Helsinki University of Technology
Department of Signal Processing and Acoustics
Espoo 2008 Report 6

ADVANCED OPTIMIZATION ALGORITHMS FOR SENSOR ARRAYS AND MULTI-ANTENNA COMMUNICATIONS

Traian Abrudan

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Signal Processing and Acoustics for public examination and debate in Auditorium S2 at Helsinki University of Technology (Espoo, Finland) on the 21st of November, 2008, at 12 o'clock noon.

Helsinki University of Technology
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics

Teknillinen korkeakoulu
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Signaalinkäsittelyn ja akustiikan laitos

Distribution:
Helsinki University of Technology
Department of Signal Processing and Acoustics
P.O. Box 3000
FIN-02015 HUT
Tel. +358-9-451 3211
Fax. +358-9-452 3614
E-mail: Mirja.Lemetyinen@hut.fi
Web page: http://signal.hut.fi

ISBN 978-951-22-9606-4 (Printed)
ISBN 978-951-22-9607-1 (Electronic)
ISSN 1797-4267

Multiprint Oy
Espoo 2008

ABSTRACT OF DOCTORAL DISSERTATION	HELSINKI UNIVERSITY OF TECHNOLOGY P. O. BOX 1000, FI-02015 TKK http://www.tkk.fi

Author Traian Abrudan

Name of the dissertation

Advanced Optimization Algorithms for Sensor Arrays and Multi-antenna Communications

Manuscript submitted August 1, 2008	Manuscript revised October 10, 2008
Date of the defence November 21, 2008	
☐ Monograph	☒ Article dissertation (summary + original articles)

Faculty	Faculty of Electronics, Communications and Automation
Department	Department of Signal Processing and Acoustics
Field of research	Sensor Array Signal Processing
Opponent(s)	Prof. Athina Petropulu (Drexel University) and Prof. Erik G. Larsson, (Linköping University)
Supervisor	Prof. Visa Koivunen
Instructor	Prof. Visa Koivunen

Abstract

Optimization problems arise frequently in sensor array and multi-channel signal processing applications. Often, optimization needs to be performed subject to a matrix constraint. In particular, unitary matrices play a crucial role in communications and sensor array signal processing. They are involved in almost all modern multi-antenna transceiver techniques, as well as sensor array applications in biomedicine, machine learning and vision, astronomy and radars.

In this thesis, algorithms for optimization under unitary matrix constraint stemming from Riemannian geometry are developed. Steepest descent (SD) and conjugate gradient (CG) algorithms operating on the Lie group of unitary matrices are derived. They have the ability to find the optimal solution in a numerically efficient manner and satisfy the constraint accurately. Novel line search methods specially tailored for this type of optimization are also introduced. The proposed approaches exploit the geometrical properties of the constraint space in order to reduce the computational complexity.

Array and multi-channel signal processing techniques are key technologies in wireless communication systems. High capacity and link reliability may be achieved by using multiple transmit and receive antennas. Combining multi-antenna techniques with multicarrier transmission leads to high the spectral efficiency and helps to cope with severe multipath propagation.

The problem of channel equalization in MIMO-OFDM systems is also addressed in this thesis. A blind algorithm that optimizes of a combined criterion in order to be cancel both inter-symbol and co-channel interference is proposed. The algorithm local converge properties are established as well.

Keywords Optimization, unitary matrix, array signal processing, blind separation, equalization, MIMO, OFDM

ISBN (printed) 978-951-22-9606-4	ISSN (printed) 1797-4267
ISBN (pdf) 978-951-22-9607-1	ISSN (pdf)
Language English	Number of pages 95p. + app. 94p.

Publisher Helsinki University of Technology, Department of Signal Processing and Acoustics

Print distribution Helsinki University of Technology, Department of Signal Processing and Acoustics

☒ The dissertation can be read at http://lib.tkk.fi/Diss/2008/isbn9789512296071

VÄITÖSKIRJAN TIIVISTELMÄ	TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi
Tekijä Traian Abrudan	
Väitöskirjan nimi Kehittyneet optimointialgoritmit antenniryhmien signaalinkäsittelyssä ja moniantennikommunikaatiossa	
Käsikirjoituksen päivämäärä 01.08.2008	Korjatun käsikirjoituksen päivämäärä 10.10.2008
Väitöstilaisuuden ajankohta 21.11.2008	
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)

Tiedekunta	Elektroniikan, tietoliikenteen ja automaation tiedekunta
Laitos	Signaalinkäsittelyn ja akustiikan laitos
Tutkimusala	Tietoliikenteen signaalinkäsittely
Vastaväittäjä(t)	Prof. Athina Petropulu (Drexel University) and Prof. Erik G. Larsson, (Linköping University)
Työn valvoja	Prof. Visa Koivunen
Työn ohjaaja	Prof. Visa Koivunen

Tiivistelmä

Sensori- ja antenniryhmien sekä monikanavaisten signaalien käsittelyssä esiintyy usein optimointiongelmia, joissa on rajoitteita optimoitaville parametreille. Monikanavaisten signaalien tapauksessa rajoitteet kohdistuvat tyypillisesti matriiseihin, ja erityisesti unitaariset matriisit ovat keskeisessä osassa moniantennitietoliikenteen sekä sensoriryhmien biolääketieteellissä sovelluksissa, tutkajärjestelmissä, koneoppimisessa ja radioastronomiassa.

Tässä väitöskirjassa on kehitetty algoritmeja optimointiin unitaarisuusrajoituksen alla. Kehitetyt jyrkimmän laskeuman ja liittogradientti menetelmät perustuvat Riemannin geometriaan ja käyttävät hyväkseen unitaaristen matriisien Lien ryhmän ominaisuuksia. Kehitetyt menetelmät löytävät optimiratkaisun tehokkaasti ja toteuttavat unitaarisuusrajoituksen tarkasti. Unitaariseen viivahakuun on myös kehitetty tehokkaita hakualgoritmeja, jotka hyödyntävät rajoiteavaruuden rakennetta laskentatarpeen vähentämiseksi.

Moniantenniteknologiat ovat keskeisessä osassa tulevaisuuden laajakaistaisissa langattomissa tietoliikennejärjestelmissä. Niiden avulla voidaan radiolinkkien kapasiteettia ja laatua parantaa merkittävästi. Käytettäessä lisäksi monikantoaaltotekniikoita saavutetaan erinomainen spektritehokkuus sekä luotettava toiminta huolimatta radiokanavan monitie-etenemisestä.

Tässä työssä on johdettu sokea kanavakorjainalgoritmi moniantenni ja -kantoaaltotekniikkaan perustuviin MIMO-OFDM vastaanottimiin. Kehitetty algoritmi kumoaa symbolien välisen keskinäisvaikutuksen sekä samalla kanavalla esiintyvän interferenssin käyttäen rajoitettua optimointia. Työssä osoitetaan kehitetyn algoritmin suppenevan paikalliseen optimiratkaisuun.

Asiasanat Optimointi, unitaari matriisi, antenniryhmien signaalinkasittely, sokea signaalien erottelu, ekvalisaatio, MIMO, OFDM

ISBN (painettu) 978-951-22-9606-4	ISSN (painettu) 1797-4267
ISBN (pdf) 978-951-22-9607-1	ISSN (pdf)
Kieli Englanti	Sivumäärä 95s. + liit. 94s.
Julkaisija Teknillinen korkeakoulu, Signaalinkäsittelyn ja akustiikan laitos	
Painetun väitöskirjan jakelu Teknillinen korkeakoulu, Signaalinkäsittelyn ja akustiikan laitos	
☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/2008/isbn9789512296071	

Acknowledgements

The research work for this doctoral thesis was carried out at the Department of Signal Processing and Acoustics, Helsinki University of Technology, during the years 2001–2008. The Statistical Signal Processing group led by Prof. Visa Koivunen is part of SMARAD (Smart and Novel Radios Research Unit) Centre of Excellence in research nominated by the Academy of Finland.

First, I wish to express my sincere gratitude to my supervisor, Prof. Visa Koivunen, for his continuous support and encouragement during the course of this work. His guidance was crucial and helped me overcome many obstacles I encountered in my research work. It has been an honour to work with such a dedicated scientist and outstanding group leader.

I am also grateful to all my co-workers, especially Dr. Jan Eriksson and Dr. Marius Sîrbu, with whom I co-authored several publications. Their comments, suggestions and constructive criticism have greatly contributed to the technical quality of the thesis.

I would like to thank my thesis pre-examiners, Prof. Corneliu Rusu and Prof. Keijo Ruotsalainen for their comments and for the effort they have put in revising the manuscript. Furthermore, Prof. Iiro Hartimo, the former director of GETA Graduate School in Electronics, Telecommunications and Automation, and Marja Leppäharju, the GETA coordinator are highly acknowledged. The department secretaries Mirja Lemetyinen and Anne Jääskeläinen deserve many thanks for assisting with all the practical issues and arrangements.

This research was funded by the Academy of Finland and the GETA graduate school. I use this opportunity to thank Nokia Foundation, Jenny and Antti Wihuri Foundation, and Elisa Foundation for the financial support they provided during my studies.

I would like to thank all my colleagues in the department for the interesting discussions, especially Dr. Timo Roman, Dr. Mihai Enescu, Eduardo Zacarías, Karol Schober, Dr. Fabio Belloni, Jussi Salmi, Mário Costa, Dr. Stefan Werner, Dr. Cássio Ribeiro, Dr. Andreas Richter, Tuomas Aittomäki, Mei Yen Cheong, and Prof. Risto Wichman.

Special thanks go to my dear friends from Otaniemi for the fantastic times we spent together. I will remember these wonderful years all my life.

I also thank all my Romanian friends in Finland who made me feel not so far from home. Futhermore, I am very grateful to my good friends in my home city Cluj-Napoca, as well as my dear old friends from my childhood place Feleacu, for the memorable holidays we spent together. They certainly contributed to my good mood throughout all these years.

Finally, I am deeply grateful and I dedicate this thesis to my parents Traian and Leontina, and my sister Luiza, and thank them for their love and irreplaceable support. In final, dedic această lucrare parinţilor mei Traian şi Leontina, şi surorii mele Luiza şi le mulţumesc din suflet pentru dragostea dăruită şi pentru suportul lor de neînlocuit. I am grateful to my love, Orquidea Ribeiro, for bringing immeasurable happiness into my life. I thank her from the heart for support and encouragement. Estou grato ao meu amor, Orquidea Ribeiro, pela imensurável felicidade que deu à minha vida. Agradeço-lhe, do fundo do coração, pelo apoio e encorajamento.

Espoo, October 2008

Traian Abrudan

Contents

List of original publications

(I) T. Abrudan, J. Eriksson, V. Koivunen, "Steepest Descent Algorithm for Optimization under Unitary Matrix Constraint", *IEEE Transactions on Signal Processing* vol. 56, no. 3, Mar. 2008, pp. 1134–1147.

(II) T. Abrudan, J. Eriksson, V. Koivunen, "Conjugate Gradient Algorithm for Optimization Under Unitary Matrix Constraint", Submitted for publication. Material presented also in the technical report: "Conjugate Gradient Algorithm for Optimization Under Unitary Matrix Constraint". *Technical Report 4/2008*, Department of Signal Processing and Acoustics, Helsinki University of Technology, 2008. ISBN 978-951-22-9483-1, ISSN 1797-4267.

(III) T. Abrudan, J. Eriksson, V. Koivunen, "Optimization under Unitary Matrix Constraint using Approximate Matrix Exponential", *Conference Record of the Thirty Ninth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 28 Oct.–1 Nov. 2005, pp. 242–246.

(IV) T. Abrudan, J. Eriksson, V. Koivunen, "Efficient Line Search Methods for Riemannian Optimization Under Unitary Matrix Constraint", *Conference Record of the Forty-First Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 4–7 Nov. 2007, pp. 671–675.

(V) T. Abrudan, J. Eriksson, V. Koivunen, "Efficient Riemannian Algorithms for Optimization Under Unitary Matrix Constraint", *IEEE International Conference on Acoustics Speech and Signal Processing* Las Vegas, NV, 31 Mar.– 4 Apr. 2008, pp. 2353–2356.

(VI) T. Abrudan, M. Sîrbu, V. Koivunen, "Blind Multi-user Receiver for MIMO-OFDM Systems", *IEEE Workshop on Signal Processing Advances in Wireless Communications*, Rome, Italy, 15–18 Jun. 2003, pp. 363–367.

(VII) T. Abrudan, V. Koivunen, "Blind Equalization in Spatial Multiplexing MIMO-OFDM Systems based on Vector CMA and Decorrelation Criteria", *Wireless Personal Communications*, vol. 43, no. 4, Dec. 2007, pp. 1151–1172.

List of abbreviations

[Publication p]	pth original publication
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
ADSL	Asymmetric Digital Subscriber Line
ANSI	American National Standards Institute
ARIB	Association of Radio Industries and Businesses
B3G	Beyond 3G
BLAST	Bell Labs lAyered Space-Time architecture
BPSK	Binary Phase Shift Keying
BRAN	Broadband Radio Access Networks
BSS	Blind Source Separation
CCI	Co-Channel Interference
CDMA	Code Division Multiplexing Access
CG	Conjugate Gradient
CM	Constant Modulus
CMA	Constant Modulus Algorithm
CP	Cyclic Prefix
CR	Cognitive Radios
CRLB	Cramér-Rao Lower Bound
CSI	Channel State Information
DAB	Digital Audio Broadcast
DFT	Discrete Fourier Transform
DVB	Digital Video Broadcast
DVB-H	Digital Video Broadcast - Handheld
DVB-T	Digital Video Broadcast - Terrestrial
E-UTRA	Evolved Universal Terrestrial Radio Access
EASI	EquivAriant source Separation via Independence
ED	Eigen-Decomposition
EEG	ElectroEncephaloGraphy
EKG	ElektroKardioGramm, known also as ECG (ElectroCardioGram)
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix

flop	floating point complex operation
GCEIR	Global Channel-Equalizer Impulse Response
HIPERLAN	HIgh PERformance Radio Local Area Networks
HOS	Higher-Order Statistics
ICA	Independent Component Analysis
IEEE	Institute of Electrical and Electronics Engineers
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
IMT-2000	International Mobile Telecommunications-2000
ISA	Independent Subspace Analysis
ISI	Inter-Symbol Interference
IVLB	Intrinsic Variance Lower Bound
i.i.d.	independent and identically distributed
JADE	Joint Approximate Diagonalization of Eigen-matrices
JD	Joint Diagonalization
LAN	Local Area Network
LDPC	Low-Density Parity-Check
LS	Least Squares
LTE	Long Term Evolution
MCA	Minor Component Analysis
MEG	MagnetoEncephaloGraphy
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
MSE	Mean-Square Error
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
PAST	Projection Approximation Subspace Tracking
PASTd	Projection Approximation Subspace Tracking with deflation
PHY	PHYsical layer
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RADICAL	Robust Accurate Direct ICA aLgorithm
RQI	Rayleigh Quotient Iteration
RX	receiver
SA	Steepest Ascent
SD	Steepest Descent
SDR	Software Defined Radio
SIC	Successive Interference Cancellation
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-output
SOCS	Second-Order Cyclostationarity Statistics
SINR	Signal-to-Interference-plus-Noise Ratio

SNR	Signal-to-Noise Ratio
SOS	Second-Order Statistics
SVD	Singular Value Decomposition
TX	transmitter
UMTS	Universal Mobile Telecommunications System
V-BLAST	Vertical-BLAST
VCM	Vector Constant Modulus
VCMA	Vector Constant Modulus Algorithm
VSC	Virtual Sub-Carriers
ZF	Zero Forcing
ZP	Zero Padding
WiMAX	Worldwide Inter-operability for Microwave ACceSs
WLAN	Wireless Local Area Network
WSS	Wide-Sense Stationary
w.r.t.	with respect to

List of symbols

a, A	complex-valued scalars
$\mathbf{a}$	complex-valued vector
$\mathbf{A}$	complex-valued matrix
$\mathcal{A}$	real-valued scalar function
$\boldsymbol{\mathcal{A}}$	complex-valued matrix function
$\mathbf{A}^*$	conjugate of matrix $\mathbf{A}$
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
min	minimum value
max	maximum value
arg min	minimizing argument
arg max	maximizing argument
$E[\mathbf{A}]$	expected value of matrix $\mathbf{A}$
$\text{trace}\{\mathbf{A}\}$	trace of matrix $\mathbf{A}$
sign	sign operator
modulo	modulo operation
$\circ$	function composition operation
$\triangleq$	defined as
$:=$	operation attributing the value of the right operand to the left operand
$==$	equality testing
$\neq$	not equal
$\jmath$	imaginary unit
$\Re\{\mathbf{A}\}$	real part of a matrix $\mathbf{A}$
$\Im\{\mathbf{A}\}$	imaginary part of a matrix $\mathbf{A}$
$\angle a$	angle of complex scalar a
$\lvert a \rvert$	absolute value of complex scalar a
$\lVert\mathbf{A}\rVert_F$	Frobenius norm of a matrix $\mathbf{A}$
$\langle\mathbf{A},\mathbf{B}\rangle_E$	Euclidean inner product between $\mathbf{A}$ and $\mathbf{B}$
$\langle\mathbf{A},\mathbf{B}\rangle_{\mathbf{W}}$	Riemannian inner product between $\mathbf{A}$ and $\mathbf{B}$ at $\mathbf{W}$
$\frac{\partial\mathcal{J}}{\partial\mathbf{A}}$	partial derivative of function $\mathcal{J}$ w.r.t. matrix $\mathbf{A}$
∇^{E}	Euclidean gradient
∇^{R}	Riemannian gradient

$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{R}^n$	n-dimensional real vector space
$\mathbb{C}^n$	n-dimensional complex vector space
$\mathbb{R}^{n\times p}$	set of $n \times p$ real matrices
$\mathbb{C}^{n\times p}$	set of $n \times p$ complex matrices
$GL(n)$	general linear group (Lie group of $n \times n$ invertible matrices)
$SL(n)$	special linear group (Lie group of $n \times n$ matrices with unit determinant)
$O(n)$	orthogonal group (Lie group of $n \times n$ orthogonal matrices)
$SO(n)$	special orthogonal group (Lie group of $n \times n$ orthogonal matrices with unit determinant)
$U(n)$	unitary group (Lie group $n \times n$ unitary matrices)
$SU(n)$	special unitary group (Lie group of $n \times n$ unitary matrices with unit determinant)
$\mathrm{St}(n,p)$	Stiefel manifold (the set of $n \times p$ orthonormal matrices)
$\mathrm{Gr}(n,p)$	Grassmann manifold (the set of p-dimensional subspaces on the n-dimensional Euclidean space)
exp	standard matrix exponential
$\mathfrak{so}(n)$	Lie algebra of $SO(n)$
$\mathfrak{u}(n)$	Lie algebra of $U(n)$
$\mathbf{A}_R$	real part of matrix $\mathbf{A}$
$\mathbf{A}_I$	imaginary part of matrix $\mathbf{A}$
$\mathbf{A}_{\mathbb{R}}$	block-matrix build from the real and imaginary parts of matrix $\mathbf{A}$
$\mathbf{a}_p$	GCEIR corresponding to the pth transmitted data stream
$b_1, \ldots b_q$	complex coefficients
$\boldsymbol{\mathcal{C}}(t)$	curve parametrized by t, on a differentiable manifold
$\boldsymbol{\mathcal{C}}'(t)$	first-order derivative of $\boldsymbol{\mathcal{C}}(t)$ w.r.t. parameter t
$\bar{\mathbf{C}}$	MIMO channel matrix
$\mathbf{C}_{pq}$	Sylvester matrix containing the coefficients of the (l,p) MIMO branch
$\mathbf{c}_{pq}$	vector of coefficients of the (l,p) MIMO channel branch
c_{pq}	coefficient of the (l,p) MIMO channel branch
$\mathbf{D}_k$	diagonal matrix with eigenvalues of a skew-Hermitian matrix, at iteration k
d	integer delay
$d_1,\ d_2$	range limits of the integer delay
$\mathbf{E}_{qp}$	Sylvester matrix containing the coefficients of the sub-equalizer (q,p)
$\mathbf{e}_p$	vector with coefficients of the equalizer of the pth data stream
$\mathbf{e}_{qp}$	vector with coefficients of the pth sub-equalizer from qth receive antenna
e_{qp}	coefficient of the pth sub-equalizer from qth receive antenna
$\mathbf{F}$	normalized IDFT matrix
$\mathcal{F}$	almost periodic function
$\mathbf{G}_k$	Riemannian gradient translated to identity, at iteration k
$\tilde{\mathbf{G}}_k$	Riemannian gradient, at iteration k

$\mathbf{H}$	Riemannian search direction
$\mathbf{H}_k$	Riemannian search direction translated to identity, at iteration k
$\tilde{\mathbf{H}}_k$	Riemannian search direction at iteration k
$\mathbf{H}_k^{\mathrm{E}}$	Euclidean ascent direction at iteration k
$\mathbf{H}_k^{\mathrm{R}}$	Riemannian ascent direction at iteration k
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{I}$	group identity element
$\mathcal{J}$	differentiable cost function
$\hat{\mathcal{J}}$	cost function along a curve on the manifold
$\hat{\mathcal{J}}'$	first-order derivative of the cost function along a curve on the manifold
$\mathcal{J}_{\mathrm{geod}}$	cost function along a geodesic curve on $U(n)$
$\mathcal{J}_{\mathrm{proj}}$	cost function along a curve on $U(n)$ described by the projection operator
$\mathcal{J}_{\mathrm{Cayley}}$	cost function along a curve on $U(n)$ described by the Cayley transform
$\mathcal{J}^{\mathrm{VCMA}}$	VCMA cost function
$\mathcal{J}^{\mathrm{xcorr}}$	cross-correlation cost function
k	time instance or iteration number
$\mathcal{L}$	Lagrangian function
L	order of the global channel-equalizer filter
L_{c}	maximum channel order
L_{e}	equalizer order
M	number of subcarriers
N	length of the OFDM block
N_{a}	order of the approximation of the first-order derivative of the cost function
n	number of rows of an matrix with orthonormal columns
$\mathcal{O}(n^p)$	function of n, such that $\lim_{n\to\infty} \lvert n^{-p}\mathcal{O}(n^p)\rvert < \infty$
P	number of transmit antennas
$\boldsymbol{\mathcal{P}}\{\mathbf{A}\}$	operator that projects an arbitrary matrix $\mathbf{A}$ into the unitary group
$\hat{\boldsymbol{\mathcal{P}}}$	local parametrization arising from a projection operator
p	natural number
q	natural number
Q	number of receive antennas
$\mathbf{R}_k$	unitary rotation, at iteration k
$\mathbf{R}_{pl}$	cross-correlation matrix between the pth and lth equalized outputs
r	order of the cost function
r_2	energy dispersion constant
$\mathbf{s}_p$	block of constellation symbols of the pth transmitted data stream
$\hat{\mathbf{s}}_p$	estimated block of constellation symbols of the pth transmitted data stream
s_p	constellation symbol of the pth transmitted data stream
T	almost period
$T_{\mathbf{W}}$	tangent space at point $\mathbf{W}$
$\mathbf{T}_{\mathrm{CP}}$	cyclic prefix addition matrix
t	real parameter
t_{a}	approximation range of the first-order derivative of the cost function

$\mathcal{U}$	unitarity criterion
$\mathbf{U}_k^{\mathrm{E}}$	Euclidean gradient of the unitarity criterion at iteration k
$\mathbf{U}_k$	eigenvectors of a skew-Hermitian matrix, at iteration k
$\mathbf{u}$	vector obtained by stacking the P transmitted OFDM blocks
$\tilde{\mathbf{u}}_p$	OFDM block corresponding to the pth transmitted data stream
$\mathbf{u}_p$	vector of samples corresponding to the pth transmitted data stream
u_p	sub-symbol within the OFDM block of the pth transmitted data stream
$\mathbf{v}_q$	noise vector at the qth receive antenna
$\mathbf{v}$	noise vector obtained by stacking the Q noise vectors
$\boldsymbol{\mathcal{W}}$	local parametrization on the unitary group
$\boldsymbol{\mathcal{W}}_{\mathrm{geod}}$	geodesic curve on the unitary group
$\boldsymbol{\mathcal{W}}_{\mathrm{proj}}$	curve on the unitary group described by the projection operator
$\boldsymbol{\mathcal{W}}_{\mathrm{Cayley}}$	curve on the unitary group described by the Cayley transform
$\mathbf{W}$	unitary matrix
$\mathbf{W}_k$	unitary matrix, at iteration k
$\tilde{\mathbf{W}}_k$	matrix close to unitary, at iteration k
$\mathbf{w}$	unit-norm complex vector
w	unit-length complex number
$\mathbf{y}$	vector obtained by stacking Q received data vectors
$\mathbf{y}_q$	received data vector at the qth receive antenna
$\tilde{\mathbf{y}}_q$	enlarged received OFDM block at the qth receive antenna
$\mathbf{Z}$	arbitrary $n \times n$ complex matrix
$\mathbf{z}_p$	equalized data vector corresponding to the pth data stream
z_p	equalized data sample corresponding to the pth data stream
γ_k	weighting constant of the previous conjugate direction, at iteration $k+1$
$\boldsymbol{\delta}_i$	vector which has one unit element on position i and zero in rest
ϵ	small real number
θ	arbitrary phase rotation
λ	weighting constant of the composite cost function
λ_i	Lagrange multipliers
λ_k	weighting constant of the extra-penalty function, at iteration k
μ	step size parameter
μ_k	step size parameter at iteration k
$\tau\mathbf{H}$	parallel transport of the tangent vector $\mathbf{H}$ along a geodesic
$\omega_1, \ldots, \omega_n$	imaginary parts of the eigenvalues of a skew-Hermitian matrix
$\omega_{\max}$	dominant eigenvalue of a skew-Hermitian matrix

Chapter 1

Introduction

1.1 Motivation of the thesis

The field of multi-channel and array signal processing develops and applies powerful mathematical and statistical techniques that process multi-channel signals. Typically, *space* and *time* are employed as explaining variables. Spatial dimension is enabled by using multiple sensors, or multiple emitters located at different positions. Sensor array signal processing algorithms have the ability to fuse data collected at several sensors in order to perform a given estimation task [125]. Conversely, by using multiple transmitters, special space-time structure may be constructed into the transmitted signals depending on the task at hand. This type of array signal processing techniques are needed to solve real-world problems. The most common applications include communications and radar applications such as smart antennas and adaptive beamforming, interference cancellation, high-resolution direction of arrival estimation, channel sounding, MIMO radar and sonar and sensor networks. Other important practical applications include source localization and classification, tracking, surveillance and navigation. Array signal processing is commonly used also in biomedical applications (e.g. EEG, MEG, EKG, cancer diagnosis), machine vision, as well as geophysical and astronomical applications (radio telescopes).

The most common application of space-time processing is spatial filtering (beamforming). A beamformer is a processor used in conjunction with an array of sensors in order to provide a versatile form of spatial filtering [201]. The main goal is to estimate signals arriving from the desired directions in the presence of noise and interference signals, or to transmit signals in the desired directions. Antenna arrays with smart signal processing algorithms may also be used to identify spatial signal signatures such as the direction of arrival or location. Spatial processing techniques are also used in acoustic signal processing, track and scan radars, as well as in cellular systems. The basic idea of sensor array signal processing is given in Figure 1.1.

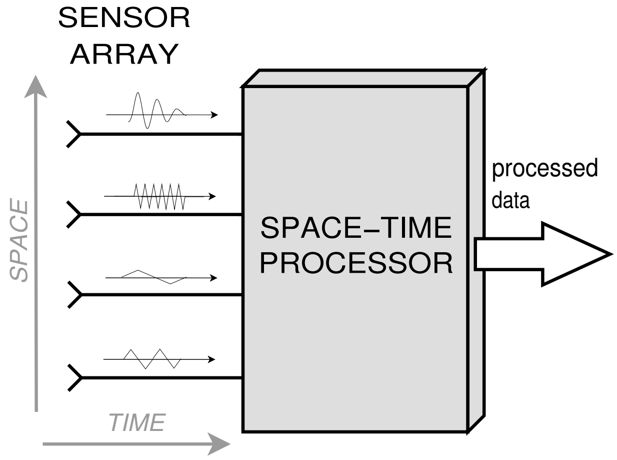

Figure 1.1: Sensor array signal processing.

The algorithms used in the space-time processor need to be smart and adaptive [105] in order to deal with time and space varying signals. In wireless receivers, optimization techniques are commonly needed. Minimizing certain error criterion, or maximizing certain gain function needs to be done iteratively, in an online fashion. Numerical optimization may be the only solution because a closed-form solution may not exist, or if it exists, it is hard to find. In multi-channel signal processing the optimization is often performed subject to matrix constraints. An elegant way to solve this type of constraint optimization problems is to view the error criterion as a multi-dimensional surface contained on the space of the free parameters. The constraint is viewed as a second surface. The feasible solutions may be found in the parameter space determined by the intersection of the two surfaces. This parameter space is usually a non-Euclidean space, i.e., a differential manifold. Consequently, powerful geometric optimization methods are needed. Classical optimization techniques operating on the usual Euclidean space suffer from low convergence speed, and/or deviation from the constraint. In order to overcome these impairments, state-of-the-art Riemannian optimization algorithms may be employed. They provide efficient solutions and allow better understanding of the problem. By using the Riemannian geometry approach, the initial constrained optimization problem is converted into an unconstrained one, in a different parameter space [18]. The constraints are fully satisfied, in a natural way. The geometric properties of the constrained space may be exploited in order to reduce to computational complexity.

Unitary/orthonormal matrices play a crucial role in array and multi-channel signal processing applications. They are involved in almost all modern transceiver techniques such as limited-feedback MIMO systems [50, 51, 121, 143–145], space-time codes [103, 116, 118, 120, 141, 144], smart antennas [85, 86, 183, 186], blind equalization and source separation [45, 156, 187, 192, 209] and MIMO radars and sonars [34, 82, 167, 191]. Unitary matrices also play an important role in biomedical applications that employ sensor arrays or pattern recognition systems (EKG, EEG, MEG, cancer prevention and diagnosis, modelling of the human body) [20, 164], machine learning [46, 73–79, 151, 160, 161], computer vision [133, 146] and optimal control [56, 211, 223]. For this reason, reliable algorithms for optimization under unitary matrix constraint are needed. In wireless communication, a major practical issue to be considered is that the terminals may possess different signal processing capabilities and limited power resources. Thus, algorithms with reasonable computational complexity should be employed in order to cope with high data rates and time-space-frequency selective channels.

In anticipation of the growing demands for voice and multimedia applications, the future beyond third generation (B3G) and fourth generation (4G) wireless services promise considerably higher effective data rates and enhanced user mobility. In this respect, emerging wireless systems like B3G Long Term Evolution (LTE) [10], IMT-2000 [8] and WiMAX [1] are being deployed. Array signal processing plays again a crucial role. By using multiple transmit/receive antennas, i.e. the so called Multiple-Input Multiple-Output (MIMO) systems [36, 194], high capacity, link reliability and enhanced network coverage may be achieved. These benefits give a strong motivation to develop reliable multi-antenna transceiver structures.

Radio spectrum is a scarce and expensive resource. It is desirable to increase the data rates without expanding the bandwidth or using additional power. Multicarrier techniques such as Orthogonal Frequency Division Multiplexing (OFDM) modulation [106, 200] use the limited spectrum very efficiently. In addition, they are very robust to channel multipath propagation and allow simple equalization. Future wireless systems such as LTE [10], IMT-2000 [8] employ OFDM in physical layer. OFDM has already been adopted in many standards such as ADSL (Asymmetric Digital Subscriber Line) [3], wireless local and metropolitan area networks (WLAN, WMAN, HIPERLAN/2) [1, 4–6, 9], and European digital audio and video broadcast (DAB, DVB) [2, 7]. The recently proposed IEEE 802.11n standard for WLAN [4] combines OFDM with MIMO in order to increase the throughput and the operation range.

In order to fully achieve the benefits of MIMO-OFDM, accurate channel estimation is needed. In MIMO systems, this is more difficult than in the single antenna case due to the the large number of channel parameters. In mobile MIMO-OFDM systems, however, plenty of pilot symbols are needed in order to deal with the space-time-frequency selective channels. Blind methods [128] achieve higher effective data rates because they do not require any training data or pilot signals. They rely only on statistical or structural properties of the transmitted signal. Semi-blind methods use a reduced amount of training data and are more feasible for fast-fading scenarios. Moreover, they allow solving all the ambiguities that blind receivers are subjected to.

1.2 Scope of the thesis

The objective of this thesis is to develop efficient optimization algorithms for multi-channel and sensor array signal processing applications such as the future multi-antenna transceivers. Special attention needs to be paid to practicality, so that no unrealistic assumptions are made in deriving the algorithms. Reasonable computational complexity of the algorithms is required in order to be able to operate at very high sampling rates needed in real-time applications.

Optimization algorithms are used in most of the modern wireless receivers. Constrained optimization problems arise frequently in sensor array and multi-channel signal processing applications. The core of and adaptive algorithm is usually an optimization algorithm, possibly with some constraints. In particular, optimization under unitary matrix constraint is needed in smart antennas, closed-loop MIMO systems, space-time codes, blind signal separation, blind subspace methods and blind beamforming. Novel optimization techniques that posses fast convergence to the desired solution need to be developed. Riemannian geometry provides powerful tools for solving this type of problems. They prove to be computationally feasible and outperform classical Euclidean approaches in terms of convergence speed. For this reason, the main objective of this dissertation is to develop optimization algorithms stemming from Riemannian geometry that have the ability to find the optimal solution in a numerically efficient manner.

The problem of blind channel equalization in MIMO-OFDM system is also addressed in this thesis. The goal is to develop computationally efficient blind algorithms for MIMO-OFDM that are able to cancel both inter-symbol and co-channel interference. They must achieve fast convergence and good tracking capabilities when used in a semi-blind mode. The convergence properties and the conditions for symbol recovery need to be established as well.

1.3 Contributions

This work contributes to the fields of multi-channel and array signal processing, optimization theory and multi-antenna communications.

There are two main contributions of this thesis, as explained below. The first main contribution is to numerical optimization techniques for array and multi-channel signal processing applications. Two novel Riemannian techniques for optimization under unitary matrix constraint are proposed.

This type of optimization arises frequently in multi-antenna transceivers algorithms. Typical applications are blind equalization, source separation and smart antennas.

Steepest descent (SD) and conjugate gradient (CG) algorithms operating on the Lie group of unitary matrices are derived. Novel line search methods specially tailored for this type of optimization are also introduced. The proposed algorithms exploit the geometrical features of the Lie group in order to reduce the computational cost. They outperform the classical Euclidean approaches for constrained optimization both in terms of convergence speed and computational complexity. Moreover, they generalize the existing Riemannian optimization algorithms for optimization under orthogonal constraint which are designed only for real-valued matrices and signals. In communications and array signal processing we deal with complex-valued signals and matrices. The complex-valued case has been addressed only in [137], since most of the authors consider the extension from the real to the complex case trivial. We show that this simplistic assumption is not always true. The complexity of the proposed algorithms is significantly lower than the differential geometry approach in [137]. They are directly applicable to joint diagonalization (JADE) [45], which is a widely used technique for blind source separation (BSS). The proposed algorithms achieve faster convergence in comparison to the approach based on Givens rotations, originally proposed in [45]. Other possible application include high-resolution direction finding and blind subspace methods. The proposed SD and CG algorithms together with the two proposed line search methods are the first ready-to-implement algorithms for optimization under unitary matrix constraint.

The second main contribution of this dissertation is in the area of multi-antenna OFDM communications systems. An optimization algorithm based on a combined criterion is proposed in order to be cancel both inter-symbol interference (ISI) and co-channel interference (CCI) in a blind manner, i.e., without known training or pilot symbols. It possesses reduced computational complexity since it does not involve any matrix inversions or decompositions. The local converge properties of the algorithm are established as well. The proposed blind algorithm can be used for channel tracking using the data symbols only. It is suitable for MIMO-OFDM systems, under slow to moderate fading conditions such as in wireless LANs and continuous transmissions (television and radio).

1.4 Structure of the thesis

The thesis consists of an introductory part and seven original publications. The publications are listed at page ix and appended at the end of the manuscript, starting at page 97. The introductory part of this thesis is organized as follows. The first two chapters deal with Riemannian optimization algorithms. In particular, the problem of optimization under unitary matrix constraint is addressed. The next two chapters address the problem of blind equalization in MIMO-OFDM systems.

In Chapter 2, an overview of optimization techniques stemming from differential geometry is provided. Constrained optimization is regarded as a geometric problem. The main focus is on optimization under unitary matrix constraint and the existing techniques. A comprehensive review of applications of differential geometry to array and multi-channel signal processing is provided.

Chapter 3 proposes algorithms optimization under unitary matrix constraint. The Lie group of unitary matrices $U(n)$ is described as a real manifold. Motivation why the existing algorithms designed for real-valued matrices cannot be straight-forwardly applied to complex-valued matrices is given. Two computationally feasible optimization algorithms operating on the Lie group $U(n)$ are introduced. Steepest Descent (SD) and a Conjugate Gradient (CG) algorithms exploiting the geometrical properties of $U(n)$ are provided. Two efficient line search methods specially tailored for the proposed algorithms are also introduced.

In Chapter 4, blind channel identification and equalization algorithms for multi-antenna OFDM systems are reviewed. A classification of these algorithms is also provided.

Chapter 5 addresses the problem of blind equalization in spatial multiplexing MIMO-OFDM systems. An algorithm which optimizes a composite criterion in order to mitigate both inter-channel and co-channel interference is proposed.

Chapter 6 provides a summary of the contributions and the results of the thesis. Future research directions are also discussed.

1.5 Summary of publications

In this subsection, brief overview over the author's original publications is given.

In [Publication I], a Riemannian Steepest Decent (SD) algorithm for optimization under unitary matrix constraint is derived. The algorithm benefits from the geometrical features of the Lie group of unitary matrices $U(n)$ in order to reduce the computational complexity. Recent advances in numerical techniques for computing the matrix exponential needed in the update are exploited. Armijo line search method is efficiently used in the step size selection. The computational complexity and stability issues are addressed in detail. Detailed implementation tables and numerical solutions are provided, unlike other more general algorithms that require solving matrix equations (sometimes differential equations) and do not provide any feasible numerical solutions. The proposed algorithm is tested in a blind source separation application for MIMO systems, using joint diagonalization.

In [Publication II], a Riemannian Conjugate Gradient (CG) algorithm for optimization under unitary matrix constraint is derived. Two efficient line search methods exploiting the almost periodic property of the cost function along geodesics on $U(n)$ are also proposed. Detailed description of the implementation is provided. The proposed CG algorithm and line search methods are tested in a blind source separation application for MIMO systems using joint diagonalization. They are also used to compute the eigendecomposition of a Hermitian matrix iteratively by maximizing the Brockett criterion [42, 183, 184].

In [Publication III], a Riemannian steepest descent based on Taylor approximation of the geodesics is proposed. The Riemannian SD algorithm is applied to high resolution direction finding. A comparison to Euclidean approaches is provided. The algorithm is also applied to an existing blind receiver for MIMO-OFDM systems [228] in order to reduce complexity.

In [Publication IV], introduces two novel line search methods on $U(n)$. The first method uses a polynomial approximation of the derivative of the cost function. The second method is based on a DFT approximation. They are used together with the Riemannian SD algorithm in [Publication I].

[Publication V] compares Riemannian SD and CG algorithms on $U(n)$ using Armijo line search method. The algorithms are applied to two different cost functions. The first one is the Brockett criterion and is used to perform the diagonalization of a Hermitian matrix. The second one is the JADE criterion and is used to perform the blind signal separation of communications signals in a MIMO system.

In [Publication VI], a blind equalization algorithm for spatial multiplexing MIMO-OFDM systems is proposed. The algorithm is based on a composite criterion designed to cancel both the inter-symbol and co-channel interference. The composite criterion is comprised of a Vector Constant Modulus (VCM) criterion and a decorrelation criterion. Identifiability conditions for the MIMO channel are also provided.

In [Publication VII] a blind equalization algorithm for spatial multiplexing MIMO-OFDM systems is proposed. The algorithm is the final version of the algorithm in [Publication VI], subsequently developed in [13] and [12]. The VCMA criterion is modified in order to be able to cope with the correlation introduced by the cyclic prefix (CP). The resulting composite criterion consists of a modified VCM criterion and a decorrelation criterion. Local convergence properties of the algorithm have been established. Conditions for the blind equalization and co-channel signal cancellation are also provided.

All the simulation software for all the original publications included in this dissertation was written solely by the author.

In [Publications I–VIII], the original ideas and the derivations of the algorithms were developed by the first author. Simulations were made by the first author as well. The co-authors provided guidance during the development of the algorithms, establishing their properties and the design of the experiments. They have also provided valuable comments that substantially improved the rigorousness and the technical quality of the papers.

Chapter 2

Overview of geometric optimization techniques

In this chapter, different approaches for solving optimization problems [27, 83, 153, 163] subject to differentiable equality constrains are reviewed. We are mainly interested in minimizing cost functions, but all the algorithms considered can be easily adapted to maximization problems. Classical constrained optimization algorithms operating on Euclidean spaces, as well as non-Euclidean approaches are presented in Section 2.1. Different algorithms for optimization under unitary matrix constraint are revisited in Section 2.2. A detailed literature review of Riemannian optimization algorithms and their applications in signal processing is provided in Section 2.3.

2.1 Constrained optimization from a differential geometry perspective

Constrained optimization problems arise frequently in many sensor array and multi-channel signal processing applications. Most of the adaptive algorithms [105] require minimizing an error criterion in an iterative manner, subject to some matrix equality constraint. Solving this type of problems is typically done numerically. In fact, numerical optimization may be the only way to solve certain problems. One of the reasons for the numerical optimization may be that a closed-form solution does not exist, or if it exists it is very hard to find. Iterative optimization algorithms are also suitable in the case where only small corrections to a previous solution need to be applied sequentially.

Classical approaches solve this problem on the Euclidean space and they are mainly of two types. The first approach supposes optimizing the unconstrained cost function by using classical gradient algorithms. In order to satisfy the constraint, a restoration procedure needs to be applied after every iteration. In general, the resulting algorithms convergence slowly due to fact that most of the effort is spent on enforcing the constraint and not on the actual optimization.

The second classical approach for constraint optimization is based on the method of Lagrange multipliers [97]. The method introduces an additional set of unknown scalar parameters called Lagrange multipliers. The number of multipliers is equal to the number of scalar constraints. A new cost function called Lagrangian is constructed. It is comprised of the original cost function and the set of constraints weighted by the Lagrange multipliers. This function is jointly optimized w.r.t. to both the original variables and the new variables, i.e., the Lagrange multipliers. Closed-from solutions may be found in simple cases by solving a system of equations. Often, increasing the number of unknowns is undesirable, especially when the dimension of the optimization problem is already large. Many practical applications involve complicated expressions of the cost function (e.g. the JADE criterion [45]) and non-linear matrix constraints (e.g. the unitary matrix constraint). In such cases, the Lagrangian approach requires solving a large system of non-linear equations which may be mathematically intractable. For this reason, simplified extra-penalty methods [153] stemming from the Lagrangian approach have been proposed. Instead of using several Lagrange multipliers, they use a single scalar parameter to weight the extra-term which penalizes the deviation from the constraint. The corresponding composite cost function is minimized iteratively by using a classical steepest descent method. This approach may lead to inaccurate solutions due to the fact that the deviation from constraint accumulates after each iteration. Additional stabilization procedures are usually needed.

More reliable and modern solutions to some specific classes of constrained optimization problems [18,87,131,184,196] may be obtained by using tools of differential geometry [59,101]. The initial constrained optimization problem on Euclidean space is converted into an unconstrained one, on a different parameter space. This new parameter space can be viewed as a system of coordinates where only the values of coordinates satisfying the constraint are allowed. Geometrically, the set determined by the constraints can be understood as a lower-dimensional space embedded on the initial Euclidean space (see Figure 2.1). Usually smooth (differentiable) constraints may determine *differentiable manifolds* [59]. In general, these manifolds are non-Euclidean spaces and may be elegantly explored by using state-of-the-art tools from differential geometry. Efficient numerical optimization algorithms may be derived by taking into account the *geometrical structure* of the manifold. In addition to that, matrix manifolds [18] possess rich *algebraic structure* arising from the special properties of matrices.

The classical gradient-based algorithms such as steepest descent, conjugate gradient and Newton methods can be naturally extended from Euclidean space to differentiable manifolds. Pioneering work by Luenberger

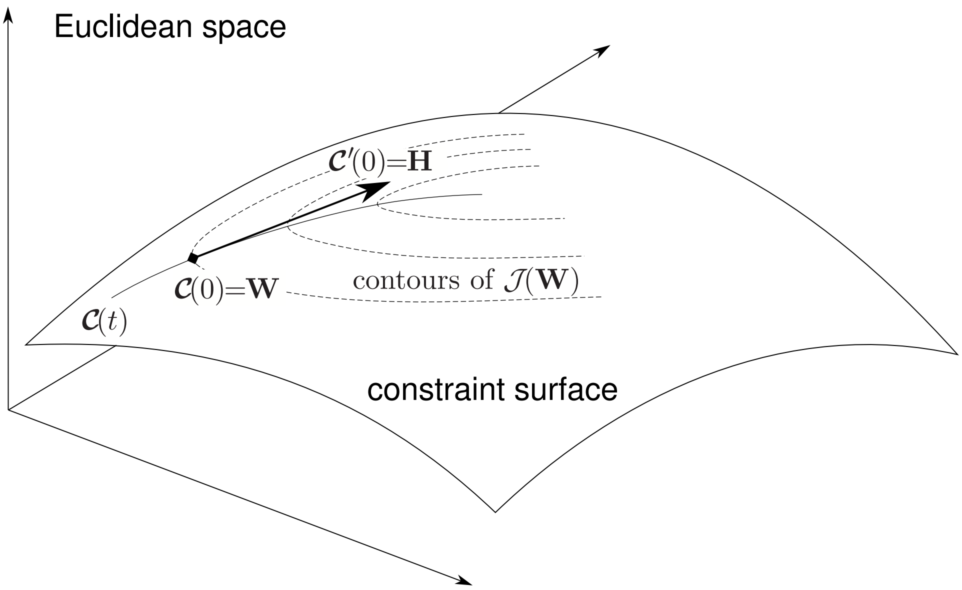

Figure 2.1: Optimization on manifolds.

[131] and Gabay [87] treats the constrained optimization problem in a differential geometry context and establishes interesting connections with the Lagrange multipliers method considering the first and the second-order optimality conditions. Optimizing a differentiable cost function $\mathcal{J}(\mathbf{W})$ on a differentiable manifold supposes choosing a point $\mathbf{W}$ on the manifold and a search direction $\mathbf{H}$ tangent to the manifold, at $\mathbf{W}$ (see Figure 2.1). The next iteration consists of moving along a curve $\mathcal{C}(t)$ which emanates from $\mathbf{W}$ in the direction of $\mathbf{H}$. This curve is called *local parametrization* and can be used to describe a neighborhood around a given point on the manifold. Therefore, $\mathcal{C}(t)$ must be contained on the manifold and fulfill the condition $\mathcal{C}(0) = \mathbf{W}$. Additionally, its derivative at $\mathbf{W}$ must coincide with the search direction $\mathbf{H}$, i.e., $\mathcal{C}'(0) = \mathbf{H}$. Optimizing the cost function in one dimension along the curve $\mathcal{C}(t)$ is required at every iteration. It means that a line search is performed. By moving along such curves, the constraint is automatically satisfied at each iteration.

Choosing the appropriate local parametrization and the appropriate search direction must be computationally feasible. The most natural local parametrizations are the *geodesics*, which on Riemannian manifolds are locally the length minimizing curves. They correspond to the straight lines on the Euclidean space. On certain manifolds such as Lie groups [89, 107, 123, 208], the geodesics are computationally attractive. Using other local parametrizations is also possible. A well-known *non-geodesic* approach is to use a *retraction*, which is a map that locally projects the tangent plane onto the manifold [16, 18, 131, 137]. Choosing the search direction is a compromise between high complexity and fast convergence. This direction can be, for example, the steepest descent (or ascent) direction, or other direction, for example the one corresponding to a conjugate gradient or Newton algorithm.

The steepest descent (SD) or steepest ascent (SA) algorithm on differentiable manifolds [41, 57, 69, 77, 87, 131, 150, 151, 161, 183, 196, 219, 220] is relatively simple, but its convergence is only linear [87, 183, 196, 219, 220]. Its asymptotic rate of convergence is related to the eigenvalues of the Hessian associated with the Lagrangian function of the constrained optimization problem evaluated at the solution [131]. Developing a SD algorithm on a differentiable manifold requires defining intrinsically the gradient field of the cost function on the manifold. The gradient can be defined only after endowing the differentiable manifold with a Riemannian metric, which turns it into a *Riemannian manifold*. The Riemannian gradient is a vector tangent to the manifold, along which the cost function increases the fastest.

The conjugate gradient (CG) algorithm on differentiable manifolds [57, 69, 122, 183, 184, 229] is still a relatively simple algorithm and it achieves superlinear convergence [183]. In general CG is considerably simpler than a Newton algorithm which would require computing second-order derivatives. CG captures the second-order information from two successive first-order derivatives which are properly combined. The additional complexity compared to the SD is due to the fact that CG requires transporting gradient vectors from one point to another, i.e., performing *parallel transport*. This operation is not as simple as in a vector space. The parallelism is not relative to straight lines as on the Euclidean space, but to an *affine connection* (often the Levi-Civita connection) [59, 107, 154]. The connection makes clear the two basic ideas of covariant differentiation and parallel transport [59, 107, 154]. In certain manifolds the parallel transport can be done in a very simple manner, and the resulting CG algorithm is comparable to the SD in terms of computational complexity.

Riemannian Newton algorithm [17, 87, 135, 155, 183, 196] achieves quadratic convergence [17, 87, 183]. Newton algorithm is a prohibitively expensive algorithm even on Euclidean space, when the dimension of the optimization problem is large. On Riemannian manifolds, the complexity is even higher, since it would require computing the second covariant derivative and inverting it. For this reason, CG [69, 122, 183, 184, 229], or modified versions of the Newton algorithm [137] are often preferred. Moreover, the Newton algorithm may converge to any type of stationary points, not only the extrema of interest. Trust-region methods on Riemannian manifolds have been recently proposed in the literature [15, 18].

In conclusion, in order to extend the optimization algorithms from the Euclidean space to Riemannian manifolds, the straight lines are replaced by geodesics, the classical differentiation by covariant differentiation and the idea of vector addition by exponential map and parallel transport [184]. The most common differentiable manifolds which arise in signal processing applications are *homogeneous spaces* [107, 208] such as *the Stiefel manifold*, *the Grassmann manifold* [69,74,75,137,183]. The Stiefel manifold $\mathrm{St}(n,p)$ is the set of $n \times p$ (real or complex) matrices with mutually orthogonal columns. The Grassmann manifold $\mathrm{Gr}(n,p)$ is the set of all p-dimensional subspaces of the n-dimensional Euclidean space ($\mathbb{R}^n$ or $\mathbb{C}^n$). They are also represented by $n \times p$ orthonormal matrices which can be in this case any arbitrarily rotated basis of a subspace. The Stiefel manifold of $n \times n$ orthogonal/unitary matrices is a special case due to the fact that these matrices are algebraically closed under the standard matrix multiplication operation, i.e., they form a matrix *Lie group* [89, 107, 123, 208]. A matrix Lie group is a differentiable manifold and a matrix group at the same time. Orthogonal matrices form the *orthogonal group* $O(n)$. Similarly, unitary matrices form the *unitary group* $U(n)$. Another relevant Lie group is the *general linear group* $GL(n)$, which is the group of $n \times n$ invertible matrices. In general, the additional group structure brings computational benefits which may be exploited in practical algorithms, as it will be shown later in Chapter 3. For this reason, in the literature special attention has been paid to optimization algorithms operating on Lie groups [17,40,41,57,69,76,77,79,135,150,151,155,161,183, 214, 229, 230].

In particular, in this thesis, the problem of optimization under unitary matrix constraint is addressed. Computationally efficient optimization algorithms which exploit the geometry of the Lie group $n \times n$ unitary matrices $U(n)$ are proposed.

2.2 Optimization under unitary matrix constraint - different approaches

Consider a differentiable real-valued cost function $\mathcal{J} : \mathbb{C}^{n \times n} \to \mathbb{R}$. The problem of optimization under unitary matrix constraint may be formulated as:

$$\text{minimize } \mathcal{J}(\mathbf{W}) \text{ subject to} \tag{2.1}$$
$$\mathbf{W}^H\mathbf{W} = \mathbf{I}_n. \tag{2.2}$$

Two main approaches for solving the above optimization problem have been proposed in the literature. The first one requires solving the *constrained* optimization problem on the Euclidean space by using classical gradient-based algorithms. The second one requires solving an *unconstrained* optimization problem on the differentiable manifold determined by the constrained set, which is the Lie group of $n \times n$ unitary matrices by $U(n)$. The two main approaches are presented below together with the methods which they include. Similar classification is provided in [Publication I],[Publication III], and [162].

2.2.1 Classical Euclidean approach for optimization under unitary matrix constraint

The classical Euclidean approach for solving constraint optimization problems include the unconstrained gradient-based method with constraint enforcement and the method of Lagrange multipliers (or methods stemming from it).

Euclidean gradient algorithms with constraint enforcement

An unconstrained classical gradient-based method is used to minimize the cost function $\mathcal{J}(\mathbf{W})$. The constraint is satisfied by using an additional restoration procedure, which needs to be applied after every iteration. This is a well-known technique for optimization under unitary/orthonormal matrix constraint [64, 113, 156, 157, 171, 172, 216]. Each iteration k of the algorithm consists of the two following steps:

$$\begin{aligned} \tilde{\mathbf{W}}_{k+1} &= \mathbf{W}_k - \mu_k \mathbf{H}_k^{\mathrm{E}} && (2.3)\\ \mathbf{W}_{k+1} &= \boldsymbol{\mathcal{P}}\{\tilde{\mathbf{W}}_{k+1}\}, && (2.4)\end{aligned}$$

where $\mathbf{H}_k^{\mathrm{E}}$ is an ascent direction and the step size $\mu_k > 0$ sets the convergence speed. The search direction $\mathbf{H}_k^{\mathrm{E}}$ may be the gradient direction on the Euclidean space at $\mathbf{W}_k$, i.e., $\nabla^{\mathrm{E}}\mathcal{J}(\mathbf{W}_k)$. Other search directions could be used, for example the one corresponding to a conjugate gradient, or Newton method. Due to the additive type of update, the result of (2.3) is generally not a unitary matrix. The new iterate $\tilde{\mathbf{W}}_{k+1}$ deviates from the unitary property at every iteration. The constraint needs to be restored by projecting $\tilde{\mathbf{W}}_{k+1}$ onto the space of unitary matrices. A projection operator $\boldsymbol{\mathcal{P}} : \mathbb{C}^{n\times n} \to U(n)$ is used in (2.4) in order to obtain a unitary matrix. Few algorithms in the literature [11, 113, 216] find the unitary matrix which is the closest to the original one under the Euclidean norm. The corresponding projection is known as "the symmetric orthogonalization" procedure [113, Ch. 6, Sect. 6.6], and it may be written as

$$\boldsymbol{\mathcal{P}}\{\tilde{\mathbf{W}}_{k+1}\} = \tilde{\mathbf{W}}_{k+1}(\tilde{\mathbf{W}}_{k+1}^H \tilde{\mathbf{W}}_{k+1})^{-1/2}, \quad (2.5)$$

or in terms of left and right singular vectors of $\tilde{\mathbf{W}}_{k+1}$ [108]. Other non-optimal projections (under Euclidean norm) may be used, for example the one based on Gram-Schmidt orthogonalization [64, 156, 157, 171, 172, 216].

The approach does not take into account the group property of unitary matrices, i.e., the fact that unitary matrices are closed under the multiplication operation. This is not the case under addition. The departure from the unitary property may be significant and most of the effort will be spent on enforcing the constraint, instead of moving towards the optimum. Consequently, this algorithm achieves lower convergence speed, as demonstrated in [Publication I] and [Publication III]. The Euclidean gradient algorithms combined with projection methods do not take into account the curvature of the constrained surface, and therefore achieve only linear convergence [183]. Moreover, this approach makes the idea of line search optimization meaningless.

Lagrange multipliers and related methods

The method of Lagrange multipliers [97] is the basic tool for optimizing a function of multiple variables subject to one or more scalar constraints. The method introduces a set of new real scalar unknowns λ_i called *Lagrange multipliers*. A composite cost function $\mathcal{L}(\mathbf{W})$ called *Lagrangian* is constructed by using the original cost function $\mathcal{J}(\mathbf{W})$ and an extra-term containing the constraints weighted by the Lagrange multipliers. This function is jointly optimized w.r.t. both the elements of $\mathbf{W}$ and the Lagrange multipliers λ_i. In this way, finding stationary points of the constrained cost function of $\mathcal{J}(\mathbf{W})$ is equivalent to finding the stationary points of the unconstrained cost function $\mathcal{L}(\mathbf{W})$. The unitary matrix constraint (2.2) is equivalent to n^2 real scalar constraints. Consequently, a number of n^2 real Lagrange multipliers are required to construct the Lagrangian. The original cost function $\mathcal{J}(\mathbf{W})$ has already $2n^2$ free real variables. Therefore, the Lagrangian has $3n^2$ free variables. Increasing the number of variables is undesired, especially when n is large. A large system of non-linear matrix equations needs to be solved in order to find the stationary points. Often, for practical cost functions, solving such a system of equations is non-trivial already for $n \geq 2$. An example of such cost function is the JADE criterion [45], whose optimization has been considered in [Publication I] and [Publication II].

For this reason, a simplified technique stemming from the method of Lagrange multipliers has been proposed in the literature [205]. This technique uses a gradient-based iterative method to minimize a composite cost function on the Euclidean space. The composite cost function is comprised of the original cost function $\mathcal{J}(\mathbf{W})$ and an extra-term $\mathcal{U}(\mathbf{W}) = \|\mathbf{W}^H\mathbf{W} - \mathbf{I}_n\|_F$, which penalizes the deviation from the unitary constraint. In this work we will refer to this method as "the extra-penalty method". The kth iteration of the extra-penalty method is of the form

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu_k[\mathbf{H}_k^{\mathrm{E}} + \lambda_k \mathbf{U}_k^{\mathrm{E}}] \tag{2.6}$$

The direction $\mathbf{H}_k^{\mathrm{E}}$ represents the Euclidean gradient of the original cost function $\mathcal{J}(\mathbf{W})$ at $\mathbf{W}_k$, whereas the direction $\mathbf{U}_k^{\mathrm{E}}$ is the Euclidean gradient of the penalty function $\mathcal{U}(\mathbf{W})$ at $\mathbf{W}_k$. The later term is used to penalize the deviation from the unitary property, and in a way it resembles the Tikhonov regularization method [147]. A single scalar weighting parameter λ_k is used to weight the direction $\mathbf{U}_k^{\mathrm{E}}$ in a manner similar to the method of Lagrange multipliers. For this reason, the method has also been called bigradient method in the literature [205]. Most of the extra-penalty type of algorithms in the literature dealing with optimization under orthonormal constraints are specialized to certain tasks such as subspace tracking or Independent Component Analysis (ICA) [61, 64, 66, 205]. They are computationally efficient for the task they are designed for. In general, they cannot be applied to other optimization problems with unitary/orthonormal constraint. Some of the optimization algorithms are restricted only to the case of optimizing $\mathcal{J}(\mathbf{w})$, where $\mathbf{w}$ is a unit-norm vector. In the unit-norm vector case, a scalar parameter λ_k always exists such that the unit-norm constraint is satisfied. Finding the optimum λ_k which keeps the constraint satisfied limits the choice of the step size $\mu_k > 0$, and consequently the convergence speed. This limitation is due to the additive update which alters the constraint. Algorithms based on additive update which are able to keep the constraint satisfied up to machine precision exist in the literature [60]. They are based on Householder transforms [99]. These algorithms are specialized to minor component analysis (MCA). The extra-penalty method is in general numerically unstable, the deviation from the unitary constraint accumulates over time. Self-stabilized algorithms have been proposed [61–66], in order to avoid this unstable behavior. They usually discretize the differential equation which describes the motion on the Riemannian manifold determined by the constraint. The discretization leads to numerical errors which are compensated by inserting additional stabilizing factors at various points within update (see [Publication III]). The numerical stability is improved and this fact has been proved in [61] by using an asymptotic analysis which shows that the error does not accumulate over time.

In conclusion, solutions based on the method of Lagrange multipliers may be non-trivial to compute for large dimensions n. Moreover, the mathematical tractability is highly dependent on the expression of the cost function, since a system of $3n^2$ nonlinear equations with $3n^2$ unknowns needs to be solved. The solution obtained by using the extra-penalty method satisfies the constraint just approximately, when only one scalar parameter λ_k is used. This fact has been demonstrated in [Publication I] and [Publication III]. Finally, these methods are not directly applicable to a general problems of optimization under unitary matrix constraint.

2.2.2 Differential geometry based optimization algorithms

The unitary matrix constraint (2.2) is a smooth constraint which determines a differentiable manifold. This manifold can be seen as a "constrained surface" embedded on an higher-dimensional Euclidean space. Differential geometry-based optimization algorithms move along directions which are tangent to the manifold. Depending on the choice of the local parametrization there are two types of algorithms *projection-based algorithms* and *geodesic algorithms*.

An important aspect to be considered is that the unitary matrices are closed under the standard matrix multiplication operation, i.e., they form the Lie group of $n \times n$ unitary matrices $U(n)$. This fact brings additional properties which may be exploited in optimization. The most important property is that geodesics are described by simple formulas, therefore they can be efficiently computed.

Unlike the Lagrange multipliers method which introduces new unknown variables, differential geometry-based algorithms exploit the reduced dimension of the manifold. The original problem involves a cost function of $2n^2$ variables and n^2 constraints. The differential geometry-based approach involves only n^2 variables (which is the dimension of $U(n)$), and no constraints. Moreover, the mathematical tractability does not depend on the expression of cost function.

Non-geodesic differential geometry based gradient algorithms

This type of algorithms use as a local parametrization a projection operator (2.5). They move along straight lines tangent to the manifold and deviate from the unitary constraint at every iteration. This is due to the fact that the the manifold is a “curved space”. Therefore, this algorithm has the same drawback as its Euclidean counterpart, i.e., the constraint restoration procedure needs to be applied after every iteration. The algorithm consists of the two following steps at each iteration:

$$\tilde{\mathbf{W}}_{k+1} = \mathbf{W}_k - \mu_k \mathbf{H}_k^{\mathrm{R}} \tag{2.7}$$

$$\mathbf{W}_{k+1} = \boldsymbol{\mathcal{P}}\{\tilde{\mathbf{W}}_{k+1}\}. \tag{2.8}$$

The search direction $-\mathbf{H}_k^{\mathrm{R}}$ is a descent direction tangent to the manifold at $\mathbf{W}_k$. The step size $\mu_k > 0$ determines the convergence speed. Compared to its Euclidean counterpart, they depart less from the constrained surface since they move along the search directions which are tangent to the manifold. The deviation is only due to the curvature of the manifold, and not because of the inaccurate search direction. Consequently, these type of algorithms achieve better convergence speed, as demonstrated in [Publication I] and [Publication III].

From a differential geometry point of view, the projection operator is as local parametrization on the manifold, i.e., a mathematical description of the neighborhood of an arbitrary point $\mathbf{W}_k \in U(n)$. The corresponding curve starting from $\mathbf{W}_k \in U(n)$ with the initial tangent vector $-\mathbf{H}_k^{\mathrm{R}}$ is given by $\hat{\boldsymbol{\mathcal{P}}}(\mu) = \boldsymbol{\mathcal{P}}\{\mathbf{W}_k - \mu \mathbf{H}_k^{\mathrm{R}}\}$. Unfortunately, due to the nature of the projection operator, the line search optimization needed in the step size selection is computationally expensive.

Riemannian gradient algorithms along geodesics

A natural way to optimize a cost function on a Riemannian manifold is to move along geodesics. Geodesics are locally the length minimizing paths on a Riemannian manifold [59]. Intuitively, they correspond to straight lines on Euclidean space. Riemannian algorithms for optimization under unitary matrix constraint use the exponential map as a local parametrization. The resulting algorithms employ a multiplicative update rule, i.e., a rotation is applied to the previous value to obtain the new one. Each iteration k of the algorithm consists of the following step:

$$\mathbf{W}_{k+1} = \exp(-\mu_k \mathbf{H}_k^{\mathrm{R}})\mathbf{W}_k = \mathbf{R}_k \mathbf{W}_k \tag{2.9}$$

where $-\mathbf{H}_k^{\mathrm{R}}$ is the directional vector of the geodesic and is represented by a skew-Hermitian matrix. Consequently, its matrix exponential $\mathbf{R}_k = \exp(-\mu_k \mathbf{H}_k^{\mathrm{R}})$ is a unitary matrix. Since $\mathbf{W}_k$ is a unitary matrix, $\mathbf{W}_{k+1}$ will remain unitary at every iteration. In this way, the constraint is maintained automatically and no enforcing procedure is necessary.

Optimization algorithms with unitary constraints such as the ones in [137] are more general in the sense that they are designed for the Stiefel and the Grassmann manifolds. Therefore, when dealing with the case on $n \times n$ unitary matrices they do not take into account its Lie group structure which brings numerous computational benefits. We fully exploit these benefits in the gradient algorithms proposed in [Publication I], [Publication II], [Publication III], [Publication IV] and [Publication V]. The most important advantage is the convenient expression for the geodesics and parallel transport. Geodesics are expressed in terms of matrix exponential of skew-Hermitian matrices. Hence, they are easier to compute compared to the projection-based method [137] which requires the computing SVD of arbitrary matrices. Recent progress in numerical methods for calculating the matrix exponential may be exploited [47, 117, 142]. For more details see [Publication I], Section V. Another important property of the Lie group is that transporting vectors from a tangent space to another may be done in a very simple manner. The parallel transport which is needed for the conjugate gradient algorithm may be done simply by using left/right matrix multiplications [Publication I]. Moreover, the geodesic search when adapting the size μ_k at every iteration may be efficiently done by using for example the Armijo rule [163]. Efficient line search methods for step size adaptation have been proposed in [Publication IV]. They exploit the *almost periodic* property of the cost function along geodesics on $U(n)$.

In conclusion, geodesic algorithms for optimization under unitary matrix constraint fully exploit the geometric and algebraic structure of the Lie group of unitary unitary matrices to reduce complexity.

2.2.3 Optimization under unitary matrix constraint – an illustrative example

The goal of this example is to illustrate how different optimization algorithms operate under unitary matrix constraint. We use a simple toy problem for illustration purposes. We minimize the cost function $\mathcal{J}(w) = |w + 0.2|^2$ under unitary constraint, which in this case is the unit circle $w^*w = 1,\ w \in \mathbb{C}$. The unit-length complex numbers form the Lie group of 1×1 unitary matrices $U(1)$.

We consider a 3-dimensional representation of the cost function $\mathcal{J}(w)$ with respect to the real and the imaginary parts of w, i.e., $x = \Re\{w\}$ and $y = \Im\{w\}$. The cost function $\mathcal{J} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is quadratic in x and y and is represented by the paraboloid $\mathcal{P}$ in Figure 2.2. The unitary constraint is represented by the cylinder obtained by translating the unit circle $U(1)$ along the vertical axis. The resulting cylinder is the space $U(1)$ represented in three dimensions (embedded on $\mathbb{R}^3$). The parameter space where the cost function satisfies the constraint $x^2+y^2 = 1$ is represented by the intersection of the cost function surface with the cylinder, i.e., the ellipse $\mathcal{E}$ represented by thick curve. This ellipse represents the constrained parameter space of the cost function $\mathcal{J} : U(1) \rightarrow \mathbb{R}$.

There are a significant differences between minimizing the cost function $\mathcal{J}(w)$ on $\mathbb{R}^2$ or on $U(1)$. On the Euclidean space, i.e., $\mathbb{R}^2$ the minimum is attained at the point m^{E} of coordinates $(x, y) = (-0.2\,, 0)$ (the minimum of the paraboloid $\mathcal{P}$ in Figure 2.2). This point does not satisfy the constraint, therefore is an undesired minimum. On the Riemannian space the minimum is attained at the point m^{R} of coordinates $(x, y) = (-1,\ 0)$ (the minimum on the ellipse $\mathcal{E}$ in Figure 2.2). This point satisfies the constraint and is the desired minimum. The steepest descent direction on the Euclidean space $-\nabla^{\mathrm{E}}\mathcal{J}(x_0, y_0)$ at a given point p satisfying the constraint is tangent to the meridian of the paraboloid, and points in the direction of the undesired minimum. The Riemannian steepest descent direction $-\nabla^{\mathrm{R}}\mathcal{J}(x_0, y_0)$ is tangent to the ellipse $\mathcal{E}$, and points in the direction of the desired minimum.

The two-dimensional representation of the cost function in Figure 2.3 shows how different algorithms operate under unitary constraint. Five different algorithms are considered. The first two algorithms are the unconstrained and the constrained SD on $\mathbb{R}^2$, respectively. The constrained version enforces the unit norm after every iteration, as described in Subsection 2.2.1. The third algorithm is the extra-penalty method, as described in Subsection 2.2.1. The fourth and the fifth algorithms operate on $U(1)$, and they are the non-geodesic SD described in Subsection 2.2.2 and the geodesic SD algorithm in Subsection 2.2.2, respectively. We may notice in Fig. 2.3 that the unconstrained SD (marked by $\diamondsuit$) takes the steepest descent direction in $\mathbb{R}^2$, and goes straight to the undesired minimum. By enforcing the unit norm constraint, we project radially the current point on the unit circle ($\Box$). In

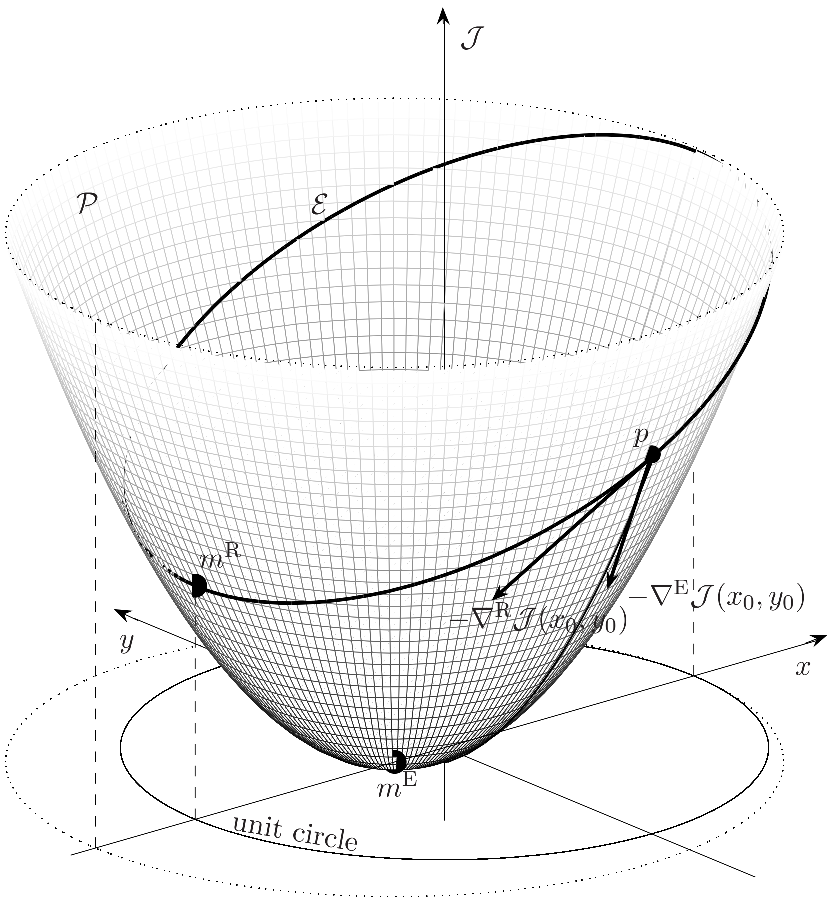

Figure 2.2: Three-dimensional visualization of the cost function given by $\mathcal{J}(x, y) = (x + 0.2)^2 + y^2$, represented by the paraboloid $\mathcal{P}$. The unitary constraint $x^2 + y^2 = 1$ in $\mathbb{R}^3$ is the cylinder obtained by translating the unit circle along the vertical axis (not shown in the figure). The constrained parameter space of the cost function $\mathcal{J}(x, y)$ is obtained by intersecting the paraboloid with the cylinder, i.e., the ellipse $\mathcal{E}$ represented by thick curve. The minimum of the *unconstrained* cost function m^{E} is the minimum in the Euclidean space. The minimum of the *constrained* cost function m^{R} is the minimum in the Riemannian space $U(1)$. The Euclidean steepest descent direction $-\nabla^{\mathrm{E}}\mathcal{J}(x_0, y_0)$ at the point (x_0, y_0) is tangent to the meridian of the paraboloid crossing (x_0, y_0). The Riemannian steepest descent direction $-\nabla^{\mathrm{R}}\mathcal{J}(x_0, y_0)$ at the point (x_0, y_0) is tangent to ellipse.

each step the constraint has to be enforced in order to avoid undesired minimum. The extra-penalty SD algorithm ($\triangledown$) converges somewhere between

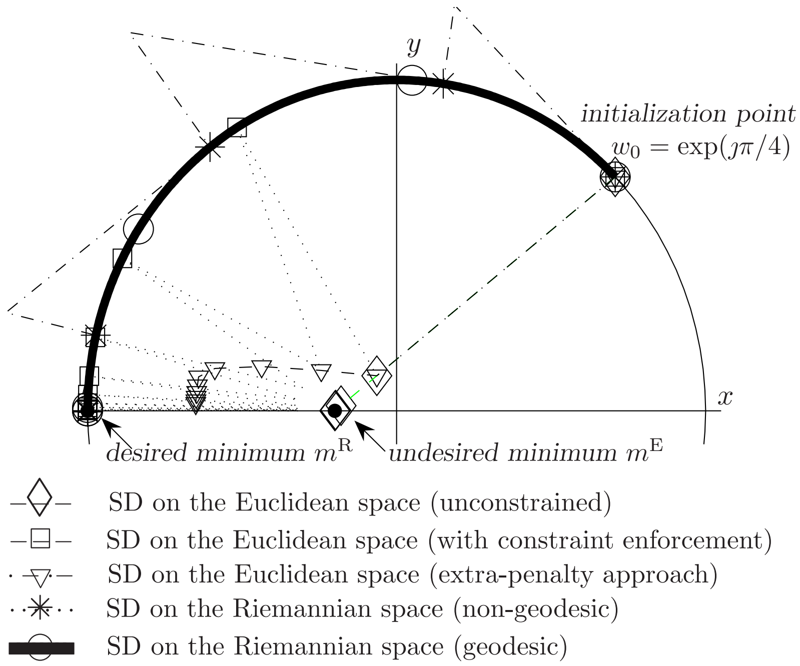

Figure 2.3: Minimizing the cost function $\mathcal{J}(x,y) = (x+0.2)^2 + y^2$ on the unit circle $U(1)$: Euclidean vs. Riemannian steepest descent (SD) methods.

the desired and the undesired minimum, depending on the factor which is used to weight the extra-penalty. The non-geodesic SD algorithm on $U(1)$ ($\ast$) [137] takes the steepest descent direction on $U(1)$, and moves along a straight line tangent to the unit circle. Due to the non-zero curvature of $U(1)$ the constraint needs to be enforced after every iteration by projection. The geodesic SD algorithm ($\bigcirc$) uses a multiplicative update which is a phase rotation in this case. Consequently, the constraint is satisfied at every step in a natural way. Although this low-dimensional example is rather trivial, it has been included for illustrative purposes. In the case of multi-dimensional unitary matrices, a similar behavior of the algorithms considered here is encountered.

2.3 Applications of differential geometry to array and multi-channel signal processing

Differential geometry has become a highly important topic in the signal processing community. It does not only provide powerful tools for solving certain problems, but it allows a better understanding of the problems [89].

A comprehensive overview of the geometric methods and their applications in engineering is provided in [88]. A recent review of applications of differential geometry to signal processing may be found in [138]. The applications may be classified by the nature of the task to be solved as follows: *optimization on manifolds*, *tracking on manifolds*, *statistics on manifolds* and *quantization on manifolds*. In this section, a detailed presentation of this classification, as well as application-oriented literature review are provided.

2.3.1 Optimization and tracking on manifolds

When the optimization needs to be performed subject to differentiable equality constraints, the optimization on differentiable manifolds arises naturally [18]. In this way, the initial constrained optimization problem becomes an unconstrained one, on an appropriate differentiable manifold. Among the most relevant contributions to this area throughout the past thirty-five years are brought by Luenberger in 1972 [131], Gabay in 1982 [87], Smith in 1993 [183] and Udrişte in 1994 [196]. They consider the optimization on general differentiable manifolds. However, the corresponding algorithms do not exploit special properties that may appear on certain manifolds, such as Lie groups and homogeneous spaces [107, 123, 208]. For this reason, the resulting algorithms may exhibit high computational complexity. Most of the work done after 1994 is dedicated to optimization on particular Riemannian manifolds. The most popular manifolds arising in practice are determined by orthonormal matrix constraints, such as the *Stiefel manifold* $\mathrm{St}(n, p)$ and *the Grassmann manifold* $\mathrm{Gr}(n, p)$ [17, 69, 74, 75, 137, 151]. The points on these manifolds may be represented by orthonormal $n \times p$ matrices. The $n \times n$ orthogonal/unitary matrices are a special case of Stiefel manifold, i.e., they form Lie groups. Special attention has been paid in the literature to optimization algorithms operating on the orthogonal group $O(n)$ [17, 40, 41, 69, 76, 77, 79, 150, 151, 161, 183, 214] and on the general linear group $GL(n)$ [229, 230]. The above algorithms are designed only for real-valued matrices. Complex-valued case has been addressed only in [137], since most of the authors consider the extension from the real to the complex case trivial. Most of the communication and sensor array signal processing applications deal with complex-valued matrices and signals.

In this thesis we focus on Riemannian algorithms operating on the unitary group $U(n)$. Therefore, these algorithms are designed for complex-valued matrices. They fully exploit the properties arising from the Lie group structure of the manifold. In [Publication I] and [Publication III] we propose Riemannian steepest descent algorithms on $U(n)$ and in [Publication II] we propose a Riemannian conjugate gradient algorithm on $U(n)$. In the following part of this subsection few relevant applications of optimization on manifolds are presented. They include subspace techniques, MIMO communications, blind source separation, and array signal processing in general.

Subspace estimation and tracking

Subspace techniques are fundamental tools in signal processing. Among the most common applications are high-resolution frequency estimation, direction finding used for smart antennas [170, 178], beamforming, delay estimation and channel equalization. A thorough literature survey of the algorithms for tracking the extreme eigenvalues and/or eigenvectors in signal processing up to year 1990 may be found in [53]. Since then, many new algorithms for subspace estimation and tracking with different complexity and performance have been developed. One of the most common algorithms is Projection Approximation Subspace tracking (PAST and PASTd) [11, 215, 216] which is mostly used for direction of arrival estimation. Other algorithms in the literature are designed for principal or minor subspace tracking in blind source separation [60, 63, 64, 66, 205].

Many authors approach the subspace estimation and tracking as an optimization problem on the Grassmann manifold $\mathrm{Gr}(n,p)$ [17–19, 85, 86, 122, 186]. The subspaces represent points on $\mathrm{Gr}(n,p)$ and their time variation correspond to a trajectory on the manifold. In addition to the observations, a stochastic model for the dynamics of the subspaces may be employed. Most of the algorithms in the literature use simple models to predict the motion or the rotation of the subspaces. In these models, the best estimate of the subspace at one time instance is simply the current value of the subspace. These algorithms do not take into account any information on the dynamic behavior, i.e., "the observed motion of the subspaces". This lack of predictability comes from the fact that the subspaces cannot be associated with vector quantities moving in a finite-dimensional vector space, described by a conventional state-space model, such as the Kalman filter. Algorithms which may take into account the subspace dynamics have been proposed, and they operate on the Grassmann manifold [85, 86, 186]. In general, the complexity is relatively high, but they posses very good tracking capabilities. As suggested in [138], a potential research area is to extend the state-space model to nonlinear settings by replacing the addition operation by a Lie group action, such as multiplication. An example of the group action is the rotation applied to an orthonormal matrix.

Some applications require estimating the exact set of eigenvectors, not only the subspace they span. In this case, the optimization should be performed on the Stiefel manifold. Classical algorithms for computing eigenvectors [53] are formulated on the Riemannian manifolds, such as Jacobi-type methods [112] (see [Publication III] and [62]) and Rayleigh Quotient Iteration (RQI) [18, 19, 136, 148, 183]. In this work, we propose Riemannian algorithms for computing the complete basis of eigenvectors (see [Publication I], [Publication II] and [Publication III]). They operate on the Lie group of unitary matrices $U(n)$.

MIMO communication systems

Communication systems with multiple transmit and/or receive antennas achieve high capacity and link reliability [194]. Sending data on the eigenmodes of the MIMO channel is an important practical transmission scheme. The unitary matrices play an crucial role in the beamformer design. The Stiefel manifold is relevant for this type of applications, since the exact eigenvectors are required. Orthonormal coding matrices are also used in space-time coding [111].

A blind identification approach is applied to beamforming in [45] based on a Joint Approximate Diagonalization of Eigen-matrices (JADE). The JADE approach relies on the independence of the sources, by exploiting the statistical information of the fourth order cumulants. Without knowing the array manifold, the beamforming is made robust to antenna array imperfections, so no physical modeling or calibration is needed. The diagonalization of the fourth order cumulant matrices is formulated as an optimization problem under unitary matrix constraint. This problem is addressed by proposing a steepest descent algorithm on the unitary group in [Publication I] and a conjugate gradient algorithm on the unitary group in [Publication II]. The proposed technique outperforms the classical JADE optimization approach [45] based on Givens rotations, especially when the number of signals is relatively large. This type of application has also been formulated as an optimization problem on the Stiefel manifold [149], and recently on the oblique manifold [16]. A Riemannian optimization technique on the oblique manifold has been applied for optimizing the transmit beamforming covariance matrix of MIMO radars [34,82,167] in [21]. In [192] a constrained beamformer design is considered. The problem may also be formulated as an optimization under unitary matrix constraint. In [185] the beamforming is used to maximize the signal-to-interference-plus-noise ratio (SINR). The optimal weight is determined by using linear and nonlinear conjugate algorithms. They operate on the Euclidean space and the Stiefel manifold, respectively. A LS approach to blind beamforming was adopted in [209]. The problem is decomposed into two stages. First, a whitening procedure is applied to the received array vector, which transforms the array response matrix into a unitary matrix. The second step is a unitary rotation. The rotation matrix is determined from the fourth order cumulants, similarly the the JADE algorithm [45].

Riemannian gradient algorithms for array signal processing have also been proposed in [187]. The optimum weight coefficients are vectors with constant magnitude, but variable phase. Therefore only a phase-nulling approach is used to maximize the SINR. This is done by using Riemannian conjugate gradient and Newton algorithms. In [197] a tracking solution for the downlink eigenbeamforming in Wideband CDMA is proposed. The unitary constrained optimization is performed by using Givens rotations.

The complexity may be decreased by formulating the problem on the Stiefel manifold. A blind source separation approach on single/multi-user MIMO communication systems has been adopted in [156]. The algorithm requires maximizing a multi-user kurtosis criterion under unitary matrix constraint. This is done by using classical Euclidean steepest descent combined with Gram-Schmidt orthogonalization procedure after every iteration (see Section 2.2.1). In [171] a blind source separation approach for the Bell Labs lAyered Space-Time coding (BLAST) architecture [210] is proposed. The algorithm is based on a multi-modulus algorithm, which leads to the same unitary optimization problem as in [156]. Similar constant-modulus criteria to be minimized under unitary matrix constraint are employed in [130, 157, 172].

Blind source separation

Separating signals blindly may be done by exploiting the statistical properties of the transmitted signals. Amari [25] proposed the *natural gradient* algorithm for blind separation. The learning algorithm operates on the Lie group of invertible matrices and it has been proved to be Fisher efficient by means of information geometry [24]. Cardoso and Laheld [44] developed Equivariant source Separation (EASI) via Independence. The concept of matrix multiplicative group is considered and the resulting algorithm is called *relative gradient*. The algorithm provides "isotropic convergence" similarly to the Newton algorithm. The connection between the natural gradient and the relative gradient has been established in [75, 161].

Douglas and Kung consider the blind source separation with orthogonality constrains and propose the ordered rotational KuickNet algorithm [65]. The algorithm discretizes the geodesic motion on the Stiefel manifold. Even though [25, 44, 65, 230] consider the matrix group concept, the update of the corresponding algorithms is not based on the group operation. Additive update is used instead. For this reason the constraint need to be restored by separate procedures in each iteration. A conjugate gradient algorithm for blind separation of temporally correlated signals which exploits the group properties of the group of invertible matrices $GL(n)$ is proposed in [229]. In this way, the undesired trivial solution which the Euclidean gradient algorithms would converge to (the zero matrix), is avoided.

Differential geometry-based learning algorithms on the orthogonal group for blind separation have been proposed first by Fiori *et. al* [73,76,78,79] and Nishimori [150]. Algorithms operating on the Stiefel and on the Grassmann manifolds have also been proposed [74, 75]. More recent relevant work in this area is [46, 77, 151]. Plumbley [161, 162] proposed Lie group methods for non-negative ICA, i.e., a steepest descent on the orthogonal group. All the above mentioned steepest descent algorithms are designed only for real-valued matrices and sources. These algorithms are not directly suitable for unitary matrices, as it will be shown later in Chapter 3.

A reliable alternative for solving the blind separation problem is the JADE algorithm proposed by Cardoso and Souloumiac [45]. The JADE algorithm consists of two stages. First, a *pre-whitening* of the received signal is performed. The second stage is a *unitary rotation*. This second stage is formulated as an optimization problem under unitary matrix constraint since no closed-form solution can be provided except for simple cases such as 2-by-2 unitary matrices. It should be noted that the first stage can also be formulated as a unitary optimization problem as in [Publication II] and [Publication III]. In order to solve for the unitary rotation we propose a steepest descent (SD) which fully exploits the benefits of the Lie group of unitary matrices $U(n)$ [Publication I] . For this reason the complexity per iteration is lower compared to the steepest descent in [137, 149]. In general, conjugate gradient (CG) converges faster than SD [183]. This happens also in the case of the JADE cost function [Publication II], when the input signals are not identically distributed. Moreover, the computational complexity is comparable to the one of the steepest descent in [Publication I] and [Publication III]. The reduction in complexity is achieved by exploiting the additional group properties of $U(n)$, when computing the search directions, as well as special matrix structures associated with the search directions. The *almost periodic property* of a smooth cost function along geodesics on $U(n)$ enables efficient search along geodesics when adapting the step size parameter [Publication IV], [Publication II]. General algorithms for optimization under unitary matrix constraint operating on the complex Stiefel manifold of $n \times n$ unitary matrices have also been proposed [137, 149]. The algorithms in [137, 149] are very general, in the sense that the local parametrization is chosen for the Stiefel and the Grassmann manifolds. For this reason, when applied to $n \times n$ unitary matrices they do not exploit the additional Lie group properties of $U(n)$, in order to reduce the computational complexity.

The FastICA algorithm [113] has been recently extended to Independent Subspace Analysis (ISA) in [179]. The corresponding algorithms are based on optimization on the Grassmann manifold. Other ICA and ISA algorithms operating on the flag manifold have been recently proposed in [152]. Robust Accurate Direct ICA aLgorithm (RADICAL) on the orthogonal group has been proposed in [26].

Linear algebra applications

Various linear algebra problems may be solved iteratively by using tools of Riemannian geometry. The most popular matrix decompositions such as the (generalized) eigendecomposition (ED) and the singular value decomposition (SVD) may be formulated in terms of descent equations on differentiable manifolds [14, 18, 40, 43, 52, 57, 112, 182–184]. The SVD has direct applications in Least Squares (LS) estimation, low-rank approximation, matrix inversion, and subspace techniques. For the low-rank approximation

SVD is optimal under the Frobenius norm, but this is no longer true under weighted norms. Moreover, no closed-form solution exists, in general. The weighted low-rank approximation may be formulated as a minimization problem on the Grassmann manifold in [140]. The convolutive reduced-rank Wiener filtering has also been formulated as an optimization problem on the Grassmann manifold in [139]. Several Least Squares (LS) matching problems may be solved by minimizing an error criterion on a suitable manifold [40, 43, 52, 69]. A typical example is encountered in image processing applications where matching points in one image with point on the second image are required. Matching is often very difficult because of the large number of possibilities. Hence, approximate solutions may be of interest. In this case, the matching problem reduces to finding an optimum orthogonal matrix and a permutation matrix [43]. Other image processing application are motion estimation in computer vision [133, 146], or biomedical applications [20, 164].

2.3.2 Quantization on manifolds

Quantization on manifolds requires approximating arbitrary points on the manifold by elements of a finite set of points on the manifold. The finite set of manifold-valued points is called code book. The goodness of the approximation is defined in terms of Riemannian distance. The code book design is crucial for the performance of the quantizer. It supposes maximizing the minimum distance between the code words. Quantization on Grassmann manifold has straight-forward application to limited-feedback MIMO communication systems [50,51,121,143–145] and space-time codes [103, 116, 118, 120, 141, 144].

Closed-loop MIMO systems

Often, in MIMO communication systems in order to increase capacity, a low-rate feedback channel is used to provide channel state information back to the receiver. This information needs to be quantized in order to reduce the transmission rate of the feedback link. Unitary/orthonormal matrices play again an important role. The quantization aims at describing orthonormal matrices by as few parameters as possible, with sufficient accuracy.

In [144] a Grassmann code book design strategy for MIMO systems is provided. The goal is to achieve full diversity and significant array gain in an uncorrelated fading channel by using the channel knowledge at the transmitter. The design is based on minimizing a “chordal distance”, which is a length defined on the manifold in order to describe the distortion introduced by quantization. The results are applied to performance analysis of a MIMO wireless communication system with quantized transmit beamforming and quantized receive combining. A similar Riemannian approach has also been considered in [145].

In [143] algorithms for quantized MIMO-OFDM systems are developed. The scheme uses a quantized feedback link in order to provide the channel state information (CSI) at the transmitter and achieve capacity and diversity gains. The motivation is that the existing schemes designed for flat-fading channel do not extend naturally to frequency selective channels due to an enormous feedback overhead. Two classes of algorithms for quantizing the channel information are considered. They are named "clustering algorithms" and "transform algorithms", respectively. The clustering algorithms group the subcarriers in clusters and choose a common frequency domain representation for each group. Thus the feedback rate depends on the number of groups and not on the number of subcarriers. The transform algorithms quantize the channel information in time domain where the transform essentially decorrelates the channel information. Both algorithms provide significant compression of the channel information maintaining the bit error rate close to the case of perfect channel knowledge.

A spatial multiplexing scheme with multi-mode precoding for MIMO-OFDM system is considered in [121]. Multi-mode precoding used linear transmit precoding, but adapts the number of transmit data streams or modes according to the channel conditions, therefore, it achieves high capacity and reliability. Typically, for OFDM scheme the multi-mode precoding requires complete knowledge of the transmit precoding matrices for each subcarrier at the transmitter. The authors propose an alternative way to reduce the feedback rate by quantizing the precoding matrices of a fraction of the number of the subcarriers and obtaining the other precoders using interpolation. The subcarrier mode selection, the precoder-quantizer design and the interpolation are addressed in the paper. It is found that unitary matrices cannot be interpolated by using linear interpolation techniques due to the fact that they do not form a vector space, but a group, i.e., the unitary group. Two algorithms for interpolation of unitary matrices are proposed, namely "geodesic interpolation" and "conditional interpolation", respectively. The geodesic interpolation has a result a point which lies halfway on the geodesic connecting two points. The method exploits the fact that the right singular vectors of the channel matrix are ambiguous up to a diagonal unitary matrix. These additional degrees of freedom are used to identify the smoothest interpolation path between adjacent quantized points. The conditional interpolation considers the interpolation of MIMO channel matrices acquired on the pilot subcarriers.

Recent work on channel adaptive quantization for limited feedback MIMO Beamforming systems may be found in [50]. Compared to [121, 143, 144] the quantization algorithm is designed for correlated Rayleigh fading MIMO channels. A Grassmannian switched code book is used to exploit the inherent spatial and temporal correlation of the channel.

In [51] an interpolation-based unitary precoding for spatial multiplexing

MIMO-OFDM with limited feedback is considered. The algorithm exploits the fact that OFDM transmission converts the frequency-selective channel into multiple narrow-band flat-fading sub-channels. Operating on each subcarrier may be costly, especially if their number is large. Therefore, the precoding algorithm operates on groups of subcarriers and interpolation techniques on the Grassmann manifold are proposed.

Space-time codes

Space-time codes improve the reliability of the radio links by providing a good trade-off between data rate and diversity gain [111]. The Grassmann space-time codes become more and more popular due to their ability to use all the degrees of freedom of the MIMO system, i.e., $M \times (1-(M/T))$ symbols per channel use, where M is the number of transmit antennas and T is the temporal length of the space-time code. In [120] a family of Grassmann space-time codes for non-coherent MIMO systems is proposed. The codes exploit all degrees of freedom of the Grassmann manifold $\mathrm{Gr}(T, M)$. The code design and also the decoding are based on minimum chordal distance, similarly to [144].

Unitary matrices play an important role in space-time coding. A capacity efficient scheme using isotropically random unitary space-time signals is proposed in [141]. The signals transmitted across antennas, viewed as a matrix with spatial and temporal dimensions form a unitary matrix. A unitary space-time modulation scheme via Cayley transform is proposed in [118]. No channel knowledge is required at the receiver. The scheme is suitable for wireless systems where channel tracking is unfeasible, either because of rapid changes in the channel characteristics or because of the limited system resources. Similar transmission schemes have been considered in [103]. The codes may be decoded by using a sphere decoder [203] algorithm (often with cubic complexity) near to ML performance. In [165], Cayley differential unitary space time codes for MIMO-OFDM are proposed. These codes possess excellent features. They allow effective high-rate data transmission in multi-antenna communication systems, with reasonable encoder and decoder complexity.

Differential geometry methods may be combined with information theory to help understanding and interpreting different problems. The resulting methods are called information geometry methods. The LDPC codes which are powerful and practical error correction codes may be analyzed by using information geometry [116].

2.3.3 Statistics on manifolds

When dealing with estimation of parameters which are constrained to a specific subset of the Euclidean space by some smooth constraints, the classical estimation techniques must be revisited [24, 188, 189, 191, 192, 212]. Estimators of manifold-valued parameters as well as their statistical bounds need to be derived on the space where the parameters are constrained. This space is usually a curved space, i.e., a Riemannian manifold with non-zero sectional curvature. They occur frequently in practice, for example when the parameters are defined on a sphere, or when estimating eigenvectors or subspaces. Another class of such problems occurs when the parameters can be estimated only up to certain ambiguities, like in blind channel estimation and blind source separation. Concepts such as bias and variance need to be defined in the proper parameter space, i.e., the constrained set. Estimators which are unbiased on the Euclidean space may be biased on the Riemannian space. The bias should be measured by using the distance defined on the corresponding Riemannian manifold, instead of the Euclidean distance. Also statistical performance bounds such as the Cramér-Rao Lower Bound (CRLB) are in general derived for parameters taking values in Euclidean spaces. They are no longer valid for constrained parameters defined on differential manifolds. Ignoring the constraint when deriving such bounds may lead to singular Fisher information matrix (FIM) due to too many degrees of freedom. Often the pseudo-inverse of the FIM is considered. The geometric interpretation of this was given in [212].

CRLB for estimating parameters with differentiable deterministic constraints have been derived by Stoica *et al.* [192]. The unconstrained Fisher information matrix, which is not necessarily of full rank is replaced by a constrained FIM determined from the smooth constraint. This bound is still expressed in terms of Euclidean distance. Its accuracy is expected to degrade since the curvature of the corresponding Riemannian space is neglected. An Intrinsic Variance Lower Bound (IVLB) on Riemannian manifolds has been derived in [213]. The IVLB is a lower limit on the estimation accuracy, measured in terms of the mean-square Riemannian distance. For parameters defined in Euclidean spaces (zero sectional curvature), the IVLB coincides with the classical CRLB.

Estimation Bounds on arbitrary manifolds in which no set of extrinsic coordinates exist have been established recently by Smith, in [188, 189]. The frequently encountered examples of estimating either an unknown subspace or a covariance matrix are examined in detail in [188]. Intrinsic versions of the Cramér-Rao bound on manifolds are derived for both biased and unbiased estimators. Remarkably, it is shown in [188] that from an intrinsic perspective, the sample covariance matrix is a biased and inefficient estimator. The bias term reveals the dependency of estimator on the limited sample support observed in practice. For this reason, the natural invariant metric is recommended over the flat metric (Euclidean norm) for analysis of covariance matrix estimation.

The capacity of non-coherent MIMO fading channels is derived in [231] by using a Riemannian geometry approach. The scenario of fast fading is considered, thus an accurate estimation of the fading coefficients is not available to either the transmitter or the receiver. A geometric interpretation of the capacity expression on the Grassmann manifold is given.

Monte-Carlo extrinsic estimators of manifold-valued parameters have been considered in [191]. The estimation of means and variances of manifold-valued parameters is considered, using two popular sampling methods (independent and importance sampling). The results are applied to target pose estimation on the orthogonal group and subspace estimation on the Grassmann manifold.

Chapter 3

Practical Riemannian algorithms for optimization under unitary matrix constraint

3.1 The unitary group $U(n)$ as a real manifold

Most of the Riemannian optimization algorithms in the literature [40, 41, 46, 69, 73, 76–79, 150, 151, 161, 183, 214] are designed for optimization on the orthogonal group $O(n)$, i.e., they consider only real-valued matrices. Very often, in communications and sensor array signal processing applications we are dealing with complex-valued matrices and signals. Consequently, the optimization needs to be performed under unitary matrix constraint, i.e., on the unitary group $U(n)$. Often, and unfairly, extending algorithms designed for real-valued matrices to complex-valued matrices is considered to be trivial. Commonly, this is done in a simplistic manner by changing a real-valued result into a complex-valued one, just by replacing the transpose operation with the Hermitian transpose, and skipping all the intermediate derivation steps. In many cases, the result holds, but there are cases when this simplistic approach fails, leading to wrong results. We will show at the end of this section an illustrative example where the simplistic approach fails. Thus, a proper derivation of the algorithms dealing with complex-valued matrices is required.

In order to be able to derive optimization algorithms on the unitary group $U(n)$, it is important to know that the Lie group $U(n)$ of $n \times n$ unitary matrices is a *real differentiable manifold*. We show how all complex algebraic operations can be mapped into real operations and vice versa, and describe a convenient way of real differentiation using complex-valued matrices and operations.

3.1.1 Revealing the real Lie group structure of $U(n)$

A Lie group is a differentiable manifold [59] and a group as the same time, with the property that the group operations are differentiable [107,123,208]. The $n \times n$ unitary matrices are closed under the standard matrix multiplication and they form the unitary group $U(n)$. Even though the elements of $U(n)$ are represented by complex-valued matrices, the unitary group is a *real Lie group*, i.e., it possesses *real differentiable structure*. Complex Lie groups are the ones whose multiplication operation is compatible with *complex analytic* manifold structure [123]. Some groups possess both real and complex differentiable structure (e.g. complex general linear group $GL(n)$, complex special linear group $SL(n)$, and so on) [107, Ch. 8]. Although the Lie group of unitary matrices $U(n)$ cannot be viewed as a complex Lie group, it is important to understand that the complex manifold structure is generally useless in the optimization context. This is because we are always dealing with real-valued cost functions of complex-valued argument. Such functions are not complex differentiable, unless they are constant functions. Instead, they may be differentiable w.r.t. to the real and the imaginary parts of the complex argument.

Any complex-valued matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ can be mapped into real-valued matrix $\mathbf{A}_{\mathbb{R}} \in \mathbb{R}^{2n\times 2n}$, by using its real and imaginary parts $\mathbf{A}_R \triangleq \Re\{\mathbf{A}\}$ and $\mathbf{A}_I \triangleq \Im\{\mathbf{A}\}$, respectively, as follows:

$$\mathbf{A} = \mathbf{A}_R + \jmath\mathbf{A}_I \quad \longleftrightarrow \quad \mathbf{A}_{\mathbb{R}} \triangleq \begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \\ \mathbf{A}_I & \mathbf{A}_R \end{bmatrix}. \tag{3.1}$$

It is straight-forward to verify that the mapping (3.1) is differentiable, and the following equalities hold for any $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$:

$$\begin{aligned}
(t\mathbf{A})_{\mathbb{R}} &= t\mathbf{A}_{\mathbb{R}}, \quad \forall t \in \mathbb{R}, && (3.2)\\
(\mathbf{A}+\mathbf{B})_{\mathbb{R}} &= \mathbf{A}_{\mathbb{R}} + \mathbf{B}_{\mathbb{R}}, && (3.3)\\
(\mathbf{A}\mathbf{B})_{\mathbb{R}} &= \mathbf{A}_{\mathbb{R}}\mathbf{B}_{\mathbb{R}}, && (3.4)\\
(\mathbf{A}^{-1})_{\mathbb{R}} &= (\mathbf{A}_{\mathbb{R}})^{-1}, \quad \det\{\mathbf{A}\} \neq 0, && (3.5)\\
(\mathbf{A}^{H})_{\mathbb{R}} &= \mathbf{A}_{\mathbb{R}}^{T}, && (3.6)\\
\operatorname{trace}\{\mathbf{A}_{\mathbb{R}}\} &= 2\Re\big\{\operatorname{trace}\{\mathbf{A}\}\big\}. && (3.7)
\end{aligned}$$

Based on (3.2), (3.3) and (3.4), it also follows that:

$$\big(\exp(t\mathbf{A})\big)_{\mathbb{R}} = \exp(t\mathbf{A}_{\mathbb{R}}), \quad t \in \mathbb{R}. \tag{3.8}$$

where $\exp(\cdot)$ denotes the exponential map from the Lie algebra to the Lie group [107, Ch. II, §1]. For matrix Lie groups, this coincides with the standard matrix exponential given by the convergent power series

$$\exp(\mathbf{A}) \triangleq \sum_{m=0}^{\infty} \frac{\mathbf{A}^m}{m!}. \tag{3.9}$$

In conclusion, the mapping (3.1) reveals the real Lie group structure of $U(n)$. The main benefit is that it enables using complex-valued matrices instead of large real-valued matrices containing the real and the imaginary parts in separate blocks. Other properties of $U(n)$ are presented in detail in [Publication I].

3.1.2 Differentiation of real-valued function of complex-valued argument

The *real differentiation* of functions of complex-valued matrix arguments can be conveniently described in complex terms by the following partial derivatives [39, 124]:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{A}} \triangleq \frac{1}{2}\Big(\frac{\partial \mathcal{J}}{\partial \mathbf{A}_R} - \jmath \frac{\partial \mathcal{J}}{\partial \mathbf{A}_I}\Big) \tag{3.10}$$

and

$$\frac{\partial \mathcal{J}}{\partial \mathbf{A}^*} \triangleq \frac{1}{2}\Big(\frac{\partial \mathcal{J}}{\partial \mathbf{A}_R} + \jmath \frac{\partial \mathcal{J}}{\partial \mathbf{A}_I}\Big). \tag{3.11}$$

In practice, it would be inconvenient to split all complex-valued matrices in their real and imaginary parts because this would complicate the mathematical expressions. The operators (3.10)-(3.11) and the mapping (3.1) enable direct manipulation of the complex-valued matrices, but we have to keep in mind that in fact all the operations are applied to the real and imaginary parts. The computation of derivatives (3.10) and (3.11) has been recently addressed in [109].

3.1.3 Justification of using complex-valued matrices

In this subsection we provide an example showing that the results involving real-valued matrices do not always extend in a straight-forward manner to complex-valued matrices. The "trick" of replacing the transpose operation with the Hermitian-transpose may lead to a wrong result. Moreover, the correct result is not a trivial extension. We consider the chain-rule for differentiating a real-valued function involving real and complex matrices, respectively. A comparison between the real and the complex case is provided in Table 3.1.

As it can be seen in Table 3.1, extending the result obtained in the real-valued case to a the complex-valued case is not straight-forward. Just by replacing the transpose operation with the Hermitian-transpose leads to a wrong result. The complex-conjugation operation of one of the two factors inside the trace operator would be missed, as well as taking the real part. This chain-rule example has been selected in purpose, since it is used to differentiate the cost function along geodesics when performing the line search optimization. For details, see [Publication II] and [Publication IV].

Real-valued case	Complex-valued case
Consider a real-valued scalar function of *real-valued* matrix argument $\mathcal{J}_1 : \mathbb{R}^{n\times n} \to \mathbb{R}$, and a *real-valued* matrix function of real-valued scalar argument $\boldsymbol{\mathcal{W}}_1 : \mathbb{R} \to \mathbb{R}^{n\times n}$.	Consider a real-valued scalar function of *complex-valued* matrix argument $\mathcal{J}_2 : \mathbb{C}^{n\times n} \to \mathbb{R}$, and a *complex-valued* matrix function of real-valued scalar argument $\boldsymbol{\mathcal{W}}_2 : \mathbb{R} \to \mathbb{C}^{n\times n}$.
The composition of $\mathcal{J}_1$ and $\boldsymbol{\mathcal{W}}_1$ is $\hat{\mathcal{J}}_1(t) \triangleq (\mathcal{J}_1 \circ \boldsymbol{\mathcal{W}}_1)(t) = \mathcal{J}_1(\boldsymbol{\mathcal{W}}_1(t))$	The composition of $\mathcal{J}_2$ and $\boldsymbol{\mathcal{W}}_2$ is $\hat{\mathcal{J}}_2(t) \triangleq (\mathcal{J}_2 \circ \boldsymbol{\mathcal{W}}_2)(t) = \mathcal{J}_2(\boldsymbol{\mathcal{W}}_2(t))$
Real-case result: $\frac{\mathrm{d}\hat{\mathcal{J}}_1}{\mathrm{d}t}(t_1) = \mathrm{trace}\{\big[\frac{\mathrm{d}\mathcal{J}_1}{\mathrm{d}\boldsymbol{\mathcal{W}}_1}(\boldsymbol{\mathcal{W}}_1(t_1))\big]^T\big[\frac{\mathrm{d}\boldsymbol{\mathcal{W}}_1}{\mathrm{d}t}(t_1)\big]\}$.	Complex-case result: $\frac{\mathrm{d}\hat{\mathcal{J}}_2}{\mathrm{d}t}(t_2) = 2\Re\{\mathrm{trace}\{\big[\frac{\partial\mathcal{J}_2}{\partial\boldsymbol{\mathcal{W}}_2}(\boldsymbol{\mathcal{W}}_2(t_2))\big]^T\big[\frac{\mathrm{d}\boldsymbol{\mathcal{W}}_2}{\mathrm{d}t}(t_2)\big]\}\}$.

Table 3.1: Differentiation by using the chain rule: real-valued case vs. complex-valued case. The result on the left column is the correct result obtained in the real-valued case. Replacing the the transpose operation with the Hermitian-transpose in the real-valued result would lead to the wrong complex-valued result for $\frac{\mathrm{d}\hat{\mathcal{J}}_2}{\mathrm{d}t}(t_2)$, i.e., $\mathrm{trace}\{\big[\frac{\partial\mathcal{J}_2}{\partial\boldsymbol{\mathcal{W}}_2}(\boldsymbol{\mathcal{W}}_2(t_2))\big]^H\big[\frac{\mathrm{d}\boldsymbol{\mathcal{W}}_2}{\mathrm{d}t}(t_2)\big]\}$. The correct result obtained in the complex-valued case is the result on the right column. It may be noticed that extending the real-valued case result to a complex-valued one in not trivial.

In conclusion, the algorithms dealing with complex-valued matrices need to be derived from scratch, by using the approach presented in Subsections 3.1.1 and 3.1.2. In [Publication I] and [Publication II] we derive Riemannian steepest descent and conjugate gradient algorithms on $U(n)$.

3.2 Practical optimization algorithms along geodesics on $U(n)$

When deriving optimization algorithms on Riemannian manifolds, the geometrical properties of the parameter space play a crucial role in reducing the computational complexity. The fact that $U(n)$ is a matrix Lie group enables very simple formulas for geodesics and parallel transport. Additional computational benefits arise from exploiting the special matrix structures. The tangent space at the group identity element is the Lie algebra of skew-Hermitian matrices $\mathfrak{u}(n)$. Geodesics through an arbitrary point $\mathbf{W} \in U(n)$ can be given in terms of an exponential of skew-Hermitian matrices. Recent developments in numerical methods dealing with this type of computation can be exploited [117, 224]. For details, see [Publication I]. Another useful aspect to be considered is the fact that the exponential map induces an almost periodic behavior of the cost function along geodesics. The almost periodic property may be exploited when performing the line search needed in the step size selection [Publication IV], [Publication II].

In this section we provide Riemannian algorithms which may be used in practical applications involving optimization under unitary matrix constraint. Steepest Descent (SD) and Conjugate Gradient (CG) algorithms operating on the Lie group of unitary matrices $U(n)$ are given in Section 3.2.1 and Section 3.2.2, respectively. Efficient line search methods which can be used together with the proposed algorithms are provided in Section 3.2.3.

3.2.1 Steepest Descent Algorithm along geodesics on $U(n)$

Similarly to the Euclidean space, the main advantage of the Riemannian steepest descent (SD) algorithm is that it is very simple to implement. Each iteration of the Riemannian SD algorithm consists of two subsequent stages. The first one is the computation of the Riemannian gradient which gives the steepest ascent direction on the manifold. The second one is taking a step along the geodesic emanating in the direction of the negative gradient. The Riemannian gradient of the smooth cost function $\mathcal{J}$ at an arbitrary point $\mathbf{W}_k \in U(n)$ is given by:

$$\nabla^{\mathrm{R}}\mathcal{J}(\mathbf{W}_k) = \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_k) - \mathbf{W}_k \Big[\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_k)\Big]^H \mathbf{W}_k. \tag{3.12}$$

where $\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_k)$ is defined in (3.11) and represents the gradient of the cost function $\mathcal{J}$ on the Euclidean space at a given $\mathbf{W}$ [39, 198]. The derivation of expression (3.12) is provided in [Publication I]. The geodesic emanating from $\mathbf{W}_k$ along the steepest descent direction $-\nabla^{\mathrm{R}}\mathcal{J}(\mathbf{W}_k)$ on $U(n)$ is given by:

$$\boldsymbol{\mathcal{W}}(\mu) = \exp(-\mu \mathbf{G}_k)\mathbf{W}_k, \quad \text{where} \tag{3.13}$$

$$\mathbf{G}_k \triangleq \nabla^{\mathrm{R}}\mathcal{J}(\mathbf{W}_k)\mathbf{W}_k^H \in \mathfrak{u}(n). \tag{3.14}$$

$\mathbf{G}_k$ is the gradient of $\mathcal{J}$ at $\mathbf{W}_k$ after translation into the tangent space at the identity element. Consequently, the matrix $\mathbf{G}_k$ is skew-Hermitian, i.e., $\mathbf{G}_k = -\mathbf{G}_k^H$. The skew-Hermitian structure of $\mathbf{G}_k$ brings important computational benefits when computing the matrix exponential [Publication I], as well as when performing the line search [Publication IV]. The Riemannian SD algorithm on $U(n)$ has been derived in [Publication I] and it is summarized in Table 3.2.

3.2.2 Conjugate Gradient Algorithm along geodesics on $U(n)$

Conjugate gradient (CG) algorithm achieves in general faster convergence compared to the SD, not only on the Euclidean space, but also on Riemannian manifolds. This is due to the fact that Riemannian SD algorithm has

1	Initialization: $k = 0$, $\mathbf{W}_k = \mathbf{I}$
2	Compute the Riemannian gradient direction $\mathbf{G}_k$: $\mathbf{\Gamma}_k = \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_k), \quad \mathbf{G}_k = \mathbf{\Gamma}_k \mathbf{W}_k^H - \mathbf{W}_k \mathbf{\Gamma}_k^H$
3	Evaluate $\langle \mathbf{G}_k, \mathbf{G}_k \rangle_{\mathbf{I}} = (1/2)\text{trace}\{\mathbf{G}_k^H \mathbf{G}_k\}$. If it is sufficiently small, then stop
4	Determine $\mu_k = \arg\min_\mu \mathcal{J}(\exp(-\mu \mathbf{G}_k)\mathbf{W}_k)$
5	Update: $\mathbf{W}_{k+1} = \exp(-\mu_k \mathbf{G}_k)\mathbf{W}_k$
6	$k := k + 1$ and go to step **2**

Table 3.2: Steepest descent (SD) algorithm along geodesics on $U(n)$

the same drawback as its Euclidean counterpart, i.e., it takes ninety degree turns at each iteration [183]. This is illustrated in Figure 3.1, where the contours of a cost function are plotted on the Riemannian surface determined by the constraint. CG algorithm may significantly reduce this drawback. It exploits the information provided by the current search direction $-\tilde{\mathbf{H}}_k$ at $\mathbf{W}_k$ and the SD direction $-\tilde{\mathbf{G}}_{k+1}$ at the next point $\mathbf{W}_{k+1}$. The new search direction is chosen to be a combination of these two, as shown in Figure 3.2. The difference compared to the Euclidean space is that the current search direction $-\tilde{\mathbf{H}}_k$ and the gradient $\tilde{\mathbf{G}}_{k+1}$ at the next point lie in different tangent spaces, $T_{\mathbf{W}_k}$ and $T_{\mathbf{W}_{k+1}}$, respectively. For this reason they are not directly compatible. In order to combine them properly, the parallel transport of the current search direction $-\tilde{\mathbf{H}}_k$ from $\mathbf{W}_k$ to $\mathbf{W}_{k+1}$ along the corresponding geodesic is needed. The new search direction at $\mathbf{W}_{k+1}$ is

$$-\tilde{\mathbf{H}}_{k+1} = -\tilde{\mathbf{G}}_{k+1} - \gamma_k \tau \tilde{\mathbf{H}}_k, \tag{3.15}$$

where $\tau\tilde{\mathbf{H}}_k$ denotes the parallel transport of the vector $\tilde{\mathbf{H}}_k$ into $T_{\mathbf{W}_{k+1}}$ along the corresponding geodesic (see Figure 3.2). The weighting factor γ_k is determined such that the directions $\tau\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_{k+1}$ are Hessian-conjugate [69, 183]. The exact conjugacy would require expensive computation of the Hessian matrices. In practice an approximation of γ_k is used instead, for example the Polak-Ribièrre formula [69]. For more details, see [Publication II]. The fact that $U(n)$ is a Lie group enables describing all tangent directions (steepest descent and search directions) by tangent vectors which correspond to elements of the Lie algebra $\mathfrak{u}(n)$ via right (or left) translation. Then, all tangent vectors are represented by skew-Hermitian matrices. Thus, the new search direction on the Lie algebra $\mathfrak{u}(n)$ is

$$\mathbf{H}_{k+1} = \mathbf{G}_{k+1} + \gamma_k \mathbf{H}_k, \quad \mathbf{H}_k, \mathbf{H}_{k+1}, \mathbf{G}_{k+1} \in \mathfrak{u}(n) \tag{3.16}$$

where $\mathbf{H}_k$ is the old search direction on $\mathfrak{u}(n)$. The conjugate gradient step is taken along the geodesic emanating from $\mathbf{W}_k$ in the direction $-\tilde{\mathbf{H}}_k = -\mathbf{H}_k\mathbf{W}_k$, i.e.,

$$\boldsymbol{\mathcal{W}}(\mu) = \exp(-\mu \mathbf{H}_k)\mathbf{W}_k. \tag{3.17}$$

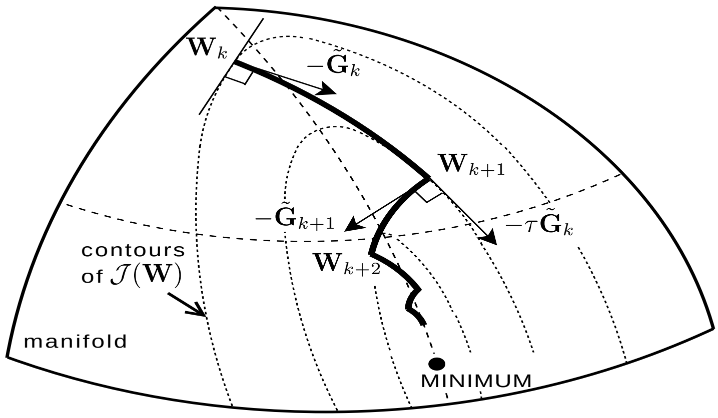

Figure 3.1: The SD algorithm takes ninety-degree turns at every iteration, i.e., $\langle -\tilde{\mathbf{G}}_{k+1}, -\tau\tilde{\mathbf{G}}_k \rangle_{\mathbf{W}_{k+1}} = 0$, where τ denotes the parallelism w.r.t. the geodesic connecting $\mathbf{W}_k$ and $\mathbf{W}_{k+1}$.

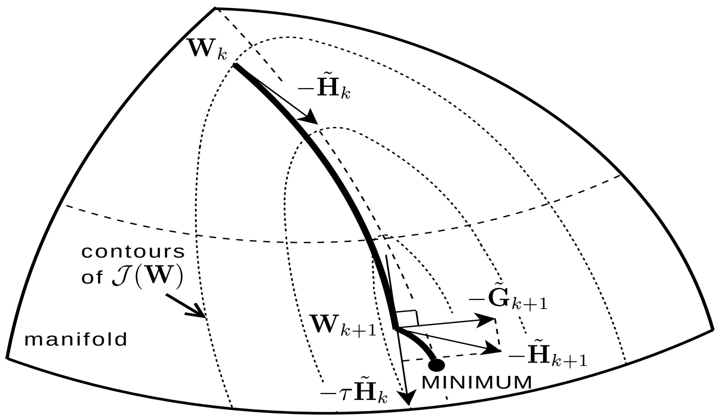

Figure 3.2: The CG takes a search direction $-\tilde{\mathbf{H}}_{k+1}$ at $\mathbf{W}_{k+1}$ which is a combination of the new SD direction $-\tilde{\mathbf{G}}_{k+1}$ at $\mathbf{W}_{k+1}$ and the current search direction $-\tilde{\mathbf{H}}_k$ transported to $\mathbf{W}_{k+1}$ along the geodesic connecting $\mathbf{W}_k$ and $\mathbf{W}_{k+1}$. The new Riemannian steepest descent direction $-\tilde{\mathbf{G}}_{k+1}$ at $\mathbf{W}_{k+1}$ will be orthogonal to the current search direction $-\mathbf{H}_k$ at $\mathbf{W}_k$ transported to $\mathbf{W}_{k+1}$, i.e., $\langle -\tilde{\mathbf{G}}_{k+1}, -\tau\tilde{\mathbf{H}}_k \rangle_{\mathbf{W}_{k+1}} = 0$.

A geodesic search needs to be performed in order to choose a suitable value of μ [Publication IV], [Publication II]. The step size selection is crucial for the performance of the CG algorithm. The Riemannian CG algorithm on $U(n)$ has been derived in [Publication II] and it is summarized in Table 3.3.

1	Initialization: $k = 0$, $\mathbf{W}_k = \mathbf{I}$
2	Compute the Riemannian gradient direction $\mathbf{G}_k$ and the search direction $\mathbf{H}_k$: if (k modulo n^2) $== 0$ $\boldsymbol{\Gamma}_k = \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_k)$ $\mathbf{G}_k = \boldsymbol{\Gamma}_k \mathbf{W}_k^H - \mathbf{W}_k \boldsymbol{\Gamma}_k^H$ $\mathbf{H}_k := \mathbf{G}_k$
3	Evaluate $\langle \mathbf{G}_k, \mathbf{G}_k \rangle_{\mathbf{I}} = (1/2)\text{trace}\{\mathbf{G}_k^H \mathbf{G}_k\}$. If it is sufficiently small, then stop
4	Determine $\mu_k = \arg\min_\mu \mathcal{J}(\exp(-\mu \mathbf{H}_k)\mathbf{W}_k)$
5	Update: $\mathbf{W}_{k+1} = \exp(-\mu_k \mathbf{H}_k)\mathbf{W}_k$
6	Compute the Riemannian gradient direction $\mathbf{G}_{k+1}$ and the search direction $\mathbf{H}_{k+1}$: $\boldsymbol{\Gamma}_{k+1} = \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W}_{k+1})$ $\mathbf{G}_{k+1} = \boldsymbol{\Gamma}_{k+1} \mathbf{W}_{k+1}^H - \mathbf{W}_{k+1} \boldsymbol{\Gamma}_{k+1}^H$ $\mathbf{H}_{k+1} = \mathbf{G}_{k+1} + \gamma_k \mathbf{H}_k$
7	$k := k + 1$ and go to step **2**

Table 3.3: Conjugate gradient (CG) algorithm along geodesics on $U(n)$.

Riemannian CG algorithm achieves superlinear convergence, whereas the Riemannian SD converges only linearly [87, 183, 196]. On $U(n)$, the computational complexity of the CG is comparable the one of the SD, due to the fact that the parallel transport is easy to perform. This is not always true on general Riemannian manifolds. Both SD on $U(n)$ introduced in [Publication I] and CG on $U(n)$ introduced in [Publication II], exhibit cubic complexity in n per iteration, i.e., $\mathcal{O}(n^3)$. This property seems to be unavoidable, since a trivial multiplication of two $n \times n$ matrices already requires about $2n^3$ flops [99].

CG is considerably simpler than a Newton algorithm which would require computing costly second-order derivatives. CG algorithm captures the second-order information by computing successive first-order derivatives and combining them properly. Newton algorithms on general Riemannian manifolds [87, 155, 183, 196] are computationally expensive also due to the fact that they do not take into account the particular properties that may appear on certain manifolds, such as special matrix structures. Newton algorithms on Stiefel and Grassmann manifolds are proposed by Edelman *et. al* [69]. Other Newton algorithms on Stiefel and Grassmann manifolds have been proposed in the literature [17, 130, 137]. When applied on $U(n)$, they have complexity is of order $\mathcal{O}(n^6)$. Due to computational reasons, in this thesis we treat only SD and CG on $U(n)$ and provide computationally feasible solutions. Moreover, Newton algorithm is not guaranteed to converge, not even locally [69, 137, 183, 184]. It may converge to any stationary points unless some strict requirements (such as convexity of the cost function) are satisfied. Trust-region methods on Riemannian manifolds have been recently proposed [15, 18], to overcome this drawback. In conclusion, Riemannian CG algorithm provides a reliable alternative for optimization under unitary matrix constraint.

3.2.3 Efficient Line search methods on $U(n)$

Once the search direction corresponding to the SD, or to the CG algorithm has been chosen, a line search needs to be performed in order to select an appropriate step size. The line search supposes minimizing (or maximizing) the cost function $\mathcal{J}(\mathbf{W})$ in one dimension, along the curve $\boldsymbol{\mathcal{W}}(\mu)$ describing the local parametrization, i.e., find

$$\mu_k = \arg\min_{\mu} \hat{\mathcal{J}}(\mu), \quad \text{where} \tag{3.18}$$

$$\hat{\mathcal{J}}(\mu) \triangleq \mathcal{J}(\boldsymbol{\mathcal{W}}(\mu)) \tag{3.19}$$

Line search usually requires expensive operations, even in the case of Euclidean optimization algorithms [163], due to multiple cost function evaluations. On Riemannian manifolds, the problem becomes even harder because every evaluation of the cost function requires expensive computations of the local parametrization. For this reason, the choice of the local parametrization plays a crucial role in reducing the computational complexity. In case of $U(n)$, exponential map possesses desirable properties that may be exploited.

In [Publication I], Armijo method [163] is efficiently used to perform search along geodesics. The step size μ_k evolves in a dyadic basis. By exploiting the properties of the exponential map, the computation of the matrix exponential may often be avoided. Reduction in complexity by half is achieved when Armijo method is used together with the geodesic SD algorithm, compared to the non-geodesic SD algorithm in [137]. The computational issues are addressed in detail in [Publication I], Section V.

An important property of the exponential map which can be exploited in line search (see [Publication I] and [Publication IV]) is that it induces an *almost periodic* [80] behavior of the cost function along geodesics on $U(n)$. The almost periodic functions is a well-studied class of functions [181]. There are many definition for these type of functions [33,35,54,80,127]. We present the most intuitive one as in [80,81]. A real number T is called ϵ-*almost period* (or just *almost period*) of the function $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ if

$$|\mathcal{F}(t+T) - \mathcal{F}(t)| \leq \epsilon, \forall t \in \mathbb{R}. \tag{3.20}$$

The function $\mathcal{F}$ is called almost periodic if for any $\epsilon > 0$, the set of ϵ-almost periods is relatively dense in $\mathbb{R}$ [80]. An almost periodic function can also be expressed as a trigonometric polynomial

$$\mathcal{F}(t) = \sum_{m=1}^{q} b_m \exp(\jmath\omega_m t), \tag{3.21}$$

where $b_1, \ldots, b_q \in \mathbb{C}$, $\omega_1, \ldots, \omega_q \in \mathbb{R}$, and $q \in \mathbb{N}$. If the numbers $\omega_1, \ldots, \omega_q$ are in harmonic relation, the above expression represents the classical Fourier series of a periodic function comprised of q harmonic components. For almost periodic functions, the frequencies $\omega_1, \ldots, \omega_q$ are non-harmonic.

The almost periodic property of a cost function along geodesics on $U(n)$ is a consequence of the fact that geodesics are expressed in terms of exponential of skew-Hermitian matrices. This special property appears only on certain manifolds such as the unitary group $U(n)$ and the special orthogonal group $SO(n)$, and it does not appear on Euclidean spaces or on general Riemannian manifolds. For this reason, other local parametrizations designed for more general Riemannian manifolds, such as Stiefel and Grassmann manifolds [69, 137] exhibit higher complexity on $U(n)$.

It will be shown next that the almost periodic property appears for the exponential map. For other common parametrizations such as the Euclidean projection operator or the Cayley transform (for details see [Publication I]) the property does not appear. These parametrizations do not take into account the special structure of the tangent vectors at the group identity. Any search direction $\tilde{\mathbf{H}}_k$ at $\mathbf{W}_k$ corresponds via right translation to a skew-Hermitian matrix into the Lie algebra $\mathfrak{u}(n)$:

$$\tilde{\mathbf{H}}_k \in T_{\mathbf{W}_k}U(n) \quad \longleftrightarrow \quad \mathbf{H}_k = \tilde{\mathbf{H}}_k \mathbf{W}_k^H \in \mathfrak{u}(n). \tag{3.22}$$

Skew-Hermitian matrices have purely imaginary eigenvalues of the form $\jmath\omega_i$, $i = 1, \ldots, n$. Consider the eigendecomposition of $\mathbf{H}_k$

$$\mathbf{H}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^H, \quad \mathbf{U}_k \in U(n), \quad \mathbf{D}_k = \begin{bmatrix} \jmath\omega_1 & 0 & \ldots & 0 \\ 0 & \jmath\omega_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \jmath\omega_n \end{bmatrix} \tag{3.23}$$

where $\mathbf{U}_k$ are the eigenvectors of $\mathbf{H}_k$, and $\mathbf{D}_k$ is a diagonal matrix containing the eigenvalues along the diagonal. Next, different local parametrization on $U(n)$ will be considered (see [Publication I]) and the differences among them from the line search perspective are explained.

Exponential map

Geodesic update on $U(n)$ at iteration k is expressed in terms of exponential of skew-Hermitian matrices

$$\begin{aligned} \boldsymbol{\mathcal{W}}_{\text{geod}}(\mu) &= \exp(-\mu\mathbf{H}_k)\mathbf{W}_k & (3.24) \\ &= \mathbf{U}_k \exp(-\mu\mathbf{D}_k)\mathbf{U}_k^H \mathbf{W}_k. & (3.25) \end{aligned}$$

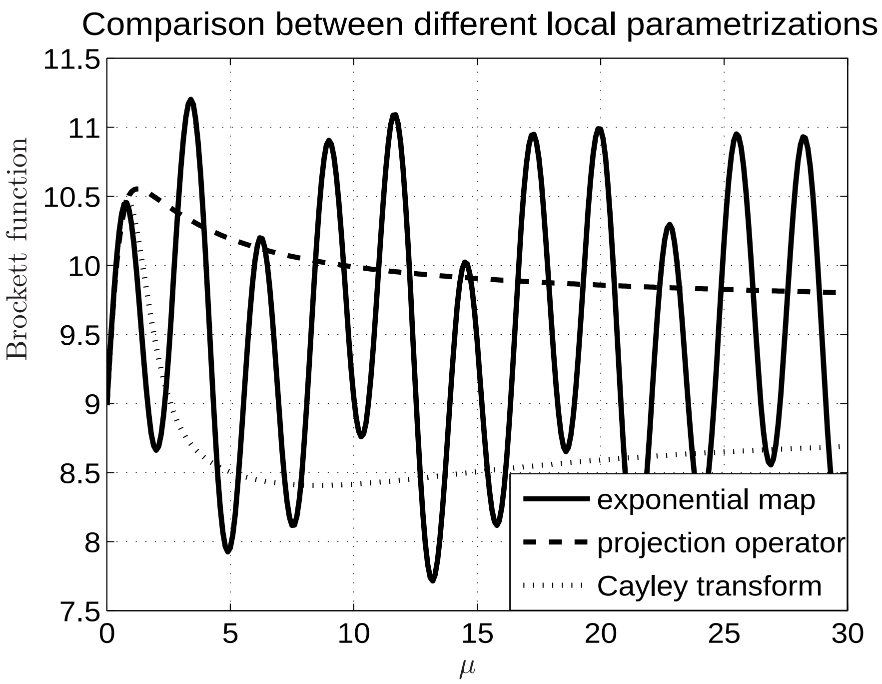

Figure 3.3: Behavior of the Brockett function [41, 184] (see also [Publication II] and [Publication V]) along different parametrizations on $U(n)$. The exponential map (3.24) is represented by continuous line. The projection operator (3.27) is represented by dashed line. The Cayley transform (3.30) is represented by dotted line. The exponential map induces an almost periodic behavior of the cost function along geodesics. For the other two parametrizations, the behavior is not almost periodic.

From (3.23) it follows that the matrix $\exp(-\mu \mathbf{D}_k)$ is a diagonal matrix whose diagonal elements are complex exponentials of the form $e^{-\jmath\omega_i\mu}$, $i = 1, \ldots, n$. Consequently, each element of $\boldsymbol{\mathcal{W}}_{\text{geod}}(\mu)$ is a sum of complex exponentials, as in Eq. (3.21). The cost function evaluated along the geodesic

$$\hat{\mathcal{J}}_{\text{geod}}(\mu) \triangleq \mathcal{J}(\boldsymbol{\mathcal{W}}_{\text{geod}}(\mu)) \tag{3.26}$$

is an *almost periodic function* [80], due to the fact that $\mathcal{J}$ is a smooth function of $\boldsymbol{\mathcal{W}}_{\text{geod}}(\mu)$. As an example, the behavior of the Brockett function [41, 184] along geodesics is shown in Figure 3.3 by continuous line. The almost periodic behavior may be exploited in line search optimization [Publication II] and [Publication IV].

Euclidean projection map

The Euclidean projection map supposes moving along a straight line tangent to the manifold and projecting back into the manifold [137]. The projection operator is defined in (2.5). It approximates the matrix exponential at the origin up the the second order, as shown in [Publication I]. The update at iteration k may be written as:

$$\begin{aligned} \boldsymbol{\mathcal{W}}_{\text{proj}}(\mu) &= \boldsymbol{\mathcal{P}}\{\mathbf{W}_k - \mu\mathbf{H}_k\mathbf{W}_k\} \qquad (3.27) \\ &= \mathbf{U}_k\boldsymbol{\mathcal{P}}\{\mathbf{I} - \mu\mathbf{D}_k\}\mathbf{U}_k^H\mathbf{W}_k. \qquad (3.28) \end{aligned}$$

From (2.5) and (3.23) it follows that the matrix $\boldsymbol{\mathcal{P}}\{\mathbf{I} - \mu\mathbf{D}_k\}$ is a diagonal matrix whose diagonal entries are of the form $(1 - \jmath\omega_i\mu)/|1 - \jmath\omega_i\mu| = e^{\jmath\angle(1-\jmath\omega_i\mu)}$, $i = 1, \ldots, n$. Consequently, the corresponding rotation angles $\angle(1 - \jmath\omega_i\mu)$ are all confined within the interval $[0, \pi/2)$, regardless how much μ is increased. This fact can be easily understood by taking as an example the unit circle $U(1)$. Moving along a straight line tangent to the circle towards infinity, and projecting back ends up to a point which is rotated ninety degrees from the starting point. Therefore, the projection operator cannot produce rotations larger than ninety degrees. The exponential map, on the other hand, can span the whole unit circle, and the cost function is periodic along the circle. The illustrative example in Figure 2.3, Section (2.2.3) may clarify these explanations. The variation of the rotation angle on $U(1)$ w.r.t. μ is shown in Figure 3.4. The exponential map is represented by continuous line and the projection operator, by dashed line. The same angle limitation of the projection operator appears also in the multi-dimensional case. By increasing μ towards infinity, the projection map will converge to a fixed matrix

$$\lim_{\mu\to\infty}\boldsymbol{\mathcal{P}}\{\mathbf{W}_k - \mu\mathbf{H}_k\mathbf{W}_k\} = \mathbf{U}_k \begin{bmatrix} \jmath\text{sign}(\omega_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \jmath\text{sign}(\omega_n) \end{bmatrix} \mathbf{U}_k^H\mathbf{W}_k. \qquad (3.29)$$

The values of the Brockett function [41, 184] (see also [Publication II] and [Publication V]) along different curves on $U(n)$ is shown in Figure 3.3. It may be noticed that the function is not almost periodic along the curve described by the projection operator (3.27) (dashed line), as in the case of the exponential map (solid line).

Cayley transform

A local parametrization based on the Cayley transform is also a second-order approximation of the matrix exponential at the origin (see [Publication I]).

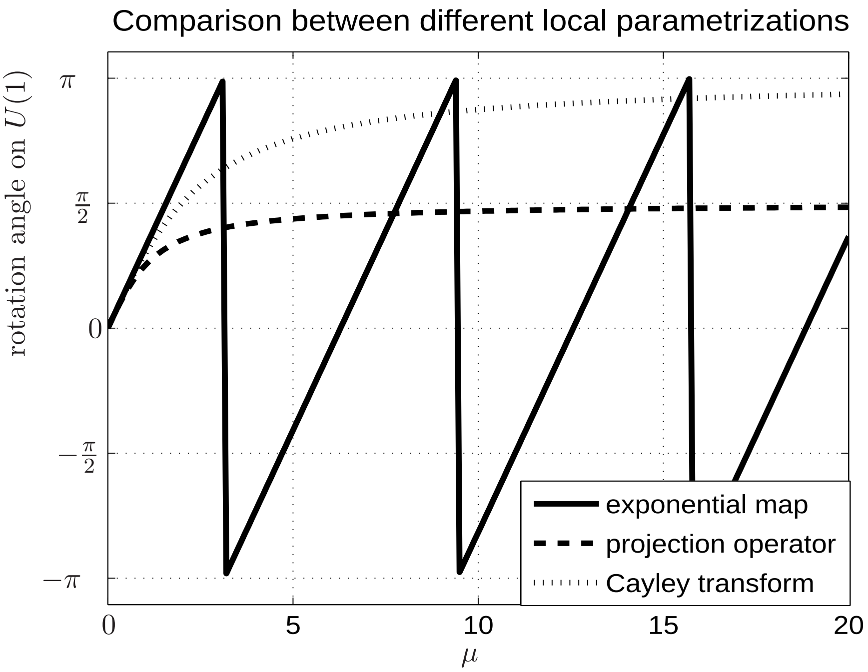

Figure 3.4: The rotation angle on the unit circle $U(1)$ for different local parametrizations. The initial point is the zero angle, and the tangent vector is unit-norm. The exponential map (3.24) spans all the range $[-\pi, \pi)$ (shown by continuous line). The phase increases linearly with μ and is periodic. The projection operator (3.27) spans the interval $[0, \pi/2)$ (shown by dashed line), and the angle has horizontal asymptote at $\pi/2$. The Cayley transform (3.30) spans the interval $[0, \pi)$ (represented by dotted line), and the angle has horizontal asymptote at π .

The corresponding update at iteration k is given by

$$\boldsymbol{\mathcal{W}}_{\text{Cayley}}(\mu) = \big[(\mathbf{I} + \frac{\mu}{2}\mathbf{H}_k)^{-1}(\mathbf{I} - \frac{\mu}{2}\mathbf{H}_k)\big]\mathbf{W}_k. \tag{3.30}$$

$$= \big[\mathbf{I} + 2\mathbf{U}_k\big(\sum_{m=1}^{\infty}(\mu\mathbf{D}_k)^m\big)\mathbf{U}_k^H\big]\mathbf{W}_k. \tag{3.31}$$

From (3.23) it follows that the matrix $\sum_{m=1}^{\infty}(\mu\mathbf{D}_k)^m$ is a diagonal matrix whose diagonal entries are of the form $\jmath\omega_i\mu/(1-\jmath\omega_i\mu)$, $i=1,\ldots,n$. Consequently, the Cayley transform has the same limitations as the projector operator. The corresponding rotation angles $\angle(1-\jmath\omega_i\mu)$ are all confined within the interval $[0,\pi)$, regardless how much μ is increased. The variation of the rotation angle on $U(1)$ w.r.t. μ is shown in Figure 3.4. In the multi-dimensional case, the Cayley transform will converge to the matrix

$$\lim_{\mu\to\infty}\big[(\mathbf{I} + \frac{\mu}{2}\mathbf{H}_k)^{-1}(\mathbf{I} - \frac{\mu}{2}\mathbf{H}_k)\big]\mathbf{W}_k = -\mathbf{W}_k. \tag{3.32}$$

Again, the behavior of the cost function along the curve described by the Cayley transform (3.30) is not almost periodic, as in the case of the exponential map. This is shown in Figure 3.3 by dotted line, taking as an example the Brockett function [41,184] (see also [Publication II] and [Publication V]).

In conclusion, the exponential map is suitable for line search methods due to its almost periodic behavior, unlike other common local parametrizations.

Practical line search methods on $U(n)$

Many of the existing geometric optimization algorithms do not include practical line search methods [69, 151], or if they do, they are too complex when applied to optimization on $U(n)$ [77, 137]. In some cases, the line search methods are either valid only for specific cost functions [183], or the resulting search is not highly accurate [77, 122, 161]. The difficulty of finding a closed-form solutions for a suitable step size is discussed in [122]. The accuracy of line search is crucial for the performance of the resulting algorithms, especially in the case of the CG algorithm which assumes exact line search.

Two efficient high-accuracy line search methods exploiting the *almost periodic* property of the cost function along geodesics on $U(n)$ are proposed in [Publication II] and [Publication IV]. The first method finds only the first local minimum (or maximum) of the cost function along a given geodesic. It is based on a low-order polynomial approximation of the first-order derivative of the cost function along geodesics and detecting its first sign change [225]. The second one finds several local minima (or maxima) of the cost function along a given geodesic and selects the best one. It approximates the almost periodic function by a periodic one [81], using the classical Discrete Fourier Transform (DFT) approach.

The almost periodic behavior of the cost function $\hat{\mathcal{J}}(\mu)$ (3.26) and its first-order derivative $d\hat{\mathcal{J}}/d\mu$ along geodesic $\boldsymbol{\mathcal{W}}_{\text{geod}}(\mu)$ (3.24) is shown in Figure 3.5.

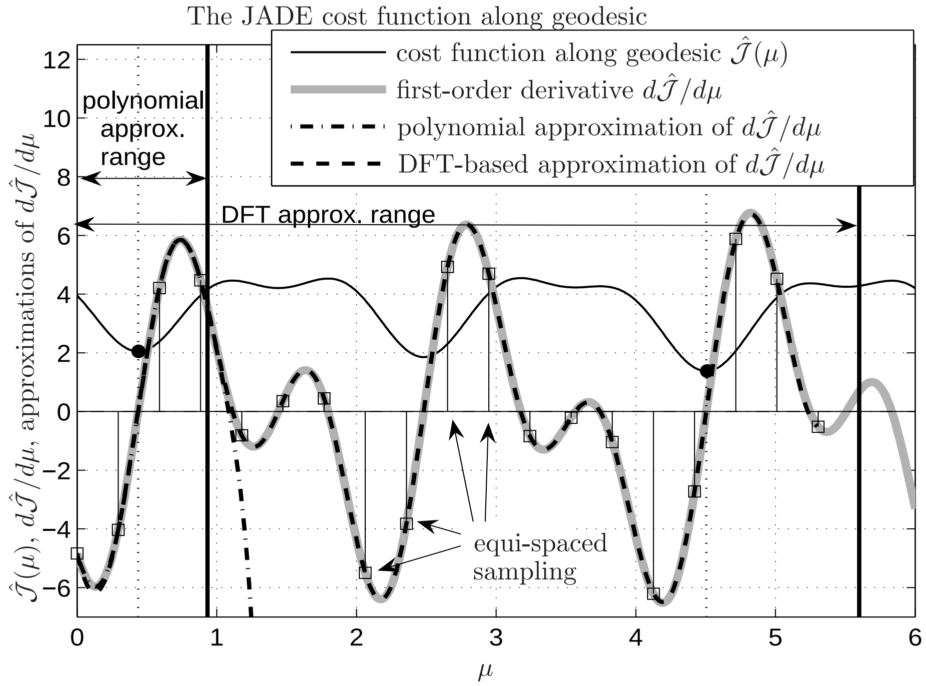

Figure 3.5: Performing the geodesic search for the JADE cost function [45]. The almost periodic behavior of the cost function $\hat{\mathcal{J}}(\mu)$ (3.26) and its first-order derivative $d\hat{\mathcal{J}}/d\mu$ along geodesic $\boldsymbol{\mathcal{W}}_{\text{geod}}(\mu)$ (3.24) may be noticed. The first proposed line search method uses a polynomial approximation of $d\hat{\mathcal{J}}/d\mu$ in order to find its smallest positive zero-crossing value which corresponds to the first local minimum of $\hat{\mathcal{J}}(\mu)$, i.e., the desired step size μ_k. The second proposed line search method uses DFT-based approximation of $d\hat{\mathcal{J}}/d\mu$ (dashed line) in order to find several local minima of $\hat{\mathcal{J}}(\mu)$ along geodesic and select the best one. Both methods sample the derivative $d\hat{\mathcal{J}}/d\mu$ at equi-spaced points in order to avoid repeated computations of the matrix exponential.

Both proposed methods find one or more zero-crossing values of the first-order derivative of the cost function $d\hat{\mathcal{J}}/d\mu$. They correspond to local minima of the cost function $\mathcal{J}(\mu)$. The zero-crossing values are related to the frequency spectrum of the first-order derivative. Due to differentiation, this spectrum corresponds to the high-pass filtered spectrum of the cost function. The approximation range is set according to the highest frequency component in the spectrum, which is related to the dominant eigenvalue of the argument of the matrix exponential. The common steps of the two line search methods in [Publication II] and [Publication IV] are summarized in Table 3.4. The main difference between the polynomial-based approach and the DFT-based approach is the choice of the approximation range t_{a} and the number of approximation points N_{a}. Other specific characteristics such as computational complexity issues are presented in detail in [Publication II] and [Publication IV]. An important common feature shared by the two methods is that the matrix exponential is evaluated at equi-spaced points (see Figure 3.5). Therefore, both proposed methods require *only one* evaluation of the matrix exponential, i.e. $\mathbf{R}_1$ in step **4** (Table 3.4). The other $N-1$ rotations are powers of $\mathbf{R}_1$.

1	Given $\mathbf{W}_k \in U(n)$, $-\mathbf{H}_k \in \mathfrak{u}(n)$, compute the eigenvalue of $\mathbf{H}_k$ of highest magnitude $\lvert\omega_{\max}\rvert$
2	Determine the order r of the cost function $\mathcal{J}(\mathbf{W})$ in the coefficients of $\mathbf{W}$, which is the highest degree that t appears in the expansion of $\mathcal{J}(\mathbf{W} + t\mathbf{Z}), \quad t \in \mathbb{R}, \mathbf{Z} \in \mathbb{C}^{n \times n}$
3	Based on $\lvert\omega_{\max}\rvert$ and r, determine an appropriate range t_{a} for approximating the first-order derivative of the cost function along geodesics, $\hat{\mathcal{J}}'(\mu)$.
4	Evaluate the rotation $\boldsymbol{\mathcal{R}}(\mu) = \exp(-\mu\mathbf{H}_k)$ at $N_{\mathrm{a}}+1$ equi-spaced points $\mu_i \in \{0, t_{\mathrm{a}}/N_{\mathrm{a}}, 2t_{\mathrm{a}}/N_{\mathrm{a}}, \ldots, t_{\mathrm{a}}\}$ as follows: $\mathbf{R}_0 \triangleq \boldsymbol{\mathcal{R}}(0) = \mathbf{I}$ $\mathbf{R}_1 \triangleq \boldsymbol{\mathcal{R}}\big(t_{\mathrm{a}}/N_{\mathrm{a}}\big) = \exp\big(-\frac{t_{\mathrm{a}}}{N_{\mathrm{a}}}\mathbf{H}_k\big)$, $\mathbf{R}_2 \triangleq \boldsymbol{\mathcal{R}}\big(2t_{\mathrm{a}}/N_{\mathrm{a}}\big) = \mathbf{R}_1\mathbf{R}_1$, $\ldots,$ $\mathbf{R}_{N_{\mathrm{a}}} \triangleq \boldsymbol{\mathcal{R}}(t_{\mathrm{a}}) = \mathbf{R}_{N_{\mathrm{a}}-1}\mathbf{R}_1$.
5	By using the computed values of $\mathbf{R}_i$, evaluate: $\hat{\mathcal{J}}'(\mu_i) \triangleq d\hat{\mathcal{J}}/d\mu = -2\Re\{\mathrm{trace}\{\frac{\partial \mathcal{J}}{\partial \boldsymbol{\mathcal{W}}^*}(\mathbf{R}_i\mathbf{W}_k)\ \mathbf{W}_k^H\mathbf{R}_i^H\mathbf{H}_k^H\}\}, \quad i = 0, \ldots, N$
6	Approximate $\hat{\mathcal{J}}'(\mu)$ by using *polynomial approximation*, or *DFT-based approximation* and find the zero-crossing values of the approximation
7	Set the step size μ_k to a root corresponding to the desired minimum (or maximum) along geodesic

Table 3.4: Geodesic search algorithm on $U(n)$.

The periodicity of functions on $SO(2)$ was discussed and exploited in the ICA context in [72]. This property was treated extensively later in [160,161] for $SO(2)$ and $SO(3)$. A one-dimensional Newton method which uses the first-order Fourier approximation the cost function along geodesics on $SO(n)$ is proposed in [161]. The main difference between the proposed DFT-based method [Publication II] and the line search method in [160,161] is that we choose to approximate the first-order derivative of the cost function along geodesics and find the corresponding zeros, instead of approximating the cost function itself and finding a local minimum as in [160,161]. Another difference is that the line search method in [160, 161] finds only one minimum. The proposed method finds multiple local minima and selects the best one. For this reason, when used with a SD algorithm the proposed DFT method leads to a performance comparable to the one of the CG algorithm, as shown in [Publication II]. The method in [160,161] exploits the *periodicity* of the cost function along geodesics which only appears on $SO(2)$ and $SO(3)$ (and it does not appear even on $U(2)$ and $U(3)$). For $n > 3$ the accuracy of the approximation decreases, since the periodicity of the cost functions is lost. Moreover, the proposed DFT-based approach uses multiple frequency components for the DFT in order to approximate the almost periodic derivative. Thus, a better spectral description of the almost periodic function is obtained. This result is demonstrated in [Publication II]. Furthermore, the proposed algorithm avoids computing second-order derivatives, unlike the method in [160, 161]. These derivatives are not always straight-forward to compute and they may be computationally expensive since large matrices (the Euclidean Hessian) may be involved.

The two proposed line search methods exploit the almost periodicity of the cost function and its derivatives in a computationally efficient manner. Other approaches are also possible. The proposed DFT-based method opens multiple possibilities for finding better local minima. One interesting approach would be to store several local minima at one iteration and use them in the next iterations. Simulated annealing approach [77] may be employed in order to reduces the dimension of the search space and at the same time avoid convergence to weak local minima. Another approach for finding better local minima (or possibly the global minimum) is to take into account more eigenvalues of the argument of the matrix exponential. The proposed approaches use only the dominant eigenvalue (no eigenvectors are required). By including several dominant eigenvalues, more precise information about the evolution of the almost periodic cost function over a wide step size range may be obtained. These aspects remain to be studied.

3.3 Discussion

In this chapter, the problem of optimization under unitary matrix constraint has been addressed. Computationally efficient SD and CG algorithms along geodesics on $U(n)$ are proposed. Two high accuracy line search methods specially tailored for the proposed algorithms are introduced. The algorithms proposed in this chapter are compared to other existing optimization algorithms (for details, see Chapter 2). Advantages and disadvantages of each algorithm are summarized in Table 3.5.

	Benefits	**Weaknesses**
Euclidean SD with enforcing constraint [113, 156, 157, 171, 172, 216]	✔easy to implement (just few equations)	✘ slow convergence ✘ computationally expensive ✘ expensive step size adaptation
Lagrange multipliers method [97]	✔closed-form solution may exist for simple cost functions and low matrix dimension n	✘ increases even more the dimension of the optimization problem (from $2n^2$ to $3n^2$) ✘ often mathematically intractable
extra-penalty method [205]	✔easy to implement (just few equations)	✘ very slow convergence ✘ low accuracy in satisfying the unitary constraint
non-geodesic SD on $U(n)$ [137, 149]	✔fast convergence (linear) ✔reduces the dimension of the optimization problem (from $2n^2$ to n^2)	✘ expensive local parametrization (projection of an *arbitrary* matrix) ✘ expensive line search (no properties to be exploited)
geodesic SD on $U(n)$ [Publication I], [Publication III], [Publication IV]	✔fast convergence (linear) ✔reduces the dimension of the optimization problem (from $2n^2$ to n^2) ✔efficient computation of the geodesics: exponential of *skew-Hermitian* matrix ✔efficient line search methods (due to the *almost periodic* behavior of the cost function along geodesics)	✘ less simple to implement (more equations are needed)
geodesic CG on $U(n)$ [Publication II], [Publication V]	✔faster convergence (superlinear) ✔reduces the dimension of the optimization problem (from $2n^2$ to n^2) ✔efficient computation of geodesics and parallel transport ✔efficient line search methods	✘ less simple to implement (more equations are needed)
geodesic and non-geodesic Newton algorithms on $U(n)$ [137, 149, 155]	✔very fast convergence (quadratic)	✘ computationally very expensive (of order $\mathcal{O}(n^6)$) ✘ may converge to undesired stationary points ✘ less simple to implement

Table 3.5: Comparison between different algorithms for optimization under unitary matrix constraint. The classical approaches operating on the Euclidean space vs. the differential geometry approaches operating on $U(n)$.

Chapter 4

Overview of blind equalization techniques for MIMO-OFDM systems

Multiple-Input Multiple-Output (MIMO) systems are a key technology of the future high-rate wireless communication systems such as 3GPP long-term evolution (LTE), IMT-2000, WiMAX and WLAN [1,4,8,10]. By using multiple transmit and received antennas, linear increase in capacity may be achieved [36, 194]. Spatial multiplexing produces parallel data streams resulting into high data rates. The transmitted streams are not necessarily orthogonal, therefore the co-channel interference problem must be considered. Moreover, the MIMO channel is selective in time, frequency and space. Space-time and space-frequency codes [22, 111, 165, 193] may be used to increase the link reliability, especially in the absence of the CSI at the transmitter [95]. They provide a good balance between the multiplexing gain and diversity gain.

When MIMO techniques are combined with spectrally efficient Orthogonal Frequency Division Multiplexing (OFDM) modulation [106, 200], the resulting MIMO-OFDM systems are very robust to multipath propagation. Typically in OFDM transmission a cyclic prefix (CP) is employed. In this way, the broadband frequency selective channel is converted into multiple orthogonal flat-fading channels. Moreover, OFDM modulation enables multiuser access schemes by allocating distinct subcarriers to different users. The resulting system is called OFDMA (Orthogonal Frequency Division Multiple Access).

Channel estimation in MIMO systems [70, 126] is a difficult problem due to the fact that the number of unknown channel parameters grows rapidly with the number of transmit/receive antennas. Consequently, pilot-aided channel estimation methods require a very large number of training data which decrease the effective data rates. Blind techniques [128] may be used to improve the effective data rates by exploiting the statistical and/or structural properties of the transmitted signals. They are very suitable in the case of continuous transmissions (e.g. DVB-T) or slowly time-varying channels (e.g. ADSL). Blind algorithms are subject to inherent ambiguities (e.g. amplitude, phase, permutation indeterminacies). A small amount of training data may be used in order to remove the ambiguities. This amount is much smaller than what is needed for the pure training-based methods. The resulting semi-blind methods may improve the convergence speed and tracking capability of the blind methods. They use both the received symbols as well as the training data. Training-based channel estimation methods need to wait until the next pilot is received. The training sequence may be distorted by the channel in a way that it is not recognized at the receiver. In conclusion, semi-blind algorithms are more feasible in practice.

The core of any semi-blind method is a blind method. In this chapter we focus on blind methods for channel estimation and equalization in MIMO-OFDM systems using P transmit antennas and Q receive antennas. Most of the methods considered here are intended for spatial multiplexing scenarios. The transmitted data streams are mutually independent and correspond to different users or multiple streams from single user. Single and multi-user SIMO cases are also considered. We classify the corresponding channel estimation and equalization algorithms into three main categories. The fist two categories include algorithms exploiting statistical properties of signals and matrices, i.e., *Second-Order Statistics (SOS)* and *Higher-Order Statistics (HOS)* approaches. The third category include deterministic algorithms that exploit structural properties of signals and matrices (or hybrid structural-statistical algorithms) .

4.1 Second-Order Statistics (SOS) based methods

In general, SOS-based blind methods exploit the correlation properties of the received signals. In general, they belong to two main classes. The first class includes SOCS (Second Order Cyclostationarity Statistics) based methods that rely on different correlation functions. The second class includes statistical subspace methods that exploit the output covariance matrix or the received data matrix.

4.1.1 SOCS based methods

Many man-made signals encountered in communications possess intrinsic periodicities caused by modulation and coding, for example. Their statistics such as mean or autocorrelation are periodic functions. Conventional WSS (wide-sense stationary) models do not take into account the valuable information contained in this periodicity. The property is called cyclostationarity and may be exploited in blind algorithms. Typically, the symbol-rate received signals are WSS, but by taking multiple samples within the symbol interval, additional information is obtained. This can be done either by oversampling in time domain, or in spatial domain by employing several symbol-rate receivers (antenna array). Cyclostationarity may be induced at the sampling rate also by shaping the signal statistics at the transmitter, but this may require redundancy.

OFDM signals possess special correlation properties. The cyclic prefix (CP), or zero-padding (ZP) induce correlation in the transmitted signals which may be exploited. Blind channel estimation methods for MIMO-OFDM systems based on the SOCS induced by the CP have been proposed in the literature [29, 67, 68, 134]. Channel correlation properties due to the Fourier transform have also been exploited in [90]. Special correlation properties may be introduced also by precoding or space-time coding [37, 221, 222]. In addition, special signal structure helps in solving the ambiguities inherent to all blind methods. In [37], a scheme for resolving the multi-dimensional ambiguity up to a diagonal complex matrix is presented. The SOCS methods in [29, 37] are immune to the common channel zeros.

4.1.2 Statistical subspace methods

Another alternative to induce structure in the transmitted signals is by inserting zero guard bands at the end of each OFDM block i.e, the so-called zero-padding (ZP) [226, 228]. Virtual subcarriers (VSC) [28, 180] may also be used, which are unmodulated subcarriers at known frequencies in the spectrum, usually in the roll-off region. Statistical subspace-methods for blind channel estimation in MIMO-OFDM have also been proposed in literature [28, 84, 91–94, 129, 180, 226, 228]. They rely on a low-rank model where the signal subspace is associated with the range space of the channel matrix. The signal and noise subspaces are obtained either via eigendecomposition of the sample estimate of the covariance matrix, or via singular value decomposition of the received data matrix [110]. Receive diversity plays an important role in building the low-rank model. Blind identification algorithms for SIMO-OFDM have been considered in [23, 48, 96]. The method in [23] is sensitive to the channel common zeros. Unlike most of the subspace methods, in [91,92,96] the model uses the covariance matrix computed in frequency domain. The advantage of this approach is the resilience to common channel zeros. Identifiability conditions for the subspace methods applied to OFDM have been formulated in [180, 227]. Statistical subspace methods are applied to OFDM signals with CP [84, 93, 94], or ZP [226, 228] in MIMO scenarios. It has been shown in [93, 94] that algorithms designed for ZP-OFDM transmission [226, 228] may be adapted to CP-OFDM transmission just by rewriting of the MIMO system model appropriately. In this way, they become compatible with most of the existing OFDM standards [2–7, 9] which use CP instead of ZP. In [67, 68] a subspace method exploiting the

CP is proposed to initialize an iterative CMA. Subspace-based methods exploiting virtual subcarriers (VSC) have been proposed in [23, 28, 180]. They do not need CP, as long as VSC are used. Space-time codes have been used in conjunction with subspace methods in [222, 226].

In general, subspace methods are able to estimate the MIMO channel up to a full-rank complex $P \times P$ ambiguity matrix [84, 91, 92, 226, 228]. The ambiguity is removed by exploiting the signal structure induced at the transmitter via precoding or space time-coding, or by using a small amount of training data, i.e., P symbols within one OFDM block of length N. Due to the fact that in practice $P \ll N$, semi-blind subspace methods still achieve increased effective data rates. An extensive review on semi-blind channel estimation methods for MIMO-OFDM systems is provided in [168]. An efficient HOS-based blind approach for solving ambiguities remaining after the blind subspace identification was proposed in [202]. By exploiting the independence between in-phase and in-quadrature components of the complex-valued signal, the algorithm is able to reduce the remaining full-rank ambiguity matrix to a diagonal matrix. Its diagonal elements correspond to complex scalar gains multiplying each of the (possibly permuted) data streams.

Most of the subspace methods involve eigendecomposition of SVD operations. In general, such matrix decompositions are computationally expensive, especially when the matrix dimensions are large [99]. Complexity reduction for the subspace methods may be often achieved. The subspace algorithm in [228] which requires both eigendecomposition and SVD has been reconsidered in [Publication III]. The SVD operation of a large tall matrix was replaced by an eigendecomposition of a small square matrix. The eigendecomposition is obtained iteratively by using Riemannian optimization technique (steepest descent on the unitary group $U(n)$). A more efficient solution would be optimizing the Brockett function [41, 184] by using the Riemannian conjugate gradient algorithm as in [Publication II].

4.2 Higher-Order Statistics (HOS) based methods

Typically, HOS-based blind methods need a larger sample support compared to the SOS-based blind methods because HOS-based estimators have higher variances. On the other hand, HOS methods may potentially have increased noise immunity, since the higher order statistics of the Gaussian noise vanish.

4.2.1 BSS methods

Most of the HOS-based blind channel estimation methods proposed in the literature [55, 102, 115, 173–177] use BSS (Blind Source Separation) principles [71] in order to separate the transmitted signals. Commonly used

Independent Component Analysis (ICA) methods [113] such as natural gradient [25] and JADE [45] are applied. They rely on the mutual statistical independence and the non-Gaussianity properties of the transmitted signals. Therefore, due to the IDFT operation at the transmitter side, they operate only in frequency domain. Assuming CP in the OFDM transmission, the MIMO channel is regarded as a set of per-tone instantaneous mixing matrices. The size of each mixing matrix is identical to the dimensionality of the MIMO channel, i.e., $P \times Q$. Consequently, the number of mixing matrices is equal to the number of tones, which leads to computationally expensive algorithms. In order to avoid this inconvenience, algorithms which exploit the correlation among subcarriers in frequency domain have been proposed in [175,177]. Correlation across subcarriers depends on the coherence bandwidth and inter-carrier spacing. These figures are directly related to channel delay spread and number of subcarriers. The channel length is considerably smaller than the IDFT length, and the channel frequency response is a result of the IDFT of a zero-padded channel impulse response [169]. Consequently, high correlation among channel coefficients corresponding to adjacent subcarriers is introduced. In conclusion, it is sufficient to obtain channel frequency response by using ICA on a number of frequency bins equal to the maximum channel length (CP length may be used as an upper bound). For the remaining tones the channel response is obtained by interpolation.

Other algorithms [115,174] apply the JADE only one one reference subcarrier, and the others subcarriers are unmixed iteratively by using a linear MMSE receiver. This may lead to error propagation across subcarriers. Additional successive interference cancellation (SIC) technique may be involved to improve the performance [174]. The SIC approach has also been used in [175] combined with a layered space-time architecture (V-BLAST) [210]. Blind source separation approach based on a natural gradient learning algorithm is developed in [102]. The algorithm in [55] uses ICA and fractional sampling. Consequently, the noise resilience is improved (more decision variables are available). On the other hand, the complexity is increased even more in comparison to the other ICA-based methods, by a factor equal to the inverse of the oversampling rate.

The major problem of BSS methods for MIMO-OFDM systems considered above is that they operate in frequency domain. This leads to several drawbacks. First, the computational complexity increases linearly with the number of subcarriers. The complexity reduction techniques become compulsory even for relatively small number of subcarriers. Second, the ambiguity problem is very hard to resolve. On each subcarrier, there is an unknown complex scalar which multiplies the channel frequency response. For this reason, convolutional coding [115,173] may be required in order to introduce redundancy, thus decreasing the effective data rate. Non-redundant linear precoding of each transmitted data stream has been also proposed [174,175]. This techniques introduce known correlation structure between subcarriers which may be exploited at the receiver. Special constellation properties have also been considered [55].

Most of the blind algorithms considered in this section are based on minimizing the JADE criterion [45]. Complexity reduction for JADE may be achieved by using Riemannian conjugate gradient as in [Publication II]. This problem has also been addressed in [Publication I], [Publication IV] and [Publication V]. The reduction in complexity is considerable if multiple JADE algorithms are employed in parallel, such as the blind algorithms considered in this section, which operate on a subcarrier basis. Moreover, when the dimensions of the mixing matrix are large, the pairwise processing used in the diagonalization stage of JADE leads to slow convergence [Publication II], [Publication IV], [Publication V]. Same problem occurs when the input data streams have different distributions, i.e., they belong to different constellations [Publication II].

4.3 Structural properties based methods

Typically, statistical blind methods require a large sample size in order to provide unbiased channel estimates. Apart form statistical properties, signals and matrices may posses special structural properties that may be exploited in blind algorithms. These properties arise either from the modulation scheme, or from different matrix structures employed in the description of system model. These special properties appear even for very small sample size. For this reason, in some cases (e.g. noise-free scenarios or constant modulus constellations) the channel estimation can be achieved even from a single received data block, in a deterministic manner. In general, deterministic methods outperform statistical methods for small number of received data blocks. By using more received data blocks, the robustness to noise and other imperfections is increased. In this way, some deterministic methods are converted into hybrid structural-statistical methods. Such hybrid methods require fewer observations compared to the pure statistical methods. Reducing the size of data records enables good channel tracking capabilities.

OFDM modulation and MIMO channel determine special signal and/or matrix structures that may be useful in blind channel estimation. In this subsection, the structural blind methods for multiple-antenna OFDM systems are classified into three main categories. The first one exploits the structure of the transmitted signals arising from the known constellation modulation scheme [30, 67, 68, 114, 126, 159, 190, 232]. The second category includes algorithms that use the OFDM guard bands [31, 49, 119]. The third category includes blind algorithms exploiting special matrix structures arising from the data model [49, 100, 126, 166, 204].

4.3.1 Modulation properties

Blind algorithms that rely on the properties of the modulation scheme have been proposed in the literature [114, 159, 232]. The main properties that are exploited are the finite-alphabet, constant envelope, or constant block energy of the transmitted signals. In addition to these properties, some of the methods exploit the receive diversity which is achieved by oversampling in time or space.

Finite-alphabet methods

Digitally modulated communications signals have a finite-alphabet (FA) structure, i.e., the transmitted symbols belong to a finite set of amplitudes and phases. Blind methods using the FA property match the received signals to the unknown channel taps and projects the soft-estimated symbols onto the constellation set. Least squares methods that estimate the channel coefficients and the transmitted symbols alternately have been proposed [126]. The FA property was first applied to OFDM in [232] in single-antenna case. The corresponding deterministic blind algorithm is able to identify the channel from a single OFDM block when PSK constellations are used at high SNR conditions. The remaining phase ambiguity may be easily resolved, since it belongs to a finite set. The algorithm is computationally expensive due to the fact that it operates on each tone. It requires an exhaustive search on a space which grows exponentially with the number of active subcarriers. A sub-optimal version of the FA method is proposed in [232], but the dimension of the search space is still exponential in the number of channel taps. An improved version was proposed in [159]. The method dramatically reduces the computational complexity because it operates on clusters of subcarriers (withing the same coherence bandwidth). For this reason, the method is sensitive to the choice of the clusters. In order to overcome this difficulty, turbo-decoding was used in conjunction with the FA property in [159]. The drawback of the method is that channel coding requires redundancy. Other sub-optimal approaches for reducing the complexity of the FA method were proposed in [114]. Receiver diversity may also be employed in order to improve the performance. A deterministic ML (maximum-likelihood) blind method for SIMO-OFDM exploiting the FA property was considered in [30]. The method exploits the receive diversity and in the absence of noise can achieve perfect channel estimation by using single received OFDM block. The algorithm exhibits very high complexity. The ML method can be decoupled into two separate LS problems, involving the channel coefficients and the transmitted symbols, respectively. Exhaustive search over a high-dimensional space is still needed for the symbol estimation part. A comprehensive review of different FA methods until year 2000 may be found in [126, Sec. 4.3]. For QAM constellations, FA methods suffer from error floor effect and high variance of the estimates. Moreover, for higher-order constellations they require HOS of the received signals, and therefore longer data records.

Constant modulus property

An iterative algorithm exploiting the constant envelope characteristic of the transmitted data symbols was proposed for SIMO-OFDM in [190]. The algorithm is based on least squares CMA (Constant Modulus Algorithm) that takes additional benefit from the receive diversity. In [67, 68] constant modulus algorithm was employed for channel equalization on MIMO-OFDM systems. The initialization is made by using the estimates provided by a subspace method. After initialization the CMA works in an adaptive fashion. The constant modulus property has also been used to resolve the multiple scaling ambiguities of blind algorithms operating in frequency domain [102].

Constant mean-block energy property

Iterative methods minimizing different criteria have been proposed for OFDM. A blind equalizer based on restoring the constant mean block energy property [218] of the received OFDM data blocks was proposed for single-antenna case in [119]. The algorithm is called VCMA (Vector Constant Modulus Algorithm). In addition, the structure of the CP and ZP guard bands are exploited. A blind equalizer for MIMO OFDM systems using VCMA and decorrelation criteria was introduced in [Publication VI]. The algorithm has been modified in order to take into account the correlation introduced when using CP in [13]. A block-Toeplitz structure of the equalizer is enforced by averaging along diagonals. Other approaches enforcing the Toeplitz structure are considered in [126]. The VCMA algorithm proposed in [Publication VI] and its improved final version in [Publication VII] will be discussed in detail in Chapter 5. The VCMA-based algorithms can also be included in the class of HOS methods, since the corresponding criteria use fourth order moments. We have included them in the class of structural methods because they attempt to restore the constant mean block energy property and do not use sample averaging, which may decrease the convergence speed. In practice, they are implemented in an adaptive fashion and the expected values are replaced by instantaneous estimates.

4.3.2 Properties of the guard interval of OFDM signal (CP or ZP)

This type of methods rely on the guard intervals used in the OFDM transmission such as CP or ZP [119]. In addition, they may exploit the receive diversity. A blind beamformer for SIMO-OFDM systems exploiting receive antennas diversity and the temporal redundancy induced by the cyclic prefix was proposed in [31]. A criterion which penalizes the MSE between the CP samples and the corresponding data samples within the received OFDM block is employed. In [49], a deterministic LS blind approach using received diversity and CP similarity is proposed. The algorithm is able to estimate the SIMO channel by using a single received OFDM block. The channel response and the array response are incorporated into a global mixing matrix, whose dimensions depend on the channel order. Therefore, the channel order must be known precisely. It is estimated by using SVD of the data matrix. The channel is found by using the SVD of a difference matrix which penalizes the difference between the CP samples and the corresponding data samples. The ambiguity is resolved also based on the CP redundancy.

4.3.3 Exploiting special matrix structures

Special matrix structures arising from the SIMO and MIMO models may be exploited in blind methods [199].

Deterministic subspace methods

The full column rank property of the channel matrix is a prerequisite for subspace-based system identification. Therefore, the model must involve a tall channel matrix (more rows than columns). This is just a necessary condition, but not a sufficient one. If the full column rank condition is not met, several received data blocks may be stacked in the top of each other. Received diversity is also a mean to build a low-rank model. In spatial multiplexing scenarios, estimating the desired base of the subspaces is not sufficient for separating the transmitted data streams. Thus, a second constraint must be employed. Finite-alphabet, constant-modulus or specific Hankel and Toeplitz matrices may be used in conjunction with the subspace methods [126]. Enforcing special matrix structures enables satisfying the second constraint needed for the estimated subspace. In this way, by properly combining the estimated basis vectors the co-channel signal cancellation may be achieved. Cyclic prefix or zero-padding may also be exploited [49]. In some cases [49, 100, 204] the projection matrix to the noise subspace is deterministic and known at the receiver, and the channel estimation may be accomplished after the first received OFDM block.

In [204], a blind SIMO channel identification algorithm based on received diversity was proposed. Special properties arising from the unitary FFT matrix used in the OFDM transmission was exploited. The SIMO channel estimation problem reduces to the usual problem of finding the minimal eigenvector of a Hermitian matrix (or minimal singular vector of a tall matrix). The algorithm may be sensitive to the common zeros on the subchannels. Moreover, in [204] the channel length is assumed to be known *a priori.*

Special structure in the signals may be induced at transmitter by using a redundant precoding scheme. In [100], a deterministic blind equalization algorithm for SIMO-OFDM is proposed. The algorithm exploits the receive antenna diversity and the structure imposed by frequency domain spreading. Therefore, the transmission system can be viewed as a multicarrier CDMA system. In the absence of noise, the equalization can be achieved in a single OFDM block. A regularization approach is used in order to cope with the problem of common zeros. The redundancy introduced by spreading improves the system reliability. It is shown that the proposed redundant scheme outperforms the uncoded scheme. Same data rate is achieved at given SNR with lower bit error rate.

Other algebraic techniques

A deterministic blind channel identification method for MIMO-OFDM has been recently proposed in [166]. It is based on an algebraic technique which decomposes the received signal vector in a four-way tensor whose dimensions are space, time and frequency. The method exhibits high complexity, due to the fact that it uses large multi-dimensional matrices. In addition, it is difficult to prove if the identifiability conditions are met in practice.

4.4 Discussion

In this chapter, the most relevant blind channel estimation and equalization methods for multi-antenna OFDM systems have been reviewed. The main focus was on blind methods applicable to SIMO and MIMO systems. The methods were classified in three main classes. The first two include statistical methods (SOS and HOS) and the third one includes methods exploiting structural properties of signals and matrices. In Table 4.1, a comparison of the three classes of methods is given. The pros and cons of each methods are considered, as well as the computational complexity. Both statistical and structural methods possess benefits and drawbacks. HOS-based methods require large sample support in order to provide unbiased estimates. In general, SOS-based methods outperform the HOS-based methods for the same amount of received data. Subspace-based methods require computationally expensive matrix decompositions. This is valid also for some of the structural methods. Moreover, some structural methods such as FA are just for theoretical study, since their computational complexity is unaffordable in practice. Finally, most of the methods require precoding or shaping the space-time signals in order to cope with the channel common zeros. In conclusion, exploiting the structural properties of matrices and signals, and at the same time taking advantage of their statistics may result into fast and computationally efficient algorithms.

	Benefits	Weaknesses
SOS methods: SOCS [29, 37, 67, 68, 134, 221, 222]	✔require less received data compared to HOS methods ✔computationally simple	✘ may require additional precoding or oversampling ✘ may be unable to identify nonminimum-phase channels ✘ sensitive to common zeros unless precoding is involved ✘ require CP/ZP
subspace based [23, 28, 84, 91–94, 96, 129, 180, 226, 228]	✔require less received data compared to HOS methods ✔frequency domain models are immune to common zeros	✘ require expensive matrix decompositions ✘ frequency domain models are computationally complex ✘ require CP/ZP and/or VSC
HOS methods: BSS [55, 102, 115, 173–177]	✔robust to Gaussian noise ✔robust to channel zeros	✘ high variance ✘ require more received data compared to SOS methods ✘ high complexity for large number of subcarriers ✘ ambiguities are harder to resolve ✘ require CP/ZP
Structural methods: Modulation properties (FA, CMA, VCMA) [30, 67, 68, 102, 190]	✔FA may achieve estimation in single OFDM block (high SNR and CM constellations) ✔CMA and VCMA are adaptive ✔for VCMA, the complexity does not grow with the number of subcarriers ✔for VCMA, ambiguities are easier to resolve	✘ FA methods are extremely complex ✘ for FA and CMA, ambiguities are harder to resolve ✘ FA and CMA exhibit high complexity for large number of subcarriers ✘ FA, CMA require CP/ZP ✘ VCMA is sensitive to common zeros ✘ CMA and VCMA may converge only locally
CP/ZP structure [31, 49, 119]	✔estimation may be achieved in single OFDM block (high SNR)	✘ ambiguities are harder to resolve ✘ require CP/ZP
Special matrix structures [49, 100, 166, 204]	✔estimation may be achieved in single OFDM block (high SNR)	✘ ambiguities are harder to resolve ✘ require expensive matrix decompositions ✘ require CP or precoding

Table 4.1: Different algorithms for blind channel estimation and equalization for SIMO/MIMO-OFDM systems. The statistical-based methods (SOS, HOS), and methods exploiting the structural properties.

Chapter 5

Blind equalizer for MIMO-OFDM systems based on vector CMA and decorrelation criteria

In this chapter, the problem of blind recovery of multiple OFDM data streams in a MIMO system is addressed. We propose an equalization algorithm for MIMO-OFDM receivers which optimizes a *composite criterion* in order to cancel both the ISI and CCI. ISI is minimized by using a *modified Vector Constant Modulus criterion* while CCI is minimized by using a *decorrelation criterion*. The composite criterion was introduced in [Publication VI]. The algorithms was subsequently improved in [12, 13] and [Publication VII]. The convergence properties of the algorithm have also been established. Conditions for the existence of the stable minima corresponding to the zero forcing receiver which performs the joint blind equalization and the co-channel signal cancellation are established in [Publication VII].

The proposed blind algorithm operates in the time domain before the DFT operation at the receiver. Therefore, it is designed to deal with three different cases: there is no CP at all, the CP is too short (compared to the channel impulse response) and CP is sufficiently long. The CP may be used for synchronization purposes, hence it is included in the algorithm derivation. However, it is not needed in finding the equalizer. The proposed blind algorithm exploits the *mutual statistical independence* among the transmitted data streams and the *the constant mean block energy* property of the OFDM signals. Hence it is applicable to spatial multiplexing systems. The VCMA criterion [217] penalizes the deviation of the block energy from a dispersion constant. The VCMA cost function may be decomposed into a constant modulus (CM) cost function [98] and an auto-correlation function of the squared magnitudes of the received signal [195]. Therefore, the original VCMA is not suitable for signals which have a periodic correlation such as OFDM signal. When CP is used, a strong auto-correlation in the transmitted signals is introduced. It may be stronger than the correlation caused by the multipath propagation channel. Consequently, the performance of the original VCMA degrades then significantly because it penalizes the correlation induced by the CP [12, 13]. The *modified VCMA* proposed in [Publication VII] is designed to deal with the auto-correlation caused by the CP and to cancel ISI simultaneously.

The VCMA was applied to blind equalization for shaped constellations in [217] and for Single-Input Single-Output (SISO) OFDM system in [119]. In [119] CP or ZP were required in order to perform the equalization. MIMO schemes have been considered in [132, 158] using the classical CMA, for BPSK signals. VCMA was employed in the context of DS-CDMA systems in [206]. In a spatial multiplexing MIMO scenario, the problem becomes more difficult. At one receive antenna we have the desired signal with its delayed replicas caused by the channel ISI in addition to the co-channel signals, i.e., CCI with their delayed replicas. In order to perform both the blind equalization and signal separation, an output decorrelation criterion is needed. This criterion assumes that the transmitted data streams are mutually independent. Hence, it is suitable for spatial multiplexing systems. It penalizes the correlation among the equalized outputs. Consequently, we come up with a cost function comprised of two criteria: a modified VCMA criterion and a decorrelation criterion.

The proposed algorithm is presented in detail in this chapter. First, the system model is given in Section 5.1. The blind equalizer is presented in Section 5.2.

5.1 System model for spatial multiplexing MIMO-OFDM system

We consider a MIMO-OFDM system with P-transmit and Q-receive antennas (Figure 5.1). We assume a spatial multiplexing scenario, where independent OFDM data streams are launched from each antennas. Each data stream consists of i.i.d. complex symbols modulated by M subcarriers. Multi-user SIMO systems have similar model. In this model we use a block formulation similar to the one in [207]. The sample index is denoted by $(\cdot)$, and the block index by $[\cdot]$. Consider the complex symbols from the pth data stream stacked in a $M \times 1$ vector $\mathbf{s}_p[k] = [s_p(kM), \ldots, s_p(kM - M + 1)]^T$. The $N \times 1$ transmitted OFDM block of the pth data stream can be written as:

$$\tilde{\mathbf{u}}_p[k] = \mathbf{T}_{\mathrm{CP}} \mathbf{F} \mathbf{s}_p[k], \tag{5.1}$$

where $\mathbf{F}$ is the $M \times M$ normalized IDFT matrix and $\mathbf{T}_{\mathrm{CP}}$ is the $N \times M$ cyclic prefix addition matrix. The sequence of $L + 1$ consecutive transmit-

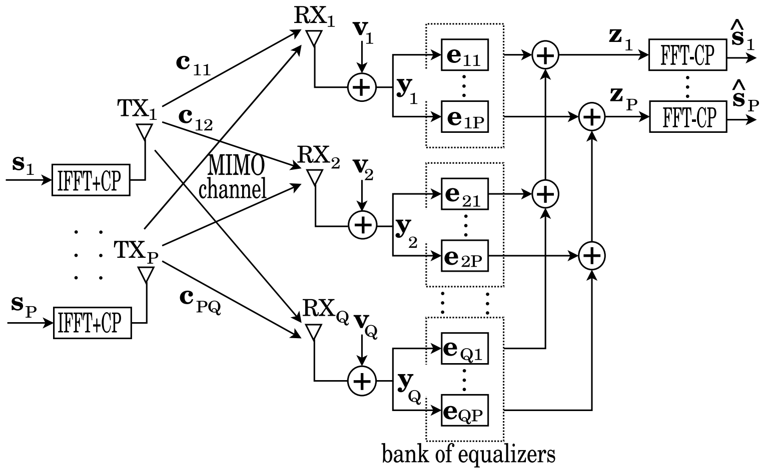

Figure 5.1: MIMO-OFDM system model

ted OFDM samples corresponding to the antenna p is denoted by $\mathbf{u}_p(k) = [u_p(k), \ldots, u_p(k-L)]^T$. The MIMO channel branches from the pth transmit to the qth receive antenna (see Figure 5.1) have maximum order L_c and they are characterized by the impulse responses $\mathbf{c}_{pq} = [c_{pq}(0), \ldots, c_{pq}(L_\mathrm{c})]$. Stacking the vectors $\mathbf{u}_p(k)$ corresponding to the P transmitted data streams in a vector $\mathbf{u}(k) = [\mathbf{u}_1^T(k), \ldots, \mathbf{u}_P^T(k)]^T$, the $L - L_\mathrm{c} + 1$ consecutive samples received at the antenna q, are:

$$\mathbf{y}_q(k) = [\mathbf{C}_{1q} \ldots \mathbf{C}_{Pq}]\mathbf{u}(k) + \mathbf{v}_q(k), \quad q = 1, \ldots, Q, \tag{5.2}$$

where $\mathbf{C}_{pq}$ are $(L-L_\mathrm{c}+1)\times(L+1)$ Sylvester convolution matrices containing the channel coefficients $\mathbf{c}_{pq}$, and $\mathbf{v}_q(k)$ is the additive white Gaussian noise at the qth receive antenna. Consider the MIMO channel matrix $\bar{\mathbf{C}}$ whose blocks (p, q) are the matrices $\mathbf{C}_{pq}$, with $p = 1, \ldots, P$ and $q = 1, \ldots, Q$. The $Q(L - L_\mathrm{c} + 1) \times 1$ array output $\mathbf{y}(k) = [\mathbf{y}_1^T(k), \ldots, \mathbf{y}_Q^T(k)]^T$ may be written as:

$$\mathbf{y}(k) = \bar{\mathbf{C}}\mathbf{u}(k) + \mathbf{v}(k), \tag{5.3}$$

where $\mathbf{v}(k) = [\mathbf{v}_1^T(k), \ldots, \mathbf{v}_Q^T(k)]^T$. The adaptive equalizers have order L_e and they are row vectors denoted by $\mathbf{e}_{qp}[k] = [e_{qp}(0), \ldots, e_{qp}(L_\mathrm{e})]$. The minimum equalizer order is chosen according to the identifiability conditions presented in [Publication VII]. In order to recover the P transmitted data streams, a bank of P equalizers

$$\mathbf{e}_p[k] = \big[\mathbf{e}_{1p}[k], \ldots, \mathbf{e}_{Qp}[k]\big],\ p = 1, \ldots, P \tag{5.4}$$

is used at each receive antenna (see Figure 5.1). By choosing $L = L_\mathrm{c} + L_\mathrm{e}$, the equalized sample corresponding to the pth data stream can be written as:

$$z_p(k) = \mathbf{e}_p[k]\mathbf{y}(k). \tag{5.5}$$

Considering the $1 \times (L+1)$ global channel-equalizer impulse response (GCEIR) $\mathbf{a}_p = \mathbf{e}_p[k]\bar{\mathbf{C}}$ corresponding to the pth data stream, the equalized sample from this data stream may be written as:

$$z_p(k) = \mathbf{a}_p\mathbf{u}(k). \tag{5.6}$$

If the equalization is achieved, the GCEIRs are equal to the standard unit vector multiplied by an unknown phase rotation, i.e., $\mathbf{a}_p = \boldsymbol{\delta}_i e^{\jmath\theta}$, where $p = 1, \ldots, P$, $i \in \{0, 1, \ldots, L\}$ and $\theta \in [-\pi, \pi)$. A permutation of the equalized data streams may also be encountered. The phase and the permutation ambiguities are inherent to all blind methods. The adaptive equalizer corresponding to each recovered data stream operates in a block mode, and it outputs a block of k samples $\mathbf{z}_p[k] = [z_p(kN), \ldots, z_p(kN-N+1)]^T$. The equalized block corresponding to the pth data stream may be written as:

$$\mathbf{z}_p[k] = \sum_{q=1}^{Q} \mathbf{E}_{qp}[k]\tilde{\mathbf{y}}_p[k], \tag{5.7}$$

where $\tilde{\mathbf{y}}_p[k] = [y_p(kN), \ldots, y_p(kN-N-L_\mathrm{e}+1)]^T$ and $\mathbf{E}_{qp}[k]$ are $N\times(N+L_\mathrm{e})$ Sylvester convolution matrices built with the coefficients $\mathbf{e}_{qp}[k]$.

5.2 Blind MIMO-OFDM equalizer

The proposed blind algorithm performs the equalization and the co-channel interference cancellation. It minimizes a composite cost function comprised of two criteria: *a modified VCMA criterion* and *a decorrelation criterion.* These two criteria are described next.

5.2.1 Modified VCMA Criterion

In single transmitter case [217] VCMA criterion penalizes the deviation of the equalized block energy from a given dispersion constant. In the multiple transmitter scenario considered in this work, the energy penalty over all data streams may be written as:

$$\mathcal{J}^{\mathrm{VCMA}}(\mathbf{e}_1[k], \ldots, \mathbf{e}_P[k]) = \sum_{p=1}^{P} E\Big[(\|\mathbf{z}_p[k]\|^2 - r_2)^2\Big], \tag{5.8}$$

where $\|\cdot\|$ denotes the l_2-norm of a vector. The block energy dispersion constant is $r_2 = E[\|\tilde{\mathbf{u}}_p[k]\|^4]/E[\|\tilde{\mathbf{u}}_p[k]\|^2]$. The original VCMA cost function [217] is not applicable to signals which have a periodic correlation such as

OFDM signal using cyclic prefix (CP). This is due to the fact that the proposed criterion penalizes the both the autocorrelation and the cross-correlations of the transmitted data streams. CP introduces autocorrelation which may be stronger than the inter-symbol interference (ISI) caused by the multipath channel. The proposed *modified VCMA* [Publication VII] can handle both the auto-correlation caused by the CP and the channel ISI simultaneously. For details, see [12, 13].

5.2.2 Output Decorrelation Criterion

The pth equalized data stream $\mathbf{z}_p[k]$ may contain interfering signals corresponding to the other data streams $\mathbf{z}_l[k]$, as well as their delayed replicas, $\mathbf{z}_l[n,d] = [z_l(kN-d),\ldots,z_l(kN-d-N+1)]^T$. The interference is measured by the cross-correlation matrix $\mathbf{R}_{pl}(d)$, between pth and lth equalized outputs for a certain delay d, i.e, $\mathbf{R}_{pl}(d) = E\big[\mathbf{z}_p[k]\mathbf{z}_l^H[n,d]\big]$. A decorrelation criterion must be employed because multiple copies of other signals may be present in the desired signal, i.e., the CCI. This criterion minimizes the squared Frobenius norm of the cross-correlation matrices. The cross-correlation cost function over all equalized data streams is:

$$\mathcal{J}^{\text{xcorr}}(\mathbf{e}_1[k],\ldots,\mathbf{e}_P[k]) = \sum_{\substack{p,l=1\\p\neq l}}^{P}\sum_{d=d_1}^{d_2}\big\|\mathbf{R}_{pl}(d)\big\|_F^2. \tag{5.9}$$

The delays d_1, d_2 are chosen according to the maximum delay introduced by the channel. The integer d spans the window of all possible delays, in order to mitigate all the delayed replicas of the interference signals.

5.2.3 Composite Criterion

The VCMA cost function (5.8) has originally been designed for single transmitter case [217]. Its global convergence has not been established, not even in the single transmitter case [104, 195]. If multiple signals are present, depending on its initialization, VCMA may converge to any of the transmitted signals, usually to the ones that have the strongest power [217]. This is due to the fact that VCMA updates the equalizers corresponding the the P data streams independently, i.e., the equalized outputs do not influence each other. Obviously, VCMA alone is not sufficient for equalization in a spatial multiplexing scenario, since the problem of co-channel signals must be considered as well [158]. We propose a composite criterion which we prove to be locally convergent. The cost functions (5.8) and (5.9) may be combined in order to cancel both ISI and CCI. A weighting parameter $0 < \lambda < 1$ is used to weight the two criteria. The composite cost function is given by:

$$\mathcal{J} = \lambda\mathcal{J}^{\text{VCMA}} + (1-\lambda)\mathcal{J}^{\text{xcorr}}. \tag{5.10}$$

The composite criterion (5.10) needs to be minimized w.r.t. to the equalizer coefficients. In addition, the unknown parameter λ needs to be found. This is a challenging optimization problem since the function to be minimized is a multivariate function of fourth-order in its complex-valued arguments. The method of Lagrange multipliers [97] is the first method one would have in mind for this type of constraint minimization problem. In that case, the unknown parameters are the equalizer coefficients and the Lagrange multipliers. The equalized outputs depend on both channel and equalizer coefficients. Consequently, the corresponding Lagrangian function includes the unknown channel impulse responses and the Lagrangian method cannot be applied. Moreover, solving the corresponding system of equations would be difficult even with known channel. So happens for the dual optimization approach [32, 38]. In addition, due to the fact that the criterion is non-convex, the dual approach does not provide an optimal solution, i.e., there is the so-called *duality gap* [32, 38]. Our derivation in [Publication VII] provides an optimal solution to the minimization problem. In order to get rid of the unknown channel impulses response, the cost function is analyzed in the space of the global channel-equalizer responses. In this way, the optimum weighting parameter λ may be found. By using the obtained value, the equalizer coefficients are updated by using a stochastic steepest descent method

$$\mathbf{e}_p[k+1] = \mathbf{e}_p[k] - \mu \nabla^{\mathrm{E}}_{\mathbf{e}_p} \mathcal{J}, \tag{5.11}$$

where $\nabla^{\mathrm{E}}_{\mathbf{e}_p} \mathcal{J}$ is the instantaneous Euclidean gradient (3.10) of the composite cost function $\mathcal{J}$ (5.10), w.r.t. the pth equalizer coefficients $\mathbf{e}_p$ (5.4) at iteration k. The gradient expression is given in [Publication VII] and it includes instantaneous estimates instead of expected values. A conjugate gradient algorithm could also be used instead, but this would require performing an exact line search. An accurate adaptive step size would be required, but this is difficult due to the due to stochastic nature of the algorithm. A closed-form expression for the optimum weighting parameter λ and an upper bound for the step size μ are provided in [Publication VII]. It has been demonstrated in [Publication VII] that stable zero-forcing solutions always exist if the value of the parameter λ is set appropriately. Local convergence properties of the algorithm have also been established in [Publication VII].

5.2.4 Conditions for symbol recovery

The proposed algorithm guarantees both signal equalization and co-channel interference cancellation under the conditions below. First, the existence of the zero-forcing (ZF) equalizer must to be established. Second, conditions for the convergence of the equalizer needs to be found.

Existence of the blind equalizer

Two conditions are necessary for the existence of the ZF equalizer [58]. The first condition is that the virtual polynomials associated to the MIMO channel have no common zeros. This is ensured by a sufficient antenna element spacing w.r.t. coherence distance or by coding. The second necessary condition for identifiability is the minimum equalizer length which depends on the channel maximum order and the number of transmit/receive antennas. For details, see [Publication VII].

Convergence properties

The global convergence properties of the original VCMA [217] have not been established so far, not even in the fractionally-spaced case [104]. The local convergence of VCMA has been investigated in [195]. In [Publication VII] we prove that the proposed algorithm which is based *a composite cost function* convergences at least locally. The differences compared to the pure VCMA [104, 217] are explained in detail in [Publication VII]. Once the existence of the ZF equalizer is guaranteed, the convergence of the blind equalizer to the ZF solution depends upon the characteristics of the surface of the composite criterion represented in the space of the global channel-equalizer impulse responses. We show that truly stable local minima of the composite criterion which correspond to the zero-forcing solutions always exist under the assumption that the parameter λ weighting the two criteria and the step size μ are appropriately selected. It has been demonstrated in [Publication VII] that in the absence of noise, any value $0 < \lambda < 1$ is appropriate. In noisy conditions, values of λ which are close to zero and one should be avoided. A closed-form expression for optimal parameter λ and an upper bound on the step size parameter μ are also provided and they depend only on the system parameters.

It has been shown in [Publication VII] that other local minima than the ones corresponding to the ZF solutions are very unlikely to exist. Even if they would exist, the convergence can be achieved by proper initialization of the algorithm. This can be achieved by using a very small amount of training data. In this case the algorithm operates on a semi-blind mode. Semi-blind methods are more feasible in practice since they also resolve the inherent ambiguities which the blind algorithms are subjected to. Initialization strategies may also be found. A basic requirement is that the initial equalizer settings are non-zero vectors. This case of zero vectors corresponds to a maximum of the cost function and the coefficients will remain identically zero. Moreover, the initial settings of the sub-equalizers corresponding to different output data streams must not be identical. This is necessary because identical initial settings may cause the same data stream to be recovered at the corresponding outputs. This case corresponds to a saddle point. For details, see the comments related to Table 1 in [Publication VII]. The saddle points may be easily avoided in practice by ensuring the fact that the initial gradient value is non-zero. In that case, different initialization setting may be chosen. Initialization at saddle points has extremely low probability since these type of stationary points involve a very special structure of the global channel-equalizer impulse responses.

5.3 Discussion

We propose a blind equalizer for spatial multiplexing MIMO-OFDM systems based on the minimization of a composite criterion. It is able to perform the blind equalization and co-channel interference cancellation without estimating the MIMO channel matrix, unlike most of the existing SOS and HOS blind methods. For this reason, the proposed blind algorithm is computationally simple. It does not require expensive operations such as matrix inversions as usually required for the ZF and MMSE equalizers, or matrix decompositions employed in subspace methods [28, 84, 91–94, 129, 180, 226, 228] or the algebraic techniques [100, 166].

When the channel order exceeds the CP length used in the OFDM transmission, the benefit of single-tap equalization is lost. The proposed equalizer operates in time domain, before the FFT operation at the receiver. Consequently the equalizer is able to deal with three situations: no CP at all, CP too short or CP sufficiently long. CP is not needed in equalization, but it may be used for synchronization purposes, for example. The channel identifiability is conditioned on the common zeros of the MIMO channel branches, but this problem can be solved via non-redundant precoding [37, 91, 92, 129, 174, 175, 221, 222]. Unitary Cayley space-time codes [103, 118, 141, 165] may also be used due to the fact that they preserve the correlation properties of the transmitted space-time signals. Grassmann space-time codes [120] are also a good alternative, since they efficiently use the degrees of freedom of the MIMO channel [231].

Compared to the BSS methods [55, 102, 115, 173–177], or the tensor-based method in [166], which perform the channel estimation in frequency domain, the proposed blind algorithm has much lower complexity. Moreover, the ambiguities (inherent to all blind methods) do not affect every subcarrier, but every transmitted data stream. The proposed algorithm is able to recover each of the the transmitted data streams up to a phase ambiguity and possible delay. User permutation ambiguity may be also encountered. These can be resolved by using P pilot symbols within a single OFDM block. Since in practice P is much smaller than the length of the OFDM block, the semi-blind version still provides high data rate. The same amount pilots is necessary for subspace-based methods, but in addition these methods require either CP, ZP or VC. The proposed blind algorithm can be used for channel tracking using the data symbols only. It is suitable for MIMO-OFDM systems, under slow to moderate fading conditions such as in wireless LANs, continuous transmissions (television and radio), and fixed wireless communications systems.

Chapter 6

Summary

Optimization techniques are a key part of many array and multi-channel signal processing algorithms. Application domains include radar, multi-antenna communications, sensor arrays, biomedical applications. Often, the optimization needs to be performed subject to matrix constraints. In particular, orthogonal or unitary matrices play a crucial role in many tasks, for example, adaptive beamforming, interference cancellation, MIMO transmission, space-time coding, and signal separation. In order to obtain optimal or close to optimal performance, optimization algorithms are needed to minimize the selected error criterion or cost function. In many practical applications numerical optimization is the only computationally feasible solution. For this reason, in this work we focus on optimization under unitary matrix constraint.

In this thesis, reliable and computationally feasible constrained optimization algorithms are proposed. Riemannian steepest descent and conjugate gradient algorithms operating on the Lie group of unitary matrices $U(n)$ are derived. They take full benefit of the geometrical properties of the group, as well as the recent advances in numerical techniques. Two novel line search methods exploiting the almost periodic property of the cost function along geodesics on $U(n)$ are also proposed. The proposed algorithms are suitable for performing the joint diagonalization of a set of Hermitian matrices, which is a fundamental problem of blind source separation. They outperform the classical JADE approach based on Givens rotations [45] in terms of converge speed, at similar cost/iteration, as demonstrated in [Publication I], [Publication II], [Publication IV], [Publication V]. SD and CG on $U(n)$ are used for computing the full set of eigenvectors of a Hermitian matrix, by optimizing the off-norm cost function [112] in [Publication III], and the Brockett cost function [41, 184] in [Publication II], [Publication V].

Multi-antenna MIMO systems and multicarrier transmission such as OFDM are the key technologies in future wireless communication systems such as B3G Long Term Evolution (LTE), IMT-2000 and WiMAX [1, 8, 10].

In this work a blind receiver for MIMO-OFDM systems is proposed. The algorithm optimizes a composite criterion in order to cancel both inter-symbol and co-channel interference. Identifiability conditions and local convergence properties of the algorithm are established.

Possible topics of future research include extending the proposed algorithms to optimization w.r.t. non-square orthonormal matrices. When the optimization needs to be performed w.r.t. an orthonormal matrix with more rows than columns, the appropriate parameter space is the Stiefel manifold of $n \times p$ orthonormal matrices, $\mathrm{St}(n, p)$. This is the case of applications that require a distinct set of orthonormal vectors, such as limited-feedback MIMO systems, unitary space-time codes, MIMO radars and sonars. If the cost function possesses symmetries, such as invariance to right multiplication of its argument by unitary matrices, the appropriate parameter space is the Grassmann manifold $\mathrm{Gr}(n, p)$ of p-dimensional subspaces of the n-dimensional Euclidean space. This is the case of all subspace estimation and tracking techniques. Therefore, a broad range of array and multi-channel applications may be addressed. The most important applications include blind equalization and source separation, smart antennas, as well as biomedical applications.

The fact that Stiefel and Grassmann manifolds are homogeneous spaces, may be beneficial in reducing the computational complexity of optimization algorithms. Stiefel and Grassmann manifolds are quotient spaces arising form the unitary group $U(n)$ (for complex-valued matrices) of from orthogonal group $O(n)$ (for real-valued matrices). There are many properties inherited from the corresponding Lie groups that may be exploited. In conclusion, the fact that the proposed algorithms focus only on $U(n)$ does not have to be seen as a limitation to $n \times n$ unitary matrices.

Line search methods are crucial for the performance of the optimization algorithms. New approaches exploiting the almost periodicity of the cost function along geodesics are possible and they remain to be studied. The proposed DFT-based line search method opens multiple possibilities for finding better local minima (or to reach the global minimum faster).

Computationally efficient Riemannian Newton algorithms may also be addressed in future work. An important goal is to achieve complexity of order $\mathcal{O}(n^3)$ per iteration by fully exploiting the properties of $U(n)$. Trust-region methods [15, 18] present particular interest due to their desirable global convergence properties, which classical Newton algorithms do not possess, in general.

Another possible research topic for the future is developing algorithms for joint blind equalization and carrier frequency-offset compensation in MIMO-OFDM systems. This may lead to computationally efficient algorithms that enable high user mobility.

Bibliography

[1] Worldwide inter-operability for microwave access (WiMAX). WiMAX Forum: http://www.wimaxforum.org/home.

[2] Digital broadcasting system television, sound, and data services: framing structure, channel coding, and modulation digital terrestrial television. ETSI Standard: EN 300 744, 1996. European Telecommunications Standardization Institute, Sophia-Antipolis, Valbonne, France.

[3] Network and customer installation interfaces – asymmetric digital subscriber line (ADSL). ANSI Standard: T1.413, 1998. American National Standards Institute.

[4] Wireless local area network (LAN) medium access control (MAC) and physical layer (PHY) specifications: high speed physical layer in the 5 GHz band. IEEE Standards: IEEE 802.11a-1999, IEEE 802.11g-2003, IEEE 802.11n-2007, 1999/2003/2007.

[5] Broadband radio access networks (BRAN); high performance radio local area networks (HIPERLAN), type 2; physical (PHY) layer. ETSI Standard: TS 101 475, 20 Dec. 2001. European Telecommunications Standardization Institute, Sophia-Antipolis, Valbonne, France.

[6] Local and metropolitan area networks: Air interface for fixed broadband access systems. IEEE Standard: 802.16, 2004.

[7] Radio broadcasting system, digital audio broadcasting (DAB) to mobile, portable, and fixed receiver. ETSI Standard: EN 300 401, 2006. European Telecommunications Standardization Institute, Sophia-Antipolis, Valbonne, France.

[8] IMT-2000 (International Mobile Telecommunications-2000) OFDMA (Orthogonal Frequency Division Multiplexing) TDD (Time Division Duplex) WMAN (Wireless Metropolitan Area Network). ITU Recommendation: ITU-R M.1457, 2007. International Telecommunication Union.

[9] OFDMA broadband mobile wireless access system. ARIB Standards: STD-T94, STD-T95, 2008. Association of Radio Industries and Businesses, Tokyo, Japan.

[10] Third Generation Partnership Project (3GPP). Evolved universal terrestrial radio access (E-UTRA); long-term evolution (LTE) physical layer. 3GPP Standard TS 36.201, http://www.3gpp.org/.

[11] K. Abed-Meraim, A. Chkeif, and Y. Hua. Fast orthonormal PAST algorithm. *IEEE Signal Processing Letters*, 7(3):60–62, Mar. 2000.

[12] T. Abrudan, A. Hjørungnes, and V. Koivunen. Toeplitz method for blind equalization in MIMO OFDM systems. In *International Zürich Seminar on Communications, IZS 2004*, pages 212–215, Zürich 2004, Switzerland, 18–20 Feb. 2004.

[13] T. Abrudan, M. Sîrbu, and V. Koivunen. A block-Toeplitz VCMA equalizer for MIMO-OFDM systems. In *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 1037–1041, Pacific Grove, CA, 9–12 Nov. 2003.

[14] P.-A. Absil, C. G. Baker, and K. A. Gallivan. A truncated-CG style method for symmetric generalized eigenvalue problems. *Journal of Computational and Applied Mathematics*, 189(1–2):274–285, May 2006.

[15] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Foundations of Computational Mathematics*, 7(3):303–330, 2007.

[16] P.-A. Absil and K. A. Gallivan. Joint diagonalization on the oblique manifold for independent component analysis. Technical Report NA2006/01, DAMTP, University of Cambridge, http://www.damtp.cam.ac.uk/user/na/reports.html, 2006.

[17] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.

[18] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, Jan. 2008.

[19] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren. A Grassmann-Rayleigh quotient iteration for computing invariant subspaces. *SIAM Review*, 44:57–73, 2002.

[20] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis*, 22(3):359–390, Jul. 2002.

[21] T. Aittomäki and V. Koivunen. Signal covariance matrix optimization for transmit beamforming in MIMO radars. In *Forty-First Asilomar Conference on Signals, Systems and Computers, 2007*, pages 182–186, 4–7 Nov. 2007.

[22] S. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16:1451–1458, Oct. 1998.

[23] H. Ali, J. H. Manton, and Y. Hua. A SOS subspace method for blind channel identification and equalization in bandwidth efficient OFDM systems based on receive antenna diversity. In *11th IEEE Signal Processing Workshop on Statistical Signal Processing*, pages 401–404, Aug. 2001.

[24] S.-I. Amari. *Lecture notes in statistics*. Differential Geometry methods in Statistics. Springer-Verlag, 1985.

[25] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

[26] M. Journée amd P.-A. Absil and R. Sepulchre. Gradient-optimization on the orthogonal group for independent component analysis. In *7th International Conference on Independent Component Analysis and Signal Separation, ICA 2007*, pages 57–64, London, UK, 9–12 Sep. 2007.

[27] A. Antoniou and W.-S. Lu. *Practical Optimization: Algorithms and Engineering Applications*. Springer, 2007.

[28] W. Bai and Z. Bu. Channel identification in MIMO-OFDM systems. In *The IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication '04*, volume 2, pages 611–614, 31 May–2 Jun. 2004.

[29] W. Bai, C. He, L. Jiang, and H. Zhu. Blind channel estimation in MIMO-OFDM systems. In *IEEE Global Telecommunications Conference, GLOBECOM '02*, volume 1, pages 317–321, 17–21 Nov. 2002.

[30] W. Bai, H. Yang, and Z. Bu. Blind channel identification in SIMO-OFDM systems. In *International Conference on Communications, Circuits and Systems, ICCCAS 2004*, volume 1, pages 318–321, 27–29 Jun. 2004.

[31] D. Bartolomé, A. I. Pérez-Neira, and A. Pascual. Blind and semiblind spatio-temporal diversity for OFDM systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 2769–2772, 13-17 May 2002.

[32] D. P. Bertsekas. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[33] A. S. Besicovitch. *Almost periodic functions*. Dover, New York, 1954.

[34] D. W. Bliss and K. W. Forsythe. Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution. In *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 54–59, 2003.

[35] H. Bohr. *Almost periodic functions*. Chelsea, New York, 1951.

[36] H. Bölckei, D. Gesbert, and A. J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Transactions on Communications*, 50(2):225–234, Feb. 2002.

[37] H. Bölcskei, R. W. Heath, and A. J. Paulraj. Blind channel identification and equalization in OFDM-based multiantenna systems. *IEEE Transactions on Signal Processing*, 50(1):96–109, Jan 2002.

[38] S. Boyd and L. Vandenberghe. *Convex Optimization*. University Press, Cambridge, UK, 2004.

[39] D. H. Brandwood. A complex gradient operator and its applications in adaptive array theory. In *IEE Proceedings, Parts F and H*, volume 130, pages 11–16, Feb. 1983.

[40] R. W. Brockett. Least squares matching problems. *Linear Algebra and its Applications*, 122/123/124:761–777, 1989.

[41] R. W. Brockett. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Linear Algebra and its Applications*, 146:79–91, 1991.

[42] R. W. Brockett. Differential geometry and the design of gradient algorithms. *Proceedings of the Symposia in Pure Math, American Mathematical Society*, 54(1):69–92, 1992.

[43] R. W. Brockett. Singular values and least squares matching. In *Proceedings of 36th IEEE Conference on Decision and Control*, pages 1121–1124, San Diego, CA, USA, Dec. 1997.

[44] J.-F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44:3017–3030, 1996.

[45] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

[46] E. Celledoni and S. Fiori. Neural learning by geometric integration of reduced 'rigid-body' equations. *Journal of Computational and Applied Mathematics*, 172(2):247–269, 2004.

[47] E. Celledoni and A. Iserles. Methods for approximation of a matrix exponential in a Lie-algebraic setting. *IMA Journal on Numerical Analysis*, 21(2):463–488, 2001.

[48] Y. Y. Cheng, Y. Lee, , and H. J. Li. Subspace-MMSE blind channel estimation for multiuser OFDM with receiver diversity. In *IEEE Global Telecommunications Conference*, volume 4, pages 2295–2299, Dec. 2003.

[49] H. Cheon and D. Hong. A blind spatio-temporal equalizer using cyclic prefix in OFDM systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2957–2960, Jun. 2000.

[50] J. Choi and R. W. Heath Jr. Channel adaptive quantization for limited feedback MIMO beamforming systems. *IEEE Transactions on Signal Processing*, 54(12):4717–4729, Dec. 2006.

[51] J. Choi, B. Mondal, and R. W. Heath Jr. Interpolation based unitary precoding for spatial multiplexing MIMO-OFDM with limited feedback. *IEEE Transactions on Signal Processing*, 54(12):4730–4740, Dec. 2006.

[52] M. T. Chu and K. R. Driessel. The projected gradient methods for least squares matrix approximations with spectral constraints. *SIAM Journal on Numerical Analysis*, 27(4):1050–1060, 1990.

[53] P. Comon and G. H. Golub. Tracking a few extreme singular values and vectors in signal processing. In *Proceedings of the IEEE*, volume 78, pages 1327–1343, Aug. 1990.

[54] C. Corduneanu. *Almost periodic functions*. Interscience tracts in pure and applied mathematics. Interscience Publishers, New York, 1961.

[55] S. R. Curnew and J. Ilow. Blind signal separation in MIMO OFDM systems using ICA and fractional sampling. In *International Symposium on Signals, Systems and Electronics, ISSSE '07*, pages 67–70, 30 Jul.–2 Aug. 2007.

[56] J. Dehaene. *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*. PhD thesis, K. U. Leuven, Oct. 1995.

[57] J. Dehaene, C. Yi, and B. De Moor. Calculation of the structured singular value with gradient-based optimization algorithms on a Lie group of structured unitary matrices. *IEEE Transactions on Automatic Control*, 42(11):1596–1600, Nov. 1997.

[58] L. Deneire, E. De Carvalho, and D. T. M. Slock. Identifiability conditions for blind and semi-blind multiuser multichannel identification. In *Ninth IEEE SP Workshop on Statistical Signal and Array Processing*, pages 372–375, 14–16 Sept. 1998.

[59] M. P. do Carmo. *Riemannian Geometry*. Mathematics: theory and applications. Birkhauser, 1992.

[60] S. C. Douglas. Numerically-robust adaptive subspace tracking using Householder transformations. In *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*, pages 499–503, Mar. 2000.

[61] S. C. Douglas. Self-stabilized gradient algorithms for blind source separation with orthogonality constraints. *IEEE Transactions on Neural Networks*, 11(6):1490–1497, Nov. 2000.

[62] S. C. Douglas. On the design of gradient algorithms employing orthogonal matrix constraints. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, volume 4, pages 1401–1404, 15–20 Apr. 2007.

[63] S. C. Douglas, S.-I. Amari, and S.-Y. Kung. A self-stabilized minor subspace rule. *IEEE Signal Processing Letters*, 5(12):328–330, Dec. 1998.

[64] S. C. Douglas, S.-I. Amari, and S.-Y. Kung. On gradient adaptation with unit-norm constraints. *IEEE Transactions on Signal Processing*, 48(6):1863–1847, Jun. 2000.

[65] S. C. Douglas and S.-Y. Kung. An Ordered–Rotation KuicNet Algorithm for separating arbitrarily-distributed sources. In *Proceedings of IEEE International Conference on Independent Component Analysis and Signal Separation*, pages 419–425, Aussois, France, Jan. 1999.

[66] S. C. Douglas and X. Sun. Designing orthonormal subspace tracking algorithms. In *Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1441–1445, 2000.

[67] Jiang Du, Qicong Peng, and Yubei Li. Adaptive blind equalization for MIMO-OFDM wireless communication systems. In *International Conference on Communication Technology Proceedings, ICCT '03*, volume 2, pages 1086–1090, 31 May–2 Jun. 2003.

[68] Jiang Du, Qicong Peng, and Hongying Zhang. Adaptive blind channel identification and equalization for MIMO-OFDM wireless communication systems. In *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, PIMRC 2003*, volume 3, pages 2078–2082, 7–10 Sept. 2003.

[69] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[70] M. Enescu. *Adaptive Methods for Blind Equalization and Signal Separation in MIMO Systems*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, 2002.

[71] J. Eriksson. *Contributions to Theory and Algorithms of Independent Component Analysis and Signal Separation*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, Aug. 2004.

[72] J. Eriksson and V. Koivunen. Characteristic-function-based independent component analysis. *Signal Processing*, 83:2195–2208, Oct. 2003.

[73] S. Fiori. 'Mechanical' neural learning for blind source separation. *Electronics Letters*, 35(22):1963–1964, 28 Oct. 1999.

[74] S. Fiori. Stiefel-Grassman Flow (SGF) learning: further results. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, volume 3, pages 343–348, 24-27 Jul. 2000.

[75] S. Fiori. A theory for learning by weight flow on Stiefel-Grassman manifold. *Neural Computation*, 13:1625–1647, 2002.

[76] S. Fiori. Unsupervised neural learning on Lie group. *International Journal of Neural Systems*, 12(3 and 4):219–246, 2002.

[77] S. Fiori. Quasi-geodesic neural learning algorithms over the orthogonal group: a tutorial. *Journal of Machine Learning Research*, 1:1–42, Apr. 2005.

[78] S. Fiori and F. Piazza. Orthonormal strongly-constrained neural learning. In *IEEE International Joint Conference on Neural Networks*, volume 2, pages 1332–1337, 4-9 May 1998.

[79] S. Fiori, A. Uncini, and F. Piazza. Application of the MEC Network to principal component analysis and source separation. In *Proceedings of International Conference on Artificial Neural Networks*, pages 571–576, 1997.

[80] A. Fischer. Structure of Fourier exponents of almost periodic functions and periodicity of almost periodic functions. *Mathematica Bohemica*, 121(3):249–262, 1996.

[81] A. Fischer. Approximation of almost periodic functions by periodic ones. *Czechoslovak Mathematical Journal*, 48(123):193–205, 1998.

[82] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela. MIMO radar: an idea whose time has come. In *IEEE Radar Conference 2004*, pages 71–78, 2004.

[83] R. Fletcher. *Practical Methods of Optimization; (2nd ed.)*. Wiley-Interscience, New York, NY, USA, 1987.

[84] H. Fu, P. H. W. Fung, and S. Sun. Semiblind channel estimation for MIMO-OFDM. In *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, volume 3, pages 1850–1854, 5-8 Sept. 2004.

[85] D. R. Fuhrmann. A geometric approach to subspace tracking. In *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers*, volume 1, pages 783–787, 2-5 Nov. 1997.

[86] D. R. Fuhrmann, A. Srivastava, and Hojin Moon. Subspace tracking via rigid body dynamics. In *Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pages 578–581, 24-26 Jun. 1996.

[87] D. Gabay. Minimizing a differentiable function over a differential manifold. *Journal of Optimization Theory and Applications*, 37(2):177–219, Jun. 1982.

[88] J. Gallier. *Geometric methods and applications*, volume 38 of *Texts in applied mathematics*. Springer, 2000.

[89] J. Gallier. *Notes on Differential Geometry and Lie Groups*. Apr. 2008. Book in progress (2008) – available online at http://www.cis.upenn.edu/~jean/gbooks/manif.html.

[90] C. Gao, Ming Zhao, Shidong Zhou, and Yan Yao. Blind channel estimation algorithm for MIMO-OFDM systems. *Electronics Letters*, 39(19):1420–1422, 18 Sept. 2003.

[91] F. Gao and A. Nallanathan. Subspace-based blind channel estimation for SISO, MISO and MIMO OFDM systems. In *IEEE International Conference on Communications, ICC '06*, volume 7, pages 3025–3030, Jun. 2006.

[92] F. Gao and A. Nallanathan. Blind channel estimation for MIMO OFDM systems via nonredundant linear precoding. *IEEE Transactions on Signal Processing*, 55(2):784–789, Jan. 2007.

[93] F. Gao, W. Wu, Y. Zeng, and A. Nallanathan. A novel blind channel estimation for CP-based MIMO OFDM systems. In *IEEE International Conference on Communications, ICC '07*, pages 258–2591, 24–28 Jun. 2007.

[94] F. Gao, Y. Zeng, A. Nallanathan, and T.-S. Ng. Robust subspace blind channel estimation for cyclic prefixed MIMO OFDM systems: algorithm, identifiability and performance analysis. *IEEE Journal on Selected Areas in Communications*, 26(2):378–388, Feb. 2008.

[95] D. Gesbert, M. Shafi, D. Shiu, and P. Smith. From theory to practice: An overview of space-time coded MIMO wireless systems. *IEEE Journal on Selected Areas on Communications*, 21(3), Apr. 2003. Special issue on MIMO systems.

[96] M. Ghogho and A. Swami. Blind channel identification for OFDM systems with receive antenna diversity. In *IEEE Workshop Signal Processing Advances in Wireless Communications*, pages 378–382, Jun. 2003.

[97] P. E. Gill and W. Murray. The computation of Lagrange-multiplier estimates for constrained minimization. *Mathematical Programming*, 17(1):32–60, Dec. 1979.

[98] D. N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transaction on Communications*, 7(2):1867–1875, Nov. 1980.

[99] G. H. Golub and C. van Loan. *Matrix computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[100] A. Gorokhov. Blind equalization in SIMO OFDM systems with frequency domain spreading. *IEEE Transactions on Signal Processing*, 48(12):3536–3549, Dec. 2000.

[101] S. Gudmundsson. An introduction to Riemannian geometry, 2004. Lecture notes available at http://www.matematik.lu.se/matematiklu/personal/sigma/index.html.

[102] B. Guo, H. Lin, and K. Yamashita. Blind signal recovery in multiuser MIMO-OFDM system. In *The 2004 47th Midwest Symposium on Circuits and Systems, MWSCAS '04*, pages 637–640, 25–28 Jul. 2004.

[103] B. Hassibi and B. M. Hochwald. Cayley differential unitary space-time codes. *IEEE Transactions on Information Theory*, 48(6):1485–503, Jun. 2002.

[104] M. A. Haun. The fractionally spaced vector constant modulus algorithm. Master's thesis, University of Illinois at Urbana-Champaign, 1999.

[105] S. Haykin. *Adaptive Filter Theory*, volume 3rd edition. Prentice Hall, 1996.

[106] J. Heiskala and J. Terry. *OFDM Wireless LANs: A theoretical and practical guide*. SAMS Publishing, 2001.

[107] S. Helgason. *Differential geometry, Lie groups and symmetric spaces*. Academic Press, 1978.

[108] N. J. Higham. Matrix nearness problems and applications. In M. J. C. Gover and S. Barnett, editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press, 1989.

[109] A. Hjørungnes and D. Gesbert. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transaction on Signal Processing*, 55(6):2740–2746, Jun. 2007.

[110] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[111] A. Hottinen, O. Tirkkonen, and R. Wichman. *Multi-antenna transceiver techniques for 3G and beyond*. Wiley, Jan. 2003.

[112] K. Hüper, U. Helmke, and J. B. Moore. Structure and convergence of conventional Jacobi-type methods minimizing the off-norm function. In *Proceedings of 35th IEEE Conference on Decision and Control*, volume 2, pages 2124–2129, Kobe, Japan, 11-13 Dec 1996.

[113] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[114] O. A. Alim I. Ghaleb and K. Seddik. A new finite alphabet based blind channel estimation for OFDM systems. In *IEEE 5th Workshop on Signal Processing Advances in Wireless Communications*, pages 102 – 105, 11–14 Jul. 2004.

[115] D. Iglesia, A. Dapena, and C. J. Escudero. Multiuser detection in MIMO OFDM systems using blind source separation. In *Proc. 6th Baiona Workshop on Signal Processing in Communications*, pages 41–46, Sept. 2003.

[116] S. Ikeda, T. Tanaka, and S.-I. Amari. Information geometry of turbo and low-density parity-check codes. *IEEE Transactions on Information Theory*, 50(6):1097–1114, Jun. 2004.

[117] A. Iserles and A. Zanna. Efficient computation of the matrix exponential by general polar decomposition. *SIAM Journal on Numerical Analysis*, 42(5):2218–2256, Mar. 2005.

[118] Yindi Jing and B. Hassibi. Unitary space-time modulation via Cayley transform. *IEEE Transactions on Signal Processing*, 51(11):2891–2904, Nov. 2003.

[119] D. L. Jones. Property-restoral algorithms for blind equalization of OFDM. In *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 619–622, Pacific Grove, CA, 9–12 Nov. 2003.

[120] I. Kammoun and J.-C. Belfiore. A new family of Grassmann space-time codes for non-coherent MIMO systems. *IEEE Communications Letters*, 7(11):528–530, Nov. 2003.

[121] N. Khaled, B. Mondal, R. W. Heath Jr., G. Leus, and F. Petré. Quantized multi-mode precoding for spatial multiplexing MIMO-OFDM system. In *2005 IEEE 62nd Vehicular Technology Conference*, volume 2, pages 867–871, Sept. 2005.

[122] M. Kleinsteuber and K. Hüper. An intrinsic CG algorithm for computing dominant subspaces. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, volume 4, pages 1405–1408, Apr. 2007.

[123] A. Knapp. *Lie groups beyond an introduction*, volume 140 of *Progress in mathematics*. Birkhauser, 1996.

[124] S. G. Krantz. *Function theory of several complex variables*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 2nd edition, 1992.

[125] H. Krim and M. Viberg. Two decades of array signal processing research: the parametric approach. *IEEE Signal Processing Magazine*, 13(4):67–94, Jul. 1996.

[126] J. Laurila. *Semi-Blind Detection of Co-Channel Signals in Mobile Communications*. PhD thesis, Technische Universtit at Wien, Wien, Austria, Mar 2000. available online at http://www.nt.tuwien.ac.at/mobile/theses_finished/.

[127] B. M. Levitan and V. V. Zhikov. *Almost periodic functions and differential equations*. Cambridge, 1982.

[128] R. Liu and L. Tong (Ed.). Special issue on blind system identification and estimation. *Proceedings of the IEEE*, 86(10):1903–2116, Oct. 1998.

[129] Xia Liu and M. E. Bialkowski. SVD-based blind channel estimation for a MIMO OFDM system employing a simple block pre-coding scheme. In *The International Conference on "Computer as a Tool", EUROCON 2007*, pages 926–929, 9-12 Sept. 2007.

[130] J. Lu, T. N. Davidson, and Z.-Q. Luo. Blind separation of BPSK signals using Newton's method on the Stiefel manifold. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 301–304, Apr. 2003.

[131] D. G. Luenberger. The gradient projection method along geodesics. *Management Science*, 18:620–631, 1972.

[132] Y. Luo, J. A. Chambers, and S. Lambotharan. Global convergence and mixing parameter selection in the cross-correlation constant modulus algorithm for the multi-user environment. *IEE Vision, Image and Signal Processing*, 148(1):9–20, Feb. 2001.

[133] Y. Ma, J. Košecká, and S. Sastry. Motion estimation in computer vision: optimization on Stiefel manifolds. In *Proceedings of the 37th IEEE Conference on Decision and Control*, volume 4, pages 3751–3756, Dec. 1998.

[134] Yi Ma, Yi Huang, Xu Zhu, and Na Yi. Blind channel estimation for OFDM based multitransmitter systems using guard interval diversity. In *2004 IEEE 59th Vehicular Technology Conference*, volume 1, pages 440–444, 17-19 May 2004.

[135] R. Mahony and J. H. Manton. The geometry of the Newton method on non-compact Lie groups. *Journal of Global Optimization*, 23:309–327, 2002.

[136] J. H. Manton. A new algorithm for computing the extreme eigenvectors of a complex Hermitian matrix. In *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, pages 225–228, 6-8 Aug. 2001.

[137] J. H. Manton. Optimization algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635–650, Mar. 2002.

[138] J. H. Manton. On the role of differential geometry in signal processing. In *International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages 1021–1024, Philadelphia, Mar. 2005.

[139] J. H. Manton and Y. Hua. Convolutive reduced-rank Wiener filtering. In *Proceeding of IEEE Conference on Acoustics, Speech and Signal Processing*, May 2001.

[140] J. H. Manton, R. Mahony, and Y. Hua. The geometry of weighted low-rank approximations. *IEEE Transactions on Signal Processing*, 51(2):500–514, Feb. 2003.

[141] T. L. Marzetta, B. Hassibi, and B. M. Hochwald. Structured unitary space-time autocoding constellations. *IEEE Transactions on Information Theory*, 48(4):942–950, Apr. 2002.

[142] C. Moler and C. van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49, 2003.

[143] B. Mondal and R. W. Heath Jr. Algorithms for quantized precoded MIMO-OFDM systems. In *Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, pages 381–385, 2005.

[144] B. Mondal, R. W. Heath Jr., and L. W. Hanlen. Quantization on the Grassmann manifold: applications to precoded MIMO wireless systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, volume 5, pages 1025–1028, Mar. 2005.

[145] B. Mondal, R. Samanta, and R. W. Heath Jr. Frame theoretic quantization for limited feedback MIMO beamforming systems. In *2005 International Conference on Wireless Networks, Communications and Mobile Computing*, volume 2, pages 1065–1070, 2005.

[146] J. B. Moore and P. Y. Lee. Differential geometry applications to vision systems. In *Symposium on Mechanical Systems Control*, pages 1–30, Berkeley, CA, Jun. 2006.

[147] A. Neumaier. Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Review*, 40:636–666, 1998.

[148] M. Nikpour, K. Hüper, and J. H. Manton. Generalizations of the Rayleigh quotient iteration for the iterative refinement of the eigenvectors of real symmetric matrices. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, volume 5, pages 1041–1044, 2005.

[149] M. Nikpour, J. H. Manton, and G. Hori. Algorithms on the Stiefel manifold for joint diagonalisation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1481–1484, 2002.

[150] Y. Nishimori. Learning algorithm for independent component analysis by geodesic flows on orthogonal group. In *International Joint Conference on Neural Networks*, volume 2, pages 933–938, Jul. 10-16 1999.

[151] Y. Nishimori and S. Akaho. Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. *Neurocomputing*, 67:106–135, Jun. 2005.

[152] Y. Nishimori, S. Akaho, S. Abdallah, and M. D. Plumbley. Flag manifolds for subspace ICA problems. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, volume 4, pages 1417–1420, 15–20 Apr. 2007.

[153] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.

[154] K. Nomizu. Invariant affine connections on homogeneous spaces. *American Journal of Mathematics*, 76(1):33–65, Jan. 1954.

[155] B. Owren and B. Welfert. The Newton iteration on Lie groups. *BIT Numerical Mathematics*, 40(1):121–145, Mar. 2000.

[156] C. B. Papadias. Globally convergent blind source separation based on a multiuser kurtosis maximization criterion. *IEEE Transactions on Signal Processing*, 48(12):3508–3519, Dec. 2000.

[157] C. B. Papadias and A. M. Kuzminskiy. Blind source separation with randomized Gram-Schmidt orthogonalization for short burst systems. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 809–812, 17-21 May 2004.

[158] C. B. Papadias and A. J. Paulraj. A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays. *IEEE Signal Processing Letters*, 4(6):178–181, Jun. 1997.

[159] T. Petermann, S. Vogeler, K. D. Kammeyer, and D. Boss. Blind turbo channel estimation in OFDM receivers. In *Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1489–1493, Nov. 2001.

[160] M. D. Plumbley. Optimization using Fourier expansion over a geodesic for non-negative ICA. In *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2004*, pages 49–56, Granada, Spain, Sept. 2004.

[161] M. D. Plumbley. Geometrical methods for non-negative ICA: manifolds, Lie groups, toral subalgebras. *Neurocomputing*, 67:161–197, 2005.

[162] M. D. Plumbley. Geometry and manifolds for independent component analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, volume 4, pages 1397–1400, 15–20 Apr. 2007.

[163] E. Polak. *Optimization: Algorithms and Consistent Approximations*. New York: Springer-Verlag, 1997.

[164] N. Prabhu, H.-C Chang, and M. deGuzman. Optimization on Lie manifolds and pattern recognition. *Pattern Recognition*, 38:2286–2300, 2005.

[165] A. Qatawneh and L. de Haro Ariet. OFDM-MIMO system using Cayley differential unitary space time coding. In *Proceedings of Sensor Array and Multichannel Signal Processing Workshop 2004*, pages 254–258, 18–21 Jul. 2004.

[166] M. Rajih, P. Comon, and D. Slock. A deterministic blind receiver for MIMO OFDM systems. In *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications, SPAWC '06*, pages 1–5, 2–5 Jul. 2006.

[167] F. C. Robey, S. Coutts, D. Weikle, J. C. McHarg, and K. Cuomo. MIMO radar theory and experimental results. In *Conference Record of the Thirty-Eighth Asilomar Conference on Signals Systems and Computers*, volume 1, pages 300–304, 2004.

[168] T. Roman. *Advanced receiver structures for mobile MIMO multicarrier systems*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, Apr. 2006.

[169] T. Roman, M. Enescu, and V. Koivunen. Joint time-domain tracking of channel and frequency offsets for MIMO OFDM systems. *Wireless Personal Communications*, 31(3–4):181–200, Dec. 2004.

[170] R. Roy and T. Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, Jul. 1989.

[171] P. Sansrimahachai, D. B. Ward, and A. G. Constantinides. Blind source separation for BLAST. In *14th International Conference on Digital Signal Processing*, volume 1, pages 139–142, 1-3 Jul. 2002.

[172] P. Sansrimahachai, D. B. Ward, and A. G. Constantinides. Multiple-input multiple-output least-squares constant modulus algorithms. In *IEEE Global Telecommunications Conference*, volume 4, pages 2084–2088, 1-5 Dec. 2003.

[173] L. Sarperi, A. K. Nandi, and Xu Zhu. Multiuser detection and channel estimation in MIMO OFDM systems via blind source separation. In *Proc. 5th Int. Symposium on Independent Component Analysis and Blind Signal Separation, ICA 2004*, pages 1189–1196, Granada, Spain, Sept. 2004.

[174] L. Sarperi, X. Zhu, and A. K. Nandi. Low-complexity ICA based blind multiple-input multiple-output OFDM receivers. *Neurocomputing*, 69(13–15):1529–1539, 2006.

[175] L. Sarperi, X. Zhu, and Asoke A. K. Nandi. Blind OFDM receiver based on independent component analysis for multiple-input multiple-output systems. *IEEE Transactions on Wireless Communications*, 6(11):4079–4089, Nov. 2007.

[176] L. Sarperi, Xu Zhu, and A. K. Nandi. Blind layered space-time equalization for MIMO OFDM systems. In *in Proc. 13th European Signal Processing Conference, EUSIPCO 2005*, 2005.

[177] L. Sarperi, Xu Zhu, and A. K. Nandi. Reduced complexity blind layered space-time equalization for MIMO OFDM systems. In *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, pages 236–240, 11–14 Sept. 2005.

[178] R .O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, Mar. 1986.

[179] H. She and K. Hüper. Generalised FastICA for independent subspace analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, volume 4, pages 1409–1412, Apr. 2007.

[180] C. Shin and E. J. Powers. Blind channel estimation for MIMO-OFDM systems using virtual carriers. In *IEEE Global Telecommunications Conference, GLOBECOM '04*, volume 4, pages 2465–2469, 29 Nov.–3 Dec. 2004.

[181] A. I. Shtern. Almost periodic functions and representations in locally convex spaces. *Russian Math. Surveys*, 60(3):489–557, 2005.

[182] S. T. Smith. Dynamical systems that perform singular value decomposition. *Systems and Control Letters*, 16(5):319–327, May 1991.

[183] S. T. Smith. *Geometric optimization methods for adaptive filtering.* PhD thesis, Harvard University, Cambridge, MA, May 1993.

[184] S. T. Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications, American Mathematical Society*, 3:113–136, 1994.

[185] S. T. Smith. Linear and non-linear conjugate gradient methods for adaptive processing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 1834–1837, Atlanta, GA, USA, May 1996.

[186] S. T. Smith. Subspace tracking with full rank updates. In *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 793–797, 2-5 Nov. 1997.

[187] S. T. Smith. Optimum phase-only adaptive nulling. *IEEE Transactions on Signal Processing*, 47(7):1835–1843, Jul 1999.

[188] S. T. Smith. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Transactions on Signal Processing*, 53(5):1610–1630, May 2005.

[189] S. T. Smith. Statistical resolution limits and the complexified Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 53(5):1597–1609, May 2005.

[190] J. H. Son and D. B. Williams. A blind symbol recovery for dual antenna OFDM systems. In *IEEE 10th Digital Signal Processing Workshop*, pages 29–34, 13-16 Oct. 2002.

[191] A. Srivastava and E. Klassen. Monte Carlo extrinsic estimators of manifold-valued parameters. *IEEE Transactions on Signal Processing*, 50(2):299–308, Feb. 2002.

[192] P. Stoica and D. A. Linebarger. Optimization result for constrained beamformer design. In *IEEE Signal Processing Letters*, volume 2, pages 66–67, Apr. 1995.

[193] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Transactions on Information Theory*, 44:744–765, Mar. 1998.

[194] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov. 1999.

[195] A. Touzni, L. Tong, R. A. Casas, and C. R. Johnson, Jr. Vector-CM stable equilibrium analysis. *IEEE Signal Processing Letters*, 7(2):31–33, Feb. 2000.

[196] C. Udrişte. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Mathematics and Its Applications. Kluwer Academic Publishers Group, Boston, MA, 1994.

[197] W. Utschick and C. Brunner. Efficient tracking and feedback of DL-eigenbeams in WCDMA. In *Proceedings of the 4th European Personal Mobile Communications Conference*, Vienna, Austria, 2001.

[198] A. van den Bos. Complex gradient and Hessian. *IEE Vision, Image and Signal Processing*, 141(6):380–383, Dec. 1994.

[199] A.-J. van der Veen. Algebraic methods for deterministic blind beamforming. *Proceedings of IEEE,*, 86(10):1987–2008, 1998.

[200] R. van Nee and R. Prasad. *OFDM for Wireless Multimedia Communications*. Artech House, 2000.

[201] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(5):4–24, Apr. 1988.

[202] S. Visuri and V. Koivunen. Resolving ambiguities in subspace-based blind receiver for MIMO channels. In *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 589–593, 3-6 Nov. 2002.

[203] E. Viterbo and J. Boutros. A universal lattice code decoder for fading channels. *IEEE Transactions on Information Theory*, 45:1639–1642, Jul. 1999.

[204] Hao Wang, Ying Lin, and Biao Chen. Data-efficient blind OFDM channel estimation using receiver diversity. *IEEE Transactions on Signal Processing*, 51(10):2613–2623, Oct. 2003.

[205] L. Wang, J. Karhunen, and E. Oja. A Bigradient optimization approach for robust PCA, MCA and source separation. In *Proceedings of IEEE Conference on Neural Networks*, volume 4, pages 1684–1689, 27 Nov.-1 Dec. 1995.

[206] X. M. Wang, W.-S. Lu, and A. Antoniou. Blind adaptive multiuser detection using a vector constant-modulus approach. In *Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 36–40, Pacific Grove, CA, 4–7 Nov. 2001.

[207] Z. Wang and G. B. Giannakis. Wireless multicarrier communications. *IEEE Signal Processing Magazine*, 17(3):29–48, May 2000.

[208] F. W. Warner. *Foundations of differentiable manifolds and Lie groups*. Graduate Texts in Mathematics. Springer-Verlag New York, LLC, Oct. 1983.

[209] M. Wax and Y. Anu. A new least squares approach to blind beamforming. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3477–3480, 21-24 Apr. 1997.

[210] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela. V-BLAST: an architecture for realizing very high data rates over rich-scattering channel. In *Proc. of 1998 URSI International Symposium on Signals, Systems and Electronics, ISSSE 98*, pages 295–300, 29 Sept.–2 Oct. 1998.

[211] C. Wu, D. Gao, and L. Zhang. The normal matrix approach to the design of multivariable robust control systems. In *IEEE TENCON*, pages 203–207, Beijing, 1993.

[212] J. Xavier and V. Barroso. The Riemannian geometry of certain parameter estimation problems with singular Fisher information matrix. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '04*, volume 2, pages 1021–1024, Montreal, Canada, May 2004.

[213] J. Xavier and V. Barroso. Intrinsic Variance Lower Bound (IVLB): An extension of the Cramér-Rao bound to Riemannian manifolds. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05*, volume 5, pages 1033–1036, Mar. 2005.

[214] I. Yamada and T. Ezaki. An orthogonal matrix optimization by dual Cayley parametrization technique. In *Proceedings of ICA*, pages 35–40, 2003.

[215] B. Yang. Projection approximation subspace tracking. *IEEE Transactions on Signal Processing*, 43(1):95–107, Jan. 1995.

[216] J.-F. Yang and M. Kaveh. Adaptive eigensubspace algorithms for direction or frequency estimation and tracking. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(2):241–251, Feb. 1988.

[217] V. Y. Yang and D. L. Jones. A vector constant modulus algorithm for shaped constellation equalization. In *Thirty-First Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 590–594, Pacific Grove, CA, 2–5 Nov. 1997.

[218] V. Y. Yang and D. L. Jones. A vector constant modulus algorithm for shaped constellation equalization. *IEEE Signal Processing Letters*, 5(4):89–91, Apr. 1998.

[219] Y. Yang. Optimization on Riemannian manifold. In *Proceedings of the 38th Conference on Decision & Control*, pages 888–893, Phoenix, Arizona USA, Dec. 1999.

[220] Y. Yang. Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *Journal of Optimization Theory and Applications*, 132(2):245–265, Feb. 2007.

[221] S. Yatawatta and A. P. Petropulu. Blind channel estimation in MIMO OFDM systems. In *IEEE Workshop on Statistical Signal Processing*, pages 363–366, 28 Sept.–1 Oct. 2003.

[222] S. Yatawatta and A. P. Petropulu. Blind channel estimation in MIMO OFDM systems with multiuser interference. *IEEE Transactions on Signal Processing*, 54(3):1054–1068, Mar. 2006.

[223] C. Yi. *Robustness analysis and controller design for systems with structured uncertainties*. PhD thesis, K. U. Leuven, May 1995.

[224] A. Zanna and H. Z. Munthe-Kaas. Generalized polar decomposition for the approximation of the matrix exponential. *SIAM Journal on matrix Analysis*, 23(3):840–862, Jan. 2002.

[225] J. Zeng. The first sign change of a cosine polynomial. *Proceedings of the American Mathematical Society*, 111(3):709–716, Mar. 1991.

[226] Y. Zeng, W. H. Lam, and T.-S. Ng. Semiblind channel estimation and equalization for MIMO space-time coded OFDM. *IEEE Transactions on Circuits and Systems I*, 53(2):463–474, Feb. 2006.

[227] Y. Zeng and T.-S. Ng. A proof of the identifiability of a subspace-based blind channel estimation for OFDM systems. *IEEE Signal Processing Letters*, 11(9):756–759, Sept. 2004.

[228] Y. Zeng and T.-S. Ng. A semi-blind channel estimation method for multiuser multiantenna OFDM systems. *IEEE Transactions on Signal Processing*, 52(5):1419–1429, May 2004.

[229] L. Zhang. Conjugate gradient approach to blind separation of temporally correlated signals. In *IEEE International Conference on Communications, Circuits and Systems, ICCCAS-2004*, volume 2, pages 1008–1012, Chengdu, China, 2004.

[230] L. Q. Zhang, A. Cichoki, and S.-I. Amari. Natural Gradient Algorithm for blind separation of overdetermined mixture with additive noise. *Signal Processing Letters*, 6(11):293–295, Nov. 1999.

[231] L. Zheng and D. N. C. Tse. Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Transactions on Information Theory*, 48(2):359–383, Feb. 2002.

[232] S. Zhou and G. B. Giannakis. Finite-alphabet based channel estimation for OFDM and related multicarrier systems. *IEEE Transactions on Communications*, 49(8):1402–1414, 2001.