduce the expression. Equivalently, the remaining terms of cumulants are

$$\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_1, x_2^*) = \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, h_1 s_1 + h_2 s_2, h_2 s_1^* - h_1 s_2^*) = h_1 h_2 \mathbf{P_1} - h_1 h_2 \mathbf{P_2} \tag{E.33}$$

$$\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_2, x_1^*) = \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, h_2^* s_1 - h_1^* s_2, h_1^* s_1^* + h_2^* s_2^*) = h_1^* h_2^* \mathbf{P_1} - h_1^* h_2^* \mathbf{P_2} \tag{E.34}$$

$$\text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, x_2, x_2^*) = \text{cum}(\mathbf{s}, \mathbf{s}^{\text{H}}, h_2^* s_1 - h_1^* s_2, h_2 s_1^* - h_1 s_2^*) = |h_2|^2 \mathbf{P_1} + |h_1|^2 \mathbf{P_2} \tag{E.35}$$

Substituting this set of results formed by equations (E.32), (E.33), (E.34) and (E.35) in equation (E.31), the following expression is obtained

$$\begin{aligned}
\mathbf{Q_x}(\mathbf{M}) = \mathbf{H}((m_{11}|h_1|^2 + m_{21} h_1 h_2 + m_{12} h_1^* h_2^* + m_{22}|h_2|^2)\mathbf{P_1} + \\
(m_{11}|h_2|^2 - m_{21} h_1^* h_2^* - m_{12} h_1 h_2 + m_{22}|h_1|^2)\mathbf{P_2})\mathbf{H}^{\text{H}}
\end{aligned} \tag{E.36}$$

By means of equations (E.24) and (E.18), it is possible to simplify the coefficients of $\mathbf{P}_1$ and $\mathbf{P}_2$. Thus, we achieve

$$\breve{\delta}_1 = \rho_1(m_{11}|h_1|^2 + m_{21} h_1 h_2 + m_{12} h_1^* h_2^* + m_{22}|h_2|^2) = \rho_1 \mathbf{h}_1^{\text{H}} \mathbf{M} \mathbf{h}_1 \tag{E.37}$$

$$\breve{\delta}_2 = \rho_2(m_{11}|h_2|^2 - m_{21} h_1^* h_2^* - m_{12} h_1 h_2 + m_{22}|h_1|^2) = \rho_2 \mathbf{h}_2^{\text{H}} \mathbf{M} \mathbf{h}_2 \tag{E.38}$$

As a result, we have obtained a new form of equation (E.19) for the (2×1) Alamouti code, i.e.

$$\mathbf{Q_x}(\mathbf{M}) = \mathbf{H} \breve{\mathbf{\Delta}}_{\mathbf{M}} \mathbf{H}^{\text{H}} = \mathbf{H} \begin{bmatrix} \breve{\delta}_1 & 0 \\ 0 & \breve{\delta}_2 \end{bmatrix} \mathbf{H}^{\text{H}} \tag{E.39}$$

where

$$\breve{\delta}_i = \breve{\delta}_{ii} = \rho_i \mathbf{h}_i^{\mathrm{H}} \mathbf{M} \mathbf{h}_i \tag{E.40}$$

Note that the EVD of matrix (E.39) corresponds to the following expression

$$\mathbf{Q_x}(\mathbf{M}) = \mathbf{U} \mathbf{\Delta_M} \mathbf{U}^{\mathrm{H}} \tag{E.41}$$

where $\mathbf{U}$ is the unitary matrix of the eigenvectors and $\mathbf{\Delta_M}$ is the diagonal matrix that contains the eigenvectors. Equalizing the equations (E.39) and (E.41) and considering that $||\mathbf{h}||^2 = |h_1|^2 + |h_2|^2$, the following expression is obtained

$$\mathbf{Q_x}(\mathbf{M}) = \mathbf{H} \breve{\mathbf{\Delta}}_\mathbf{M} \mathbf{H}^{\mathrm{H}} = \mathbf{U} \mathbf{\Delta_M} \mathbf{U}^{\mathrm{H}} = \mathbf{U} \left(||\mathbf{h}||^2 \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} \right) \mathbf{U}^{\mathrm{H}} \tag{E.42}$$

Thus, we can conclude that the eigenvectors are determinded by the following expression

$$\mathbf{\Delta_M} = ||\mathbf{h}||^2 \breve{\mathbf{\Delta}}_\mathbf{M} \tag{E.43}$$

For the different $\mathbf{M}$ matrices of the equations (E.20), (E.21), (E.22) and (E.23), we can obtain their corresponding values for matrix $\breve{\mathbf{\Delta}}_\mathbf{M}$ replacing $\mathbf{M}$ in equation (E.40)

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \Delta_\mathbf{M} = \begin{bmatrix} \rho_1 |h_1|^2 & 0 \\ 0 & \rho_2 |h_2|^2 \end{bmatrix} = \rho \begin{bmatrix} |h_1|^2 & 0 \\ 0 & |h_2|^2 \end{bmatrix} \tag{E.44}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \Rightarrow \Delta_\mathbf{M} = \begin{bmatrix} \rho_1 h_1 h_2 & 0 \\ 0 & -\rho_2 h_1 h_2 \end{bmatrix} = \rho \begin{bmatrix} h_1 h_2 & 0 \\ 0 & -h_1 h_2 \end{bmatrix} \tag{E.45}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow \Delta_\mathbf{M} = \begin{bmatrix} \rho_1 h_1^* h_2^* & 0 \\ 0 & -\rho_2 h_1^* h_2^* \end{bmatrix} = \rho \begin{bmatrix} h_1^* h_2^* & 0 \\ 0 & -h_1^* h_2^* \end{bmatrix} \tag{E.46}$$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow \Delta_\mathbf{M} = \begin{bmatrix} \rho_1 |h_2|^2 & 0 \\ 0 & \rho_2 |h_1|^2 \end{bmatrix} = \rho \begin{bmatrix} |h_2|^2 & 0 \\ 0 & |h_1|^2 \end{bmatrix} \tag{E.47}$$

Appendix F

Joint Diagonalization Procedure

The procedure used by JADE for simultaneous diagonalization of several 4^{th} order cross cumulant matrices is an extension of the Jacobi technique: a joint diagonality criterion is iteratively optimized under plane rotations. We have optimized the original code presented in [58] by considering the special structure of the Alamouti code.

For a matrix $\mathbf{C}$ of dimension 2×4 defined as

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix} \tag{F.1}$$

the first step of the procedure consists of obtaining the following 3×2 matrix

$$\mathbf{G} = \begin{bmatrix} c_{11} - c_{22} & c_{13} - c_{24} \\ c_{12} & c_{14} \\ c_{21} & c_{23} \end{bmatrix} \tag{F.2}$$

From this matrix, we obtain $\mathbf{F} = \mathbf{G}\mathbf{G}^{\mathrm{H}}$ and a 3×3 symmetric matrix given by

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{F.3}$$

where

$$e_{11} = f_{11} \tag{F.4}$$

$$e_{12} = e_{21} = real(f_{12} + f_{13}) \tag{F.5}$$

$$e_{13} = e_{31} = imag(-f_{12} + f_{13}) \tag{F.6}$$

$$e_{22} = real(f_{22} + 2f_{23} + f_{33}) \tag{F.7}$$

$$e_{23} = e_{32} = imag(f_{22} + 2f_{23} + f_{33}) \tag{F.8}$$

$$e_{33} = real(f_{22} - 2f_{23} + f_{33}) \tag{F.9}$$

The eigenvalues are computed and then the eigenvector $\mathbf{u}_1$ corresponding to the largest eigenvalue is obtained. If $u_{11} < 0$ then we let $u_1 = -u_1$. Finally, the channel matrix is estimated as

$$\mathbf{H} = \begin{bmatrix} c & -s^* \\ s & c \end{bmatrix} \tag{F.10}$$

where $c = \sqrt{\frac{1+u_{11}}{2}}$ and $s = \frac{u_{21}-u_{31}j}{2c}$.

Note that the method described in Appendix C.2 is a good choice to compute the eigenvalues of the 3×3 matrix $\mathbf{E}$.

Bibliography

[1] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications", *IEEE Journal Select. Areas Communications*, vol. 16, pp. 1451–1458, October 1998.

[2] P. Comon, C. Jutten, "Handbook of Blind Source Separation, Independent Component Analysis and Applications", *Academic Press*, 2010.

[3] S. Shahbazpanahi, A. B. Gershman, J. Manton, "Closed-form Blind MIMO Channel Estimation for Orthogonal Space-Time Block Codes", *IEEE Trans. on Signal Processing*, vol. 53, No. 12, pp. 4506–4516, December 2005.

[4] J. Vía, I. Santamaría, J. Pérez, D. Ramírez, "Blind Decoding of MISO-OSTBC Systems based on Principal Component Analysis", in *Proc. of International Conference on Acoustic, Speech and Signal Processing, 2006*, Vol. IV, pp. 545–549, 2006.

[5] J.-F. Cardoso, A. Souloumiac, "Blind Beamforming for Non-Gaussian Signals", *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] E. Beres, R. Adve, "Blind Channel Estimation for Orthogonal STBC in MISO Systems", in *Proc. of Global Telecommunications Conference, 2004*, vol. 4, pp. 2323–2328, November 2004.

[7] A. Goldsmith, "Wireless Communications", *Cambridge University Press*, 2005.

[8] I. E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *Technical Memorandum, Bell Laboratories, Lucent Technologies*, October 1995.

[9] G. G. Raleigh, J. M. Cioffi, "Spatio Temporal Coding for Wireless Communication", *IEEE Transactions on Communications*, vol. 46, no. 2, March 1998.

[10] G. J. Foschini, M. J. Gans, "On Limits of Wireless Communications in a Fading Environment", *Wireless Communication Magazine*, vol. 6, pp. 311-335, March 1998.

[11] A. Paulraj, C. B. Papadias, "Space-Time Processing for Wireless Communications", *IEEE Signal Processing Magazine*, vol. 14, No. 6, pp. 49–83, November 1997.

[12] D. Gesbert, M. Shafi, D. Shan Shiu, P. J. Smith, A. Naguib, "From Theory to Practice: An Overview of MIMO Space-time Coded Wireless Systems", *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 281–302, 2003.

[13] H. Jafarkhani, *Space Time Coding*, Cambridge University Press, 2005.

[14] C. E. Shannon, "A Mathematical Theory of Communication (Part I)", *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[15] V. Tarokh, N. Seshadri, A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744-765, March 1998.

[16] A. F. Naguib, V. Tarokh, N. Seshadri, A. R. Calderbank, "A Space-Time Coding Modem for High-Data-Rate Wireless", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1459-1478, October 1998.

[17] V. Tarokh, H. Jafarkhani, A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs", *IEEE Transactions on Information Theory*, Vol. 45, No. 5, pp. 1456–1467, July 1999.

[18] E. G. Larsson, P. Stoica, "Space-Time Block Coding for Wireless Communications", *Cambridge University Press*, 2003.

[19] S. Sandhu, A. Paulraj, "Space-Time Block Codes: A Capacity Perspective", *IEEE Communications Letter*, vol.4, no.12, December 2000.

[20] J. G. Andrews, "Fundamentals of WiMAX: Understanding Broadband Wireless Networking", *Prentice Hall Communications Engineering and Emerging Technologies Series*, 2007.

[21] B. L. Hughes, "Differential Space-Time Modulation", *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567-2578, November 2000.

[22] S. Roberts, R. Everson, "Independent Component Analysis: Principles and Practice", *Cambridge University Press*, United Kingdom, 2001.

[23] P. Comon, "Blind Identification and Source Separation in 2×3 Under-Determined Mixtures", *Transactions on Signal Processing*, vol. 52, no. 1, pp. 11–22, January 2004.

[24] A. Hyvärinen, J. Karhunen, E. Oja, "Independent Component Analysis". New York: John Wiley & Sons, 2001.

[25] J. Karhunen. "Neural Approaches to Independent Component Analysis and Source Separation", in *Proc. ESANN'96*, pp. 249–266, 1996.

[26] V. Zarzoso, A. K. Nandi, "Blind Source Separation", in *Blind Estimation Using Higher-Order Statistics*, A. K. Nandi, Ed. Boston, MA: Kluwer Academic Publishers, ch. 4, pp. 167–252, 1999.

[27] J. Vía, I. Santamaría, J. Pérez, "A Sufficient Condition for Blind Identifiability of MIMO-OSTBC Channels Based on Second Order Statistics", in *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications, 2006.* SPAWC'06. pp. 1–5, 2-5, July 2006.

[28] H. J. Pérez-Iglesias, J. A. García-Naya, A. Dapena, L. Castedo, V. Zarzoso, "Blind Channel Identification in Alamouti Coded Systems: A Comparative Study of Eigen-

decomposition Methods in Indoor Transmissions at 2.4 GHz", in *European Transactions on Telecomunications. Special Issue: European Wireless 2007*, vol. 19, Issue 7, pp. 751–759, November 2008.

[29] G. H. Golub, C. F. Van Loan, "Matrix Computations", *Johns Hopkins Studies in Mathematical Sciences*, 3rd Edition, 1996.

[30] V. Tarokh, H. Jafarkhani, A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results", *IEEE Journal on Select Areas in Communications*, Vol.17, No. 3, pp. 451–460, March 1999.

[31] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications", *Cambridge, Technology Press of Massachusetts Institute of Technology*, New York, Wiley, 1949.

[32] A. Hyvärinen, E. Oja, "Independent Component Analysis: A Tutorial", *Neural Networks*, vol. 13 (4–5), pp. 411–430, 2000.

[33] X. R. Cao, R. W. Liu, "General Approach to Blind Source Separation", *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 562–571, March 1996.

[34] P. Comon, C. Jutten, J. Hérault, "Blind Separation of Sources, Part II: Problems Statement", *Signal Processing*, vol. 24, no. 1, pp. 11–20, July 1991.

[35] A. Cichocki, S. Amari, "Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications", *John Wiley & Sons*, England, 2002.

[36] J. Hérault, C. Jutten, B. Ans, "Détection de Grandeurs Primitives dans un Message Composite par une Architecture Neuromimétique en Apprentissage Non Supervisé", in *GRETSI'85 Proceedings*, Nice, France, May 20–24, 1985, pp. 1017–1022.

[37] G. Deco, W. Brauer, "Nonlinear High-Order Statistical Decorrelation by Volume-Conserving Neural Architectures", *Neural Networks*, vol. 8, no. 4, pp. 525–535, April 1995.

[38] A. J. Bell, T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation* vol. 7, no. 6, pp. 1129–1159, 1995.

[39] S.-I. Amari, "Gradient Learning in Structured Parameter Spaces: Adaptive Blind Separation of Signal Sources", *In Proceedings WCNN96*, pp. 951–956, San Diego, 1996.

[40] C. Mejuto, L. Castedo, "A Neural Network Approach to Blind Source Separation", in *Proc. Neural Networks for Signal Processing VII*, pp. 486–595, Florida, USA, September 1997.

[41] A. Hyvärinen, E. Oja. "A Fast Fixed-Point Algorithm for Independent Component Analysis", *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[42] N. Delfosse, P. Loubaton, "Adaptive Blind Separation of Independent Sources: A Deflation Approach", *Signal Processing*, vol. 45, pp. 59–83, 1995.

[43] A. Hyvärinen, E. Oja, "A Neuron that Learns to Separate One Signal from a Mixture of independent sources", in *Proc. IEEE International Conference on Neural Networks*, vol. 1, pp. 62–67 vol. 1, 3–6 June 1996.

[44] A. Benveniste, M. Métivier, P. Priourent, "Adaptive Algorithms and Stochastics Approximations". *Springer-Verlag*, New York, 1990.

[45] E. Bingham, A. Hyvärinen, "A Fast Fixed-Point Algorithm for Independent Component Analysis of Complex-Valued Signals". *International Journal of Neural Systems*, 10(1):1-8, 2000.

[46] S. Haykin, "Neural Networks: A Comprehensive Foundation", *MacMillan Publishing Company*, 1st Edition, 1994.

[47] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine*, series 6, vol 2: 559–72, November 1901.

[48] Yu-Hen Hu, Jenq-Neng Hwang, "Handbook of Neural Network Signal Processing", *CRC Press*, 2002.

[49] J. A. García-Naya, T. Fernández-Caramés, H. J. Pérez-Iglesias, M. González-López, L. Castedo, D. Ramírez, I. Santamaría, J. Pérez-Arriaga, J. Vía, J. M. Torres-Royo, "Performance of STBC Transmissions with Real Data", *16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, July 2007.

[50] O. Macchi, "Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission", *John Wiley & Sons*, New York, 1995.

[51] H. Chernoff, E. L. Lehmann, "The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness-of-Fit", *The Annals of Mathematical Statistics 25*, pp. 579–586, 1954.

[52] C. Lélé, Pierre Siohan, R. Legouable, "The Alamouti Scheme with CDMA-OFDM/OQAM", *EURASIP Journal on Advances in Signal Processing*, 2010.

[53] K. Fang, G. Leus, L. Rugini, "Alamouti Space-Time Coded OFDM Systems in Time- and Frequency-Selective Channels", *IEEE Globecom 2006*.

[54] A. R. S. Bahai, B. R. Saltzberg, M. Ergen, "Multi-carrier Digital Communications: Theory and Applications of OFDM", Springer, 2004.

[55] D. Iglesia, A. Dapena, C.J. Escudero, "Multiuser Detection in MIMO OFDM Systems Using Blind Source Separation", *Proceedings of the Sixth Baiona Workshop on Signal Processing in Communications*, 2003.

[56] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels", Bell Labs, Lucent Technologies, I. Tech. Rep., Oct. 1995, in *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November- December 1999.

[57] G. Cardano, "Ars Magna or The Rules of Algebra", New York: *Dover*, 1993.

[58] J.-F. Cardoso, A. Souloumiac, "Jacobi Angles for Simultaneous Diagonalization", *SIAM Journal on Matrics Analysis and Applications*, vol. 17, no. 1, pp. 161–164, 1996.