KTH Electrical Engineering

Optimization Techniques for Future Cellular Systems:

Harnessing the gains from higher frequencies,
increased spectral efficiency, and densification

HADI GHAUCH

Doctoral Thesis in Electrical Engineering
Stockholm, Sweden 2016

KTH, School of Electrical Engineering
Department of Communication Theory
SE-10044 Stockholm
SWEDEN

TRITA-EE 2016:166
ISSN: 1653-5146
ISBN: 978-91-7729-147-3

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i Elektroteknik onsdag den 23 november 2016 klockan 13:00 i hörsal F3, Lindstedtsvägen 26, Stockholm.

Tryck: Universitetsservice US AB

Abstract

In this thesis, we study optimization techniques for future cellular systems. We focus on all three directions for increasing data rates, namely, larger system bandwidth offered by millimeter-wave systems, increased spectral efficiency via cellular coordination, and densification via cloud radio access networks.

The first part is concerned with the investigation of the hybrid analog-digital architecture, for millimeter-wave multiple-input multiple-output (MIMO) systems. In this architecture, the precoding and combining are done in two steps: analog and digital. After characterizing the optimal precoders/combiners, we focus on open issues such as channel estimation and the design of precoders/combiners. Exploiting channel reciprocity in time-division duplex MIMO, and the sparsity of eigenmodes, we propose an algorithm (based on the Arnoldi Iteration) for estimating the dominant subspace of the channel, and provide basic analytical bounds on its estimation error. Moreover, we propose a mechanism for optimizing the precoders/combiners, to approximate the estimated subspace.

Distributed coordination schemes for cellular networks is the aim of the second part: designing the transmit (resp. receive) filters, at the base station (resp. users), in a distributed manner. Despite an existing body of work, such algorithms require a large number of over-the-air iterations (hundreds/thousands). As the resulting overhead could potentially destroy the gains of such coordination algorithms, we propose the use of fast-converging algorithms (a few iterations), focusing on algorithms for (I) sum-rate maximization, and (II) leakage minimization. In the case of (I), we optimize a lower bound on the sum-rate, and derive the corresponding optimal transmit and receive filter update: we dub the latter as non-homogeneous waterfilling, and highlight its inherent ability to turn-off streams with low-SINR, thus greatly speeding up the convergence of the algorithm. For (II), we relax the classical leakage problem, and propose two different filter update structures: rank-preserving and rank-reducing updates. Inspired from the decoding of turbo codes, we introduce turbo iterations (within each main iteration) for transmit/receive filters, for improved convergence speed. The combined effect of introducing rank-reducing updates and the turbo iteration, results in massively faster convergence.

In the final part, we investigate densification.The additional degrees-of-freedom gains from having more base stations / antennas are contingent upon having effective means of combating interference. Due to its centralized nature, cloud radio access networks enable tight coordination of several radio-heads to form an antenna domain. We define the so-called antenna domain formation problem, as the optimal assignment of users to antenna domains. Using the total interference leakage as metric, we formulate it as an integer optimization problem, and devise an iterative solution method. Motivated by the complicated nature of the problem, we propose the use of lower bounds on the problem in question (and the interference leakage consequently). We derive the corresponding Dantzig-Wolfe decomposition, the dual problem, and show that the former provides a tighter bound on the problem.

Sammanfattning

I denna avhandling studeras optimeringstekniker för framtida cellulära system. De tre huvudinriktningarna för att öka datatakterna i systemet studeras, nämligen, ökad systembandbredd genom användningen av millimetervågsystem, ökad spektraleffektivitet genom cellulär samordning, samt basstationsförtätning genom radioaccessnät i molnet.

I den första delen av avhandlingen utreds en hybridarkitektur för millimetervågssystem med flera antenner. I denna arkitektur sker förkodning och mottagningsfiltrering i två steg: först ett analogt steg och sedan ett digitalt steg. Först karakteriseras de optimala förkodnings- och mottagningsfiltrena och sedan studeras öppna frågor såsom kanalskattning och filterdesign. Genom att utnyttja reciprociteten från tidsdelningsduplexning, samt glesheten hos egenmoderna i kanalen, föreslås en algoritm baserad på Arnoldi-iterationer för att skatta det dominerande underrummet hos kanalen. Algoritmen ger även grundläggande analytiska gränser för skattningsfelet. Slutligen föreslås en mekanism för att optimera förkodnings- och mottagningsfiltrena så att de approximerar det skattade underrummet.

I den andra delen studeras distribuerade system för samordningen av cellulära nätverk, speciellt distribuerad utformning av förkodnings- samt mottagningsfilter hos basstationer och användare. De existerande algoritmerna i litteraturen kräver ett stort antal 'over-the-air'-iterationer, typiskt hundratals eller tusentals. Eftersom den resulterande overheaden skulle förstöra vinsterna av samordningen föreslår avhandlingen istället snabbkonvergerande algoritmer som bara kräver ett fåtal iterationer. Två fall studeras: summadatataktsmaximering samt störningsläckageminimering. I det första fallet maximeras en undre gräns för summadatatakten och de optimala filteruppdateringsekvationerna härleds. Optimeringsmetoden är en form av icke-homogen vattenfyllnad och har möjlighet att stänga av dataströmmar med låga signal-till-brus-och-störnings-förhållanden, vilket avsevärt påskyndar konvergensen hos algoritmen. I det senare fallet så relaxeras det klassiska störningsläckageproble -met och två olika filteruppdateringsstrukturer föreslås: rang-bevarande och icke-rang-bevarande uppdateringsekvationer. Inspirerade av avkodningen av turbokoder så introduceras turboiterationer (inom varje yttre 'over-the-air'-iteration) för filtrena. Den kombinerade effekten av icke-rang-bevarande uppdateringar och turboiterationer ger avsevärt snabbare konvergens.

I den sista delen undersöks basstationsförtätning, vilket är det mest verkningsfulla sättet att öka datatakterna i systemet. De ökade frihetsgraderna som erhålls genom flera basstationer och/eller antenner kräver goda sätt att hantera den uppkomna störningen. Tack vare dess centraliserade natur så kan ett radioaccessnätverk i molnet tillåta en stram koordinering av flera radiohuvuden som därmed gemensamt bildar en antenndomän. Avhandlingen definierar ett antenndomänbildningsproblem som den optimala tilldelningen av användare till antenndomänerna. Metriken som används är det totala störningsläckaget och ett heltalsoptimeringsproblem formuleras tillsammans med en iterativ lösningsmetod. På grund av problemets komplice-

rade natur så optimeras en undre gräns av störningsläckaget. Avhandlingen härleder Dantzig-Wolfeuppdelningen för problemet, det duala problemet, och visar att det senare är en stramare undre gräns för problemet.

In the loving memory of my grandfather, Farid Farah, for teaching me the power of knowledge and education.

Cet ouvrage est dédié à la mémoire de mon grand-père, Farid Farah, pour engraver en moi le pouvoir de la connaissance et de l'éducation.

Acknowledgments

As my PhD draws to its end, I must acknowledge several teachers that inspired me, and shaped the person that I am today. I have tried my best to avoid this acknowledgment degenerating into an Academy Award acceptance speech. But in some instances, cheesy catchphrases were all I could muster; maybe catchphrases are catchy for a reason.

First and foremost, I would like to offer my heartfelt gratitude to my supervisor Professor Mikael Skoglund: *"During my time in KTH, you have been a role model, an inspiration, and a father figure, both as a researcher and a person."* Though you are a man of few sentences, your words have high entropy! Thank you for every 'bit' of information. I am also grateful for your constant enrichment of my musical background, with great artists. My utmost gratitude also goes to Professor Mats Bengtsson for the countless discussions - technical and otherwise: *"I am eternally grateful for your patience, for your constant availability, and for your invaluable help in teaching me the value of good writing"*. I cannot begin to apologize for the countless vacations that I have ruined! None of this would have been possible without the infinite hours put in by Assistant Professor Taejoon Kim: *"My dear friend, you have my eternal gratitude for teaching me the value of discipline and hard work."* I am also indebted to all the help and support from Associate Professor James Gross: *"Your positive attitude and energy are always inspiring! "*

I am also grateful to my thesis grading committee members, Professor Wei Yu, Professor Mari Kobayashi, Professor Fredrik Tufvesson and Professor Thomas Eriksson for taking time out of their busy schedule to be here. Many thanks for everybody that helped me with proofreading the thesis. To my close friends, Ahmad, Anthony, Arun, Charbel, David, Lucy, Marija, Nino, Mikko, Pierre, Rasmus, Rami, Sebastian, Samer, Siwar, Taejoon, Victor, Walid and Zhao: *"Thank you for enriching my life in so many ways."* I am extremely grateful to Efthymios, Efthymis, Farshad, Frederic, Joakim, Johan, German, Gabor, Halkan, Mahboob, Peter, Ragnar, Qiwei, Sahar, Sheng, Sumin and Tobias, for making my time in KTH so interesting and rich with ideas and discussions. With my utterly unreliable memory, I am sure I forgot somebody - I am terribly sorry about that.

I am utterly grateful to my extended family, namely, Chadi and Georges for teaching me my first lessons in math, Hanane and Nazem for their support over all the years. Last but not least, I wish to thank my parents George and Dunia: *" This achievement is the culmination of all your years of hard work, sacrifice and dedication to your children."* Most of all, I am eternally grateful for all the support and encouragement of my brother, Ziad. *" Your dedication to your goals will never seize to amaze and inspire me. "* Thank you for being the 'older' brother, in several ways, and for being an amazing one: *"To you I dedicate this work!"*

Hadi Ghauch
Stockholm, November 23, 2016

Contents

Chapter 1

Introduction

Mathematical Notation

The mathematical notation used throughout the thesis is summarized below. Additional notation will be specified within the individual chapters, when needed.

1.0.1 Sets

Calligraphic letters $\mathcal{X}$ are used to denote sets, and $|\mathcal{X}|$ denotes the cardinality of the discrete set $\mathcal{X}$. The following notations are used for common sets:

$\mathbb{R}$ / $\mathbb{R}_+$	set of real numbers / set of non-negative real numbers
$\mathbb{Z}$ / $\mathbb{Z}_+$	set of integers / set of non-negative integers
$\mathbb{B}$	set of binary number
$\mathbb{C}$	set of complex numbers
$\mathbb{S}_+^d$	set of $d \times d$ positive semi-definite matrices
$\mathbb{S}_{++}^d$	set of $d \times d$ positive definite matrices
$\{n\}$	set of integers from 1 to n
$\mathcal{K} \backslash i$	set $\mathcal{K}$ after removing the element i, $i \in \mathcal{K}$
$\mathcal{U}(n, k)$	set of $n \times k$ ($k \leq n$) unitary matrices
$\text{conv}(\mathcal{X})$	convex hull of a set $\mathcal{X}$

Scalars, Vectors, Matrices

We use lowercase letters to denote scalar quantities, e.g., $x, y, ...$, bold lowercase letters to denote vectors, e.g., $\boldsymbol{w} = (w_1, w_2, ..., w_n)^T$ denote an n-dimensional vector,

and bold uppercase letter to denote matrices, e.g.,

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix}.$$

Moreover, we denote by

$\boldsymbol{I}_n$	the $n \times n$ identity matrix
$\boldsymbol{0}_{n\times m}$	the $n \times m$ all zeros matrix
$\boldsymbol{0}_n$	the n-dimensional all zeros vector
$\boldsymbol{1}_n$	the n-dimensional all ones vector

Operators

We use the following superscripts/superscripts

^{T}	the transpose of a vector/matrix
^{c}	the complex conjugate of a scalar/vector/matrix
†	the conjugate transpose (hermitian) of a complex scalar/vector/matrix.
$V^\perp$	the orthogonal complement of a subspace V
$\|\boldsymbol{x}\|_2$	the l_2-norm (Euclidean norm) of $\boldsymbol{x}$
$\|\boldsymbol{x}\|_1$	the l_1-norm of $\boldsymbol{x}$
$\boldsymbol{x}_{(i)}$	the ith element of $\boldsymbol{x}$
$\mathrm{diag}(\boldsymbol{x})$	diagonal matrix with the elements of $\boldsymbol{x}$ on the main diagonal

For any given square matrix $\boldsymbol{A}$,

$[\boldsymbol{A}]_{(i:j)}$	the matrix formed by taking columns i to j
$\boldsymbol{A}_{(i)}$	the ith column
$\boldsymbol{A}_{(i,j)}$	element (i, j) of $\boldsymbol{A}$
$\mathrm{tr}(\boldsymbol{A})$	the trace
$\|\boldsymbol{A}\|_F$	the Frobenius norm
$\|\boldsymbol{A}\|$	the determinant
$[\boldsymbol{A}]_{SL}$	matrix formed by the strictly lower triangular part (zeros everywhere else)
$[\boldsymbol{A}]_U$	matrix formed by the upper triangular part (zeros everywhere else)
$\sigma_i[\boldsymbol{A}]$	ith singular value of $\boldsymbol{A}$ (sorted in decreasing order)
$\sigma_{\min}[\boldsymbol{A}]$	the smallest singular value of $\boldsymbol{A}$
$\sigma_{\max}[\boldsymbol{A}]$	the largest singular value of $\boldsymbol{A}$

Let $\boldsymbol{A} \in \mathbb{S}_+^n$ and $\boldsymbol{B} \in \mathbb{S}_+^n$. Unless otherwise stated in the corresponding chapter, we adopt the convention of sorting the eigenvalues of $\boldsymbol{A}$ in decreasing order. Then,

$\lambda_i[\boldsymbol{A}]$	the ith eigenvalue of $\boldsymbol{A}$
$\lambda_{\min}[\boldsymbol{A}]$	the smallest eigenvalue of $\boldsymbol{A}$
$\lambda_{\max}[\boldsymbol{A}]$	the largest eigenvalue of $\boldsymbol{A}$
$v_{1:d}[\boldsymbol{A}]$	matrix having as columns the eigenvectors corresponding to d-largest eigenvalues of $\boldsymbol{A}$
$\boldsymbol{A} \succeq \boldsymbol{0}$	implies that $\boldsymbol{A} \in \mathbb{S}_{+}^{n}$
$\boldsymbol{A} \succ \boldsymbol{0}$	implies that $\boldsymbol{A} \in \mathbb{S}_{++}^{n}$
$\boldsymbol{A}-\boldsymbol{B} \succeq \boldsymbol{0}$	implies that $\boldsymbol{A}-\boldsymbol{B} \in \mathbb{S}_{+}^{n}$
$\boldsymbol{A}-\boldsymbol{B} \succ \boldsymbol{0}$	implies that $\boldsymbol{A}-\boldsymbol{B} \in \mathbb{S}_{++}^{n}$

1.0.2 Random Variables

$\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{K})$ represents a random vector $\boldsymbol{x}$, that is drawn from a circularly symmetric complex Gaussian distribution, with zero mean and covariance matrix $\boldsymbol{K}$. Moreover, $\mathbb{E}[\boldsymbol{x}]$ denotes the expectation of the random variable $\boldsymbol{x}$.

1.0.3 Order and Special Functions

Let f and g be two functions defined on some subsets of real numbers. Then, $f(x) = O(g(x))$ as $x \to \infty$ if and only if there exists a positive real number M and a real number x_0 such that $|f(x)| \leq M|g(x)|$ for all $x \geq x_0$.

Common Abbreviations

3G/4G/5G	3rd/4th/5th Generation Cellular Systems
AD	Antenna Domain
AN	Aggregation Node
AWGN	Additive White Gaussian Noise
BS	Base Station
BCD	Block-Coordinate Descent
CSI	Channel State Information
Cloud-RAN	Could Radio-Access Network
DLT	Difference of Log and Trace
DW	Dantzig-Wolfe
FDD	Frequency Division Duplex

IA	Interference Alignment
IBC	Interfering Broadcast Channels
IC	Interference Channel
IMAC	Interfering Multiple-Access Channel
IP	Integer Program
KKT	Karush-Kuhn-Tucker
LP	Linear Program
LM	(Interference) Leakage Minimization
LR	Lagrange Relaxation
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MS	Mobile Station
mmWave	Millimetre-Wave
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SRM	Sum-Rate Maximization
TDD	Time Division Duplex
UE	User Equipment

1.1 Overview of Relevant Optimization Techniques

Optimization has been the backbone for decades of progress in the areas of signal processing and communication. Indeed, a plethora of problems in signal processing for wireless communication are posed as optimization problems. Such problems are too numerous to name and range from the minimum mean-squared error estimator, the maximum likelihood estimator, the optimal precoder design in single-user (and multi-user) MIMO, the user assignment problem, all the way to joint transmit/receive filter design in multi-user multi-cell cellular networks. Problems in signal processing for wireless communication are tackled by a wide range of optimization techniques, such as standard Lagrange techniques for convex optimization, (integer) linear programming, block-coordinate descent methods (or alternating optimization), primal-dual decompositions, relaxations, semi-definite programming, etc. We review some state-of-the-art optimization techniques, focusing on modern and prevalent ones in the field of wireless communication. Methods such as the Block-Coordinate Descent (BCD), and standard Lagrangian are ubiquitous for works that fall within the scope of the thesis.

1.1.1 Block Coordinate Descent

Block-Coordinate Descent (BCD), also known as Gauss-Seidel method, is a generalization of the well known Alternating Optimization (or Coordinate Descent method) technique. BCD consists of fixing all but one block of variables, while optimizing that latter block. It is the most used technique in this thesis. Hence, its survey will be more detailed than that of other optimization methods.

Mathematical Description Put into a more rigorous context, let $f(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ be a function consisting of N blocks of variables, $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$, that needs to be minimized,

$$(P) \begin{cases} \min & f(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) \\ \text{s. t.} & \boldsymbol{x}_k \in \mathcal{C}_k, \forall k \in \{N\} \end{cases} \tag{1.1.1}$$

where $\mathcal{C}_k$ is a closed convex set, and f is continuous. BCD produces a sequence of iterates, $\{\boldsymbol{x}_1^l, ..., \boldsymbol{x}_N^l\}_l$ such that,

$$\boldsymbol{x}_k^{l+1} \triangleq \underset{\boldsymbol{w}_k \in \mathcal{C}_k}{\operatorname{argmin}} \; f(\boldsymbol{w}_k, \boldsymbol{z}_k^l) \tag{1.1.2}$$

where $\boldsymbol{z}_k^l = \boldsymbol{x}_1^{l+1}, ..., \boldsymbol{x}_{k-1}^{l+1}, \boldsymbol{x}_{k+1}^l, ..., \boldsymbol{x}_N^l$, denotes the block of fixed variables, for $\boldsymbol{x}_k$ at the lth iteration.

Convergence The convergence of BCD is the object of a wide array of investigations: there are many convergence results, each with a specific set of assumptions about $f(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$. The most generic BCD convergence results were derived in [Tse01], and typically require two conditions:

> f has a unique minimum in $N-2$ blocks of variables (e.g. f needs to have a unique minimum in blocks $\boldsymbol{x}_1, ..., \boldsymbol{x}_{N-2}$), and
>
> f is quasi-convex in each block of variables.

Then, the sequence $\{f(\boldsymbol{x}_1^l, ..., \boldsymbol{x}_N^l)\}_l$ convergence to a stationary point of (P).

Other results about BCD convergence, hinge upon the idea that the minimizer found in each of the steps above, be unique. This in turn requires that the function be separable in each of the blocks, and that $f(\boldsymbol{x}_k, \boldsymbol{z}_k^l)$ is strongly convex in $\boldsymbol{x}_k$ (i.e., when fixing everything but block $\boldsymbol{x}_k$, f is strongly convex in $\boldsymbol{x}_k$). Then, the sequence $\{f(\boldsymbol{x}_1^l, ..., \boldsymbol{x}_N^l)\}_l$ converges to a stationary point of (P). When the above conditions are satisfied, BCD is a quite a powerful technique. The convergence of BCD has been extended to more generic settings, such as non-smooth optimization [RHL12].

Generalizations A generalization of BCD was recently proposed, the so-called Block-Successive Upper-bound Minimization (BSUM) [RHL12]. BSUM generalizes the BCD update in 1.1.2 as follows:

$$\boldsymbol{x}_k^{l+1} \triangleq \underset{\boldsymbol{w}_k \in \mathcal{C}_k}{\operatorname{argmin}} \; u_k(\boldsymbol{w}_k, \boldsymbol{z}_k^l), \tag{1.1.3}$$

where $u_k(\boldsymbol{x}_k, \boldsymbol{z}_k^l)$ is a well-chosen approximation of $f(\boldsymbol{x}_k, \boldsymbol{z}_k^l)$. BSUM offers a strict advantage over BCD in terms of convergence, i.e., there are cases where BCD does not converge while BSUM does.

Applications In the last decade, BCD has been one of most prevalent optimization techniques in several areas of signal processing. In the context of distributed coordination algorithms for cellular networks - a relevant topic to this thesis, BCD is the underlying method in most of the algorithms: indeed approaches such as interference leakage minimization [GCJ11], [pet09] minimum mean-squared error minimization (MMSE) [SSB$^+$09, PH11], and weighted minimum mean-squared error minimization [SRLH11], to name a few, all have that same underlying method. While such approaches are essentially precoder design problems (i.e., joint optimization of transmit/receive filters), recent work applies BCD to the problem of joint precoder design and user assignment ([HXRL13, SRL14]).

In recent years, the BCD method was also applied to areas outside signal processing, such as (group) Lasso, basis denoising pursuit, low-rank matrix recovery, hybrid Huberized support vector machine, blind source separation, sparse dictionary learning, non-negative tensor decomposition [XY13].

In its generic form, the BSUM covers several other well-known optimization methods, such as the convex-concave procedure (for optimizing difference of convex functions), the majorization minimization (e.g. the expectation minimization algorithm), the proximal point algorithm, the forward-backward splitting algorithm, the non-negative matrix factorization problem, and the re-weighted least-squares problem [HRLP16]. More recently, the BSUM framework has found application in several areas of bio-informatics such as DNA sequencing and tensor decomposition (for clustering and compression).

It is clear at this point that methods such as BCD and BSUM are extremely effective for tackling optimization problems such as (P) in (1.1.1), where the objective $f(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ is coupled in the variables. However, they are less effective when tackling problems such as (P) in (1.1.1), where the constraints are coupled: when non-separable constraints are present, i.e., $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) \in \mathcal{C}$, no BCD convergence results exist.

1.1.2 Lagrangian Techniques

The Complex Gradient: Though there are many ways to define complex derivatives, we follow the most widely adopted one, first outlined in [Bra83]. Let $f(\boldsymbol{X})$:

$\mathbb{C}^{p\times q} \to \mathbb{R}$ be a real-value matrix function of $\boldsymbol{X} \in \mathbb{C}^{p\times q}$, that is differentiable. Then, the complex gradient operator, $\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) : \mathbb{C}^{p\times q} \to \mathbb{R}$, is defined as,

$$[\nabla_{\boldsymbol{X}} f(\boldsymbol{X})]_{(k,l)} = \frac{1}{2}\left(\frac{\partial f}{\Re[\boldsymbol{X}_{(k,l)}]} + j\frac{\partial f}{\Im[\boldsymbol{x}_{(k,l)}]}\right),\ \forall (k,l) \in \{p\}\times\{q\} \tag{1.1.4}$$

Thus, $\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) = 0$ is necessary and sufficient to find stationary points of f. Under this definition, one can verify for instance that, $\nabla_{\boldsymbol{X}}\ \mathrm{tr}(\boldsymbol{X}^\dagger \boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X}$, and $\nabla_{\boldsymbol{X}}\ \log|\boldsymbol{I} + \boldsymbol{X}^\dagger\boldsymbol{A}\boldsymbol{X}| = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{I} + \boldsymbol{X}^\dagger\boldsymbol{A}\boldsymbol{X})^{-1}$.

KKT conditions for convex problems Lagrangian techniques, based on the Karush-Kuhn-Tucker (KKT) conditions, are the most fundamental tools in convex optimization. The standard form is often given in the context of scalar/vector optimization. However, as most of the thesis will deal with optimization problems with matrix functions, we shall give the standard (matrix) form for convex optimization problems:

$$\begin{cases} \min\limits_{\boldsymbol{X}\in\mathbb{C}^{p\times q}} f(\boldsymbol{X}) \\ \text{s. t. } g_i(\boldsymbol{X}) \le 0, \forall i \in \{m\}, \\ \qquad h_j(\boldsymbol{X}) = 0, \forall i \in \{n\} \end{cases} \tag{1.1.5}$$

where $f : \mathbb{C}^{p\times q} \to \mathbb{R}$ is convex and differentiable, $g_i : \mathbb{C}^{p\times q} \to \mathbb{R},\ \forall\ i \in \{m\}$ are convex and differentiable, and $h_j : \mathbb{C}^{p\times q} \to \mathbb{R},\ \forall\ j \in \{n\}$ are affine.

Let $\boldsymbol{X}^\star$ be the optimal primal value, and $\{\lambda_i^\star\}, \{\mu_j^\star\}$ the optimal dual Lagrange multipliers. The KKT conditions are written as follows [BV04, Sect 5.5]:

$$\begin{cases} \nabla f(\boldsymbol{X}^\star) + \sum_i \lambda_i^\star \nabla g_i(\boldsymbol{X}^\star) + \sum_j \mu_j^\star \nabla h_i(\boldsymbol{X}^\star) \\ g_i(\boldsymbol{X}^\star) \le 0, \forall i \in \{m\}, \quad h_j(\boldsymbol{X}^\star) = 0,\ \forall i \in \{n\} \\ \lambda_i^\star g_i(\boldsymbol{X}^\star) = 0,\ \forall i \in \{m\} \\ \lambda_i^\star \ge 0,\ \forall i \in \{m\} \quad , \quad \mu_j^\star \ne 0,\ \forall j \in \{n\} \end{cases} \tag{1.1.6}$$

where the gradient $\nabla f(\boldsymbol{X})$ follows the above definition. When the primal problem is convex (i.e., $f, g_1, ..., g_m$ are convex, and $h_1, ..., h_n$ are linear/affine), and *strong duality* holds, then the KKT condition are *necessary* and *sufficient* [BV04, Sect 5.5].

Remark 1.1. The KKT conditions are *necessary* conditions for optimality when the problem is not convex (i.e., $f, g_1, ..., g_m, h_1, ..., h_n$ are differentiable, $f, g_1, ..., g_m$ not necessarily convex, $h_1, ..., h_n$ not affine), but where strong duality holds.

Applications: Standard Lagrangian techniques were essential to tackling classical problems such as the minimum mean-squared error estimation [TV04], the optimal single-user MIMO precoder (waterfilling solution) [TV04], the optimal precoder

in multi-user MIMO (iterative waterfilling [YRBC04]), etc. However, in recent years very few problems in (modern) signal processing and communication can be formulated as (1.1.5). However, formulations such as the one above are still very common when one is dealing with practical problems. For instance, consider problems of the form,

$$\min_{\boldsymbol{X},\boldsymbol{Y}} f(\boldsymbol{X},\boldsymbol{Y}) \;\; \text{s. t. } \boldsymbol{X} \in \mathcal{X},\; \boldsymbol{Y} \in \mathcal{Y} \tag{1.1.7}$$

where $f(\boldsymbol{X},\boldsymbol{Y})$ is not jointly convex in all $\boldsymbol{X}$ and $\boldsymbol{Y}$. As mentioned earlier, such problems are ideal candidates for the BCD method: fix $\boldsymbol{Y}$ and optimize for $\boldsymbol{X}$ (first subproblem), then fix $\boldsymbol{X}$ and optimize $\boldsymbol{Y}$ (second subproblem), iteratively. Most often, each of the subproblems satisfies the standard form in (1.1.5).

This is the case for a significant fraction of the algorithms for distributed cellular coordination, namely, interference leakage minimization [GCJ11, pet09], minimum mean-squared error minimization (MMSE) [SSB$^+$09, PH11], and weighted minimum mean-squared error minimization [SRLH11]. In such cases, $\boldsymbol{X}$ and $\boldsymbol{Y}$ represent the block of transmit and receive filters, respectively. The application of the BCD algorithm to distributed coordination in cellular networks, is discussed at length in Chap. 4.4.

1.1.3 Dantzig-Wolfe Decomposition for Integer Programs

Since its inception in the seminal work of P. Wolfe and G. Dantzig, the Dantzig-Wolfe (DW) decomposition [GBD60a] has been widely adopted, for obtaining lower bounds on Integer Programs (IPs).

Mathematical Formulation Consider the following IP,

$$(P):\; \boldsymbol{x}^\star \begin{cases} \text{argmin } f(\boldsymbol{x}) \\ \text{s. t. } \boldsymbol{x} \in \mathcal{S},\; \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{c} \end{cases} \tag{1.1.8}$$

where f is a continuous function (possibly non-convex), $\mathcal{S}$ a finite set corresponding to integer constraints, and let $\{\boldsymbol{\psi}_j\}_{j=1}^J$ be the set of vertexes for $\mathcal{S}$. The DW decomposition then proceeds by relaxing the integer constraint, i.e., $\boldsymbol{x} \in \mathcal{S}$, into a convex one by taking its convex hull, i.e., $\boldsymbol{x} \in \text{conv}(\mathcal{S})$. As a result, every point in $\text{conv}(\mathcal{S})$ is represented as a *convex combination* of the *vertexes* of $\mathcal{S}$, i.e.,

$$\boldsymbol{x} \in \text{conv}(\mathcal{S}) = \left\{ \sum_{j=1}^J w_j \boldsymbol{\psi}_j \;\middle|\; \sum_j w_j = 1,\; w_j \geq 0, \forall j \right\} \tag{1.1.9}$$

$$= \left\{ \sum_{j=1}^J w_j \boldsymbol{\psi}_j \;\middle|\; \mathbf{1}_J^T \boldsymbol{w} = 1,\;\; \boldsymbol{w} \geq \mathbf{0}_J \right\} \tag{1.1.10}$$

The DW decomposition is seen as a mapping for $\boldsymbol{x}$ to $\boldsymbol{w}$ (given by the above equation). By letting $\alpha_j = f(\boldsymbol{\psi}_j)$, the cost function in (P) is equivalent to $\sum_{j=1}^{J} w_j \alpha_j$. Moreover, the linear constraint in (P) can be rewritten as,

$$\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{c} \Leftrightarrow \sum_{j=1}^{J} w_j (\boldsymbol{A}\boldsymbol{\psi}_j) \leq \boldsymbol{c} \Leftrightarrow \boldsymbol{\Gamma}\boldsymbol{w} \leq \boldsymbol{c} \tag{1.1.11}$$

$$\text{where } \boldsymbol{\Gamma} \triangleq [\boldsymbol{A}\boldsymbol{\psi}_1, ..., \boldsymbol{A}\boldsymbol{\psi}_J],\ \boldsymbol{w} = [w_1, ..., w_J]^T \tag{1.1.12}$$

Then, the DW decomposition associated with (P) is given by the following linear program

$$(P_{DW}):\ \boldsymbol{w}^\star \begin{cases} \underset{\boldsymbol{w}\in\mathbb{R}^J}{\text{argmin}}\ f_{DW}(\boldsymbol{w}) \triangleq \boldsymbol{\alpha}^T\boldsymbol{w} \\ \text{s. t.} \quad \boldsymbol{w} \geq \boldsymbol{0}_J,\ \boldsymbol{\Gamma}\boldsymbol{w} \leq \boldsymbol{c} \end{cases} \tag{1.1.13}$$

It then straightforward to show that

$$f(\boldsymbol{x}^\star) \geq f_{DW}(\boldsymbol{w}^\star),$$

thereby implying that the DW always provides a lower bound on (P). A look at the above problem immediately reveals the power of the DW decomposition: despite the combinatorial and non-convex nature of (P), the DW always results in a linear program.

Applications: The DW decomposition is a wide-spread systematic tool, for lower bounding integer programming problems. Though originally developed for linear integer programs [GBD60a], it was later extended to arbitrary integer programs [BJN$^+$96]. In the context of operations research, the DW decomposition is the most common tool for tackling problems such as the (generalized) assignment problem: There are a number of agents and a number of tasks. Any agent can be assigned to perform any task. Moreover, each agent has a budget and the sum of the costs of tasks assigned to it cannot exceed this budget. It is required to find an assignment in which all agents do not exceed their budget and total cost of the assignment is minimized. The generalized assignment problem is tightly related to the knapsack problem. We apply the DW to lower bound a variant of this problem, in Chap. 7, where the above cost function is replaced with a non-linear one.

1.2 Background and Motivation

Wireless communication is a vital component underlying most modern technological aspects in our society. Technologies such as mobile cellular access, device-to-device communication, machine-type communication, cyber-physical systems, wireless control, voice/video streaming services, the Internet of things, tactile Internet, etc, are contingent on reliable operation of wireless devices such as mobile

phones/tablets/laptops. For such reasons, wireless communication systems have permeated a huge number of standards such as 3G/4G/LTE, WiFi (IEEE 802.11 family), Bluetooth, ZigBee, etc. In most of this thesis however, the focus is put on cellular networks.

The targeted data rates for consecutive generations of cellular networks have been drastically increasing: around 100 Kbps for 2G, around 2 Mbps for 3G, around 200 Mbps for 4G, and greater than 1 Gbps for 5G. Moreover, it is estimated that the mobile data consumption (e.g. by smart phones, tablets, mobile PC) is expected to increase by a factor of 10, between 2015 and 2021 [Eri15]. Moreover, a total of 28 billion connected mobiles devices are expected across the world by 2021. This *exponential increase* in demand for data is also agreed upon (to some extent) by most major mobile operators. Thus, communication engineers have the monumental task of designing future cellular networks that are able to deliver unprecedented data rates. From a historical perspective, the increase in data rates for cellular systems over the last decades, can be broken down into three major categories [DHL+11]:

A1) increases in *spectral efficiency*

A2) exploiting *more spectrum*

A3) gains from higher *densification*

We underline the fact that the above categorization is exactly mirrored in the 5G system requirements.

The EU project METIS is one of the few efforts offering insights into the possible requirements of 5G systems [MET14]: concepts such as *goals for 5G systems*, as well as the most common *test cases* (each relating to specific aspects of 5G systems), are clearly defined. From the perspective of this thesis, the focus is one on test cases relating to ultra-dense networks, as well as the goals concerned with increasing data-rates.

5G Goals:

- 1000x data volume
- **10-100x user data rate**
- 10-100x number of devices
- 10-longer battery life
- 5x reduced end-to-end latency

Test cases related to ultra-dense networks:

- Virtual reality office
- Dense urban information society
- Shopping mall
- Stadium
- Open air festival

The project is clear on the mapping between the above test cases, and the '10-100x user data rate' goal: this is achieved via *densification*, *improved efficiency* and

new spectrum opportunities [MET14]. Note that this exactly corresponds to the above categorization, A1)-A3).

That being said, in this thesis, we attempt to address all three, wherein each part of the thesis will aim at addressing one of the above paradigms. Moreover, the individual parts will be essentially self-contained.

1.2.1 Part I: Exploiting more spectrum via Millimeter-wave Communication

Communication in the millimeter-wave band is one of the most promising solutions to the ever increasing demand for higher data rates. Such system are extremely attractive from that aspect, since they promise to deliver at least 10 times more spectrum (up to 200 times) over cellular systems in conventional bands [RSM+13]. Thus, one can see how mmWave communication with MIMO capabilities, is an enabler for multi-Gpbs speeds, required for 5G systems. Firstly, note that at mmWave frequencies, the required size and spacing of antennas is quite small. In addition to the orders-of-magnitude larger spectrum, the many transmit and receive antennas, operating at mmWave frequencies, result in arrays with larger number of antennas, and narrow beams. This is turns leads to reduced interference, high array gain at the transmitter and receiver (due many antennas), and better spectrum reuse (due to high pathloss).

In the scope of 5G systems, there is no specific allocation of spectrum for mmWave bands, yet. However, there are several strong candidates:[1]

(B_1) 28 GHz: bandwidth of 1.3 GHz

(B_2) 39 GHz: bandwidth of 1.4 GHz

(B_3) 37 and 42 GHz: bandwidth of 2.1 GHz

(B_4) $71-76$ and $81-86$ GHz: bandwidth of 10 GHz

In conjunction with MIMO transmission, its is envisioned that spatial multiplexing will be used in (B_1), (B_2) and (B_3), while beamforming will be used in (B_4). With that in mind, this thesis will provide insights for future standardization effort about mmWave communication systems, for both 5G systems and IEEE 802.11ad (Gbps WLAN).

However, MIMO communication in the millimeter-wave band comes with several inherent challenges - that are in part addressed in this thesis (Part I), namely, the high pathloss attenuation, channel modeling, channel estimation, precoder design, etc. The latter topics are still active areas of mmWave research.The background and motivation for mmWave MIMO systems are discussed at length, in Chap. 2.

[1] Based on the following presentation: Robert Heath - "Millimeter-wave MIMO as the future of 5G", 2015

1.2.2 Part II: Increasing spectral efficiency via Distributed Coordination

It has been known for the past years that coordination in cellular systems, results in increased spectral efficiency. From a theoretical perspective, concepts such Interference Alignment [CJ08], Coordinated Multi-point (CoMP) [GHH+10], as well as their numerous variants, are known to increase the spectral efficiency: in fact, in some specific scenarios, they are known to achieve theoretic bounds. However, several approaches that fall under that category require significant overhead, e.g. global CSI at the BSs, thereby making them unfit for operating in cellular networks. Thus, in this thesis, we advocate coordination via distributed algorithms, where each BS/MS is only required to have local CSI. While this has been a significant area of research, the entirety of the proposed approaches have paid little-to-no attention to the high associated communication overhead. That being said, a major contribution of Part II in this thesis is to proposing algorithms with improved convergence properties.

Distributed multi-cellular coordination in LTE standards is known under the name of Coordinated Multi-point (CoMP). It essentially consists of several base stations sharing CSI of their respective users (and potentially data of users to be served) to manage interference - a thorough description of the mechanism behind CoMP is done in Chapter 4.1.1. Moreover, since CoMP is operating in the context of cellular networks there is a stringent requirement on the communication overhead associated with the exchange of CSI (and potentially data): the size of the required backhaul link (among different BSs) grows accordingly. In view of mitigating the need for explicit exchange of CSI over a dedicated backhaul link, algorithms for distributed CSI acquisition are generally considered instead. This is detailed in Chap. 4.3.2.

CoMP is incorporated as an integral part of LTE Advanced as an effective mechanism for dynamic coordination of transmission and reception over a multiple base stations: it results in improved overall quality for the user, as well as better utilization of the network. CoMP was also included as a vital component in 3rd Generation Partnership Project (3GPP). CoMP-like ideas, i.e., coordination among multiple transmit/receive nodes, are also central to other standards such as IEEE WiMAX. With that in mind, CoMP has been identified as an essential technique for achieving the spectral efficiency specified by 3GPP and LTE-Advanced standard.

However, the importance of CoMP is much more pronounced in future cellular systems. Ultra-Dense Deployments (UDN) are identified as one of the most common operation modes for 5G systems, wherein the density of access of nodes is orders of magnitude higher than in current deployments [OBB+14]. Since interference is known to be the bottleneck on achieving the optimal performance in such dense settings, this inevitably raises the issue of effective management of the resulting interference. Needless to say, CoMP-like coordination will be a critical component in 5G systems.

We underline the fact that the framework and approaches presented in Part II

of the thesis, have been developed as technical components (TeC) of the 'Advanced inter-node coordination techniques' (Task 3.2), under the project EU-FP7 METIS2020. Moreover, the aforementioned techniques have been tested and evaluated (against LTE-based benchmarks), within a 'proper' 5G simulation setup. We refer interested readers to [MET15], where such issues are discussed in full details. Since the latter project is essentially a pre-draft for 5G standards, the techniques proposed in this thesis will provide insights into the standardization of 5G systems.

1.2.3 Part III: Higher Densification via Cloud Radio Access Networks

In the context of cellular networks, distributed optimization algorithms have been prevalent in the last decade. However, in more recent years, there has been increasing interest in the reverse side of the coin, i.e., coordination schemes of a centralized nature. Such schemes are applied in the context of Cloud Radio Access Network: Schemes falling under this category gather all the required CSI at one 'aggregation node', run the coordination algorithm in question, and propagate the results to each base station (and user). We will investigate this approach in Part III.

From a historical perspective, the most significant fraction of the gains in data rates are due to densification [DHL+11]: The small cells resulting from increased levels of densification allows for higher *reuse factor*. Moreover, the insights gained from *CoMP* [GHH+10, BJBO11] and *IA* [CJ08, MAMK08] show that coordination among (clusters of) base stations is a key to achieving higher sum-rate in the network. However, in traditional cellular networks, the communication overhead associated with such techniques has been identified as a (potentially) limiting factor of the sum-rate gains (e.g., [EALH12, EALH11, PH12, LHA13]). In contrast, due to its inherent centralized nature, *Cloud Radio Access Network (Cloud-RAN)* enables the tight coordination of antennas in a dense deployment in a rather economical way.

1.3 Thesis Scope and Contributions

Part I: Chapters 2 and 3

Chapters 2 and 3 mainly address the problem of channel estimation and and precoding for hybrid analog-digital millimeter-wave MIMO systems: while Chapter 2 focuses on studying the optimal precoder structure, under perfect CSI, Chapter 3 proposes practical algorithms for estimating the dominant subspace of the channel, and the design of analog/digital precoders and combiners accordingly.

In Chapter 2, we motivate the *hybrid precoding* architecture - where both precoding and combining are done in two stages, *analog and digital*, as the a promising candidate for mmWave MIMO systems: it offers the best trade-off between classical fully digital solutions (that require high power consumption and complexity),

and fully analog solutions (that are inherently limited to beamforming only). After surveying related prior work, we characterize the *optimal hybrid precoder* (resp. combiner) as the right (resp. left) singular vectors of the channel - similarly to the conventional MIMO case. The metric under consideration is the user rate.

After characterizing the optimal precoding strategy, we tackle in Chapter 3 the problem of channel estimation and precoding in hybrid millimeter-wave MIMO systems. For that matter, we propose a method based on the well-known *Arnoldi iteration* exploiting channel reciprocity in TDD systems and the *sparsity* of the channel's eigenmodes to estimate the right (resp. left) singular subspaces of the channel at the BS (resp. MS). We first describe the algorithm in the context of conventional MIMO systems, and derive bounds on the estimation error in the presence of distortions at both BS and MS. We later identify obstacles that hinder the application of such an algorithm to the hybrid analog-digital architecture, and address them individually. In view of fulfilling the constraints imposed by the hybrid analog-digital architecture, we further propose an iterative algorithm for subspace decomposition, whereby the above estimated subspaces are approximated by a cascade of analog and digital precoder/combiner. Finally, we evaluate the performance of our scheme against the perfect CSI, fully digital case (i.e., an equivalent conventional MIMO system), and conclude that similar performance can be achieved, especially at medium-to-high SNR (where the performance gap is less than 5%), however, with a drastically lower number of RF chains ($\sim$ 4 to 8 times less).

The contributions of the thesis for Part I are shown below.

- [GKBS16a] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace Estimation and Decomposition for Large Millimeter-Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528-542, Apr. 2016

- [GBKS15] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimeter-wave MIMO systems,", *in 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*

In addition to the above contributions, the author of the thesis also took part in the following work, related to mmWave MIMO systems. Though not explicitly included as contributions of the thesis, they are still related to this part.

- [CKGB16] W. M. Chan, T. Kim, H. Ghauch, and M. Bengtsson, "Subspace estimation and hybrid precoding for wideband millimeter-wave MIMO systems", *invited paper, IEEE ASILOMAR, Nov. 2016.*

- [HKG$^+$14] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter-wave MIMO channel tracking systems", *in 2014 IEEE GLOBECOM Workshops(GC Workshops), 2014*

PART II: Chapter 4 , Chapter 5, and Chapter 6

In this part of the thesis, we use the so-called framework of *distributed sum-utility optimization*. More specifically, Chapters 5 and Chapter 6, are special instances of it.

We define this framework in Chapter 4 as a generic method for the (joint) design of transmit and receive filters, in cellular network, and briefly describe its operation. After surveying modern techniques for interference management (e.g., interference alignment, coordinated multi-point), we describe the so-called *forward-backward iterations* - which is used in almost all distributed coordination algorithms developed in the last decade. We argue that in the the context of cellular networks, only a low number of forward-backward iterations can be carried out: despite the plethora of algorithms proposed under the umbrella of forward-backward iterations, virtually no work focused on algorithms that operate in the low-overhead regime. In that sense, the approaches presented in Chapters 5 and 6, are special cases of the aforementioned framework, where the aim is to design *fast-converging low-overhead algorithms* for distributed network-utility maximization.

In Chapter 5, after lower bounding the sum-rate using a so-called *DLT bound* (i.e., a difference of log and trace), we underline a major advantage of using such a bound: it leads to separable convex subproblems that naturally decouple at both the transmitters and receivers. Moreover, we derive the solution to the latter subproblem, that we dub *non-homogeneous waterfilling* (a variation on the MIMO waterfilling solution), and underline an inherent desirable feature: its ability to turn off streams exhibiting low-SINR, thereby greatly speeding up the convergence of the proposed algorithm. This *stream-control* feature is at the basis for the fast converging nature of the algorithm. We then show the convergence of the resulting algorithm to a stationary point of the DLT bound (a lower bound on the sum-rate). Finally, we rely on extensive simulations of various network configurations, to establish the superior performance of our proposed schemes, with respect to other state-of-the-art methods.

In Chapter 6, we propose low-overhead fast-converging algorithms, using the *interference leakage* as metric. For that purpose, we relax the well-known leakage minimization problem, and then propose two different filter update structures to solve the resulting non-convex problem: though one leads to conventional *full-rank filters*, the other results in *rank-deficient filters*, that we exploit to gradually reduce the transmit and receive filter rank, and greatly speed up the convergence. Furthermore, inspired from the decoding of turbo codes, we propose a *turbo-like* structure to the algorithms, where a separate inner optimization loop is run at each receiver, in addition to the main forward-backward iteration. In that sense, the introduction of this turbo-like structure converts the communication overhead required by conventional methods to computational overhead at each receiver - a cheap resource, allowing us to achieve the desired performance, under a minimal overhead constraint. Finally, we show through comprehensive simulations that both proposed schemes significantly outperform the relevant benchmarks, especially for

large system dimensions.

The contributions of the thesis for Part II are summarized below.

- [GKBS15] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Distributed Low-overhead Schemes for Multi-stream MIMO Interference Channels," *IEEE Transactions on Signal Processing*, vol. 63 no. 7, pp. 1737-1749, April 2015

- [GKBS16b] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Separability and Sum-rate Maximization in MIMO Interfering Networks," *IEEE Transactions on Signal and Information Processing over Networks (in revision, submitted Jun 2016)* ,

In addition to the above contributions, the author of the thesis also took part in the following works, relating to distributed coordination. Though not explicitly included as contributions of the thesis, they are still related to this part.

- [GMBS15] H. Ghauch, R. Mochaourab, M. Bengtsson, and M. Skoglund, "Distributed precoding and user selection in MIMO interfering networks," *in Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015*

- [MBGB15] R. Mochaourab, R. Brandt, H. Ghauch, and M. Bengtsson, "Overhead-aware distributed CSI selection in the MIMO interference channel,", *in 23rd European Signal Processing Conference (EUSIPCO), 2015*

Part III: Chapter 7 and Chapter 8

Part III addresses the problem of coordination in cellular networks, from the opposite perspective as that of Part II, by looking at centralized coordination in general, i.e., Cloud-RAN. More specifically, in Chapter 7, we introduce the so-called Antenna Domain Formation problem, focusing on its theoretical aspects.

In Chapter 7, we motivate the densification paradigm in cellular networks as the one that has brought about most gains in data rates. We then argue that Cloud-RAN is a promising candidate architecture that aims at effectively managing the densely deployed remote radio-heads in an economical way: sets of coordinating *radio-heads* are thus connected to a central aggregation node, and dubbed as an *antenna domain*. Each aggregation node gathers all the CSI and/or data, performs the required processing (e.g., computing precoder at each radio-head), and propagates the resulting setting to individual radio-heads. We motivate the so-called *Antenna Domain Formation (ADF)* problem, as the optimal assignment of users to antenna domains, in Cloud-RAN systems, and formulate it as an *integer optimization* problem.

After formulating the ADF problem, we focus in Chapter 8 on theoretical aspects of the problem, namely, finding lower bounds on the interference leakage in the network. We first propose a simple iterative algorithm for obtaining a solution.

Then, motivated by the lack of optimality guarantees on such solutions, we opt to find *lower bounds* on the problem, and the resulting interference leakage in the network, by deriving two different ones: The *Dantzig-Wolfe decomposition* corresponding to the ADF problem, and the dual problem. Moreover, we show that the former offers a tighter bound than the latter. We highlight the fact that the bounds in question consist of linear problems with an exponential number of variables in the total number of users, and adapt known methods aimed at solving them. In addition to shedding light on the tightness of the bounds in question, our numerical results show sum-rate gains of at least 200%, over a simple benchmark, in the medium SNR region.

The contributions of the thesis for Part III are shown below.

- [GRI+16] H. Ghauch, M. Mahboob Ur Rahman, S. Imtiaz, J. Gross, M. Skoglund, and C. Qvarfordt "Performance Bounds for Antenna Domain System," *IEEE Transactions on Wireless Communications (submitted, Jun 2016)*

- [GRIG16] H. Ghauch, M. Mahboob Ur Rahman, S. Imtiaz, J. Gross, "Coordination and Antenna Domain Formation in Cloud RAN systems," *in 2016 IEEE International Conference on Communication (ICC)*

In addition to the above contributions, the author of the thesis also took part in the following works, that fall under the umbrella of centralized coordination. Though not explicitly included as contributions of the thesis, they are still related to this part.

- [RGIG15] M. Mahboob Ur Rahman, H. Ghauch, S. Imtiaz, J. Gross, "RRH clustering and transmit precoding for interference-limited 5G CRAN downlink," *in 2015 IEEE GLOBECOM Workshops(GC Workshops), 2014*

- [GP11] H. Ghauch and C.B. Papadias, "Interference alignment: A one-sided approach," *in 2011 IEEE Global Communications Conference (GLOBECOM 2011)*

- [GKBS13] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Interference alignment via controlled perturbations," *in 2013 IEEE Global Communications Conference (GLOBECOM 2013)*

Copyright Notice

Part I

Millimeter-Wave MIMO systems

Chapter 2

Optimal Precoding in Hybrid MIMO systems

Millimetre wave (mmW) communication systems have the distinct advantage of exploiting the *huge amounts of unused (and possibly unlicensed) spectrum* in those bands - around 200 times more than conventional cellular systems [1]. Moreover, the corresponding antennae size and spacing become small enough, such that *tens-to-hundreds of antennas can be fitted on conventional hand-held devices*, thereby enabling gigabit-per-second communication. However, the large number of radio frequency (RF) chains required to drive the increasing number of antennas, inevitably incurs a tremendous increase in power consumption (namely by the analog-to-digital converters), as well as added hardware cost. One elegant and promising solution to remedy this inherent problem is to offload part of the precoding/processing to the *analog domain*, via analog precoding (resp.combining), i.e., a network of phase shifters to linearly process the signal at the BS (resp. MS). The system model under consideration in shown Fig. 2.1. This so-called problem of *analog and digital co-design* for beamforming and precoding in low-frequency regime was first investigated in [ZMK05, VvdV10]. This architecture was later studied within the context of higher frequency (mmWave) systems in [EARAS$^+$14, AEALH14, NBH10] - under the name of *hybrid precoding/architecture* - for the precoding problem. A similar setup for the case of beamforming was considered in [TPA11, WLP$^+$09, HKL$^+$13].

The hybrid analog-digital architecture has several salient features.

- *Highly directional* channels and propagation. Thus, the channel consists of a relatively small number of paths, and beamforming is highly selective.
- *Severe attenuation* due to the atmospheric absorption peak at 60 GHz, is an inherent feature for mmWave communication systems. In the the hybrid architecture, the severe pathloss is mitigated by having large number of transmit

[1] Early results on the design of communication systems in the millimetre wave (mmW) spectrum date back to [OMM$^+$00, OMI$^+$03], but have started receiving growing interest over the past few years.

and receive antennas: thus one can achieve high enough *array gain* to compensate for the small signal power due to attenuation.

- *High number of transmit/receive antennas* is facilitated in the hybrid architecture. This is due to the fact that antenna sizes at 60 GHz are quite small. Moreover, in contrast to conventional MIMO systems, the number of required analog-to-digital converters is a fraction of the number of transmit/receive antennas. Thus, the resulting power consumption is not a limiting factor for scaling up the number of antennas.

Several fundamental challenges have to resolved before any of the promised gains can be harnessed.

- *Channel estimation* for the (large) mmWave MIMO channel is one of the major obstacles. We underline the fact that classical training schemes developed for MIMO systems are not applicable to the hybrid analog-digital architecture, since the resulting overhead would be tremendously high (discussed in detail in Remark 3.5). While few works focused on the channel estimation, authors in [WLP+09,HKL+13,AEALH14] proposed schemes based on sounding of hierarchical codebooks. Moreover, in [GBKS15,GKBS16a], we proposed algorithms that estimate the dominant subspace of the channel, using the well known Arnoldi Iteration.

- *Hybrid precoding,* wherein the analog/digital precoder and combiner are designed based on the channel. Variations on the well-known Orthogonal Matching Pursuit (OMP) were proposed in [EARAS+14, AHAS+12, MRRGPH15], where the columns of the analog precoder / combiner are greedily designed. The authors in [SY15a] obtained upper and lower bounds on the minimum number of transmit and receive RF chains that are required to realize the theoretical capacity. Later on, designs based on heuristics for maximizing the rate, were proposed in [SY15b] initially, and later extended in [SY16], where the authors show optimality if the number of data streams in equal to the number of RF chains. Finally, in our work [GBKS15, GKBS16a], we approximated the optimal precoder/combiner by proposed methods based on Block-Coordinated Descent.

- *Open problems* in (hybrid) mmWave MIMO systems include (but not limited to), the widespread adoption of a statistical channel model (i.e., the analog of Rayleigh fading in conventional MIMO). Moreover, research on fundamental aspects of hybrid MIMO systems, i.e., channel capacity and achievable rates, is still not present.

Our work in this thesis falls under both channel estimation, and hybrid precoding. After a series of approximations to the mutual information, and taking into account precoding (excluding the receive combiners), [EARAS+14] derived an

optimality condition relating the analog and digital precoders to the optimal unconstrained precoder (i.e., the right singular vectors of the channel), by assuming *full channel state information (CSI) at both the BS and MS*. This assumption was later relaxed in [AEALH14] where an algorithm for estimating the dominant propagation paths was proposed, based on the previously proposed concept of *hierarchical codebooks sounding* in [WLP+09, HKL+13]. However, the algorithm requires *a priori knowledge of the number of propagation paths* (i.e. the propagation environment), its *performance is affected by the sparsity level of the channel*, and exhibits relatively elevated complexity. Finally, it appears rather inefficient to estimate the entire channel, while only a few eigenmodes are needed for transmission: this is particularly relevant in mmWave MIMO channels, since the majority of eigenmodes have negligible power.

In this chapter, we characterize the optimal precoding strategy for a single user hybrid analog-digital MIMO link (assuming perfect CSI at the transmitter and receiver): it is aligned with the dominant subspace of the MIMO channel. The approach we present here attempts to address the above limitations (presented in the next chapter). The proposed algorithm is based on the well known *Arnoldi Iteration*, exploits channel reciprocity inherent in TDD MIMO systems to gradually build an orthonormal basis for the corresponding Krylov subspace, and *directly estimates the dominant left / right singular modes of the channel, rather than the entire channel*. We then propose an iterative method for subspace decomposition, to *approximate the estimated right (resp. left) singular subspace by a cascade of analog and digital precoder (resp. combiner)*, while taking into account the hardware constraints of this so-called hybrid analog-digital architecture. The subspace estimation (SE) algorithm is based on *BS-initiated echoing*, whereby the BS sends along some beamforming vector, and the MS echoes its received signal back to the BS (using amplify-and-forward), thereby enabling the BS to obtain an estimate of the effective uplink-downlink channel. We first detail the algorithm in the context of conventional MIMO, taking into account distortions in the the system (e.g., noise, or other disturbances), derive bounds on the estimation error, and highlight its desirable features. We then adapt its structure, to fit the many operational constraints dictated by the hybrid analog-digital architecture. While we feel that aspects such as complexity, overhead and numerical stability are best left for future works, we do shed light on each of them.

2.1 System Model

2.1.1 Channel Model

We adopt the prevalent physical representation of sparse mmWave channels adopted in the literature, e.g., [AEALH14, EARAS+14], where only L scatterers are assumed to contribute to the received signal at the MS - an inherent property of the poor

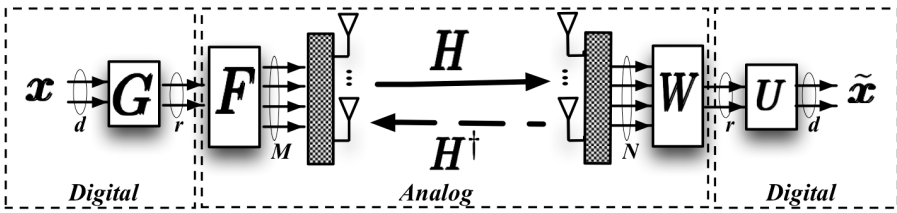

Figure 2.1: Hybrid Analog-Digital MIMO system architecture

scattering nature in mmWave channels,

$$\boldsymbol{H} = \sqrt{\frac{MN}{L}} \sum_{i=1}^{L} \beta_i \, \boldsymbol{a}_r(\chi_i^{(r)}) \boldsymbol{a}_t^{\dagger}(\chi_i^{(t)}) \tag{2.1.1}$$

where $\chi_i^{(r)}$ and $\chi_i^{(t)}$ are angles of arrival at the MS, and angles of departure at the BS (AoA/AoD) of the i^{th} path, respectively (both assumed to be uniform over $[-\pi/2,\ \pi/2]$), β_i is the complex gain of the i^{th} path such that $\beta_i \sim \mathcal{CN}(0,1),\ \forall i$. Finally, $\boldsymbol{a}_r(\chi_i^{(r)})$ and $\boldsymbol{a}_t(\chi_i^{(t)})$ are the array response vectors at both the MS and BS, respectively. For simplicity, we will use uniform linear arrays (ULAs), where we assume that the inter-element spacing is equal to half of the wavelength. We also assume a TDD system, where channel reciprocity holds. Finally, we denote the SVD of $\boldsymbol{H}$ as,

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{\Phi}_1, & \boldsymbol{\Phi}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_1^{\dagger} \\ \boldsymbol{\Gamma}_2^{\dagger} \end{bmatrix} = \boldsymbol{\Phi}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^{\dagger} + \boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^{\dagger} \tag{2.1.2}$$

where $\boldsymbol{\Gamma}_1 \in \mathbb{C}^{M \times d}$ and $\boldsymbol{\Phi}_1 \in \mathbb{C}^{N \times d}$ are semi-unitary, and $\boldsymbol{\Sigma}_1 = \text{diag}(\sigma_1, ..., \sigma_d)$ is diagonal with the d largest singular values of $\boldsymbol{H}$ (in decreasing order).

2.1.2 Signal Model

We assume a single user MIMO system with M and N antennas at the BS and MS, respectively, where each is equipped with r RF chains, and sends d independent data streams (where we assume that $d \leq r \leq \min(M, N)$). The downlink (DL) received signal at the MS is given by,

$$\boldsymbol{y}^{(r)} = \boldsymbol{HFG}\boldsymbol{x}^{(t)} + \boldsymbol{n}^{(r)} \tag{2.1.3}$$

where $\boldsymbol{H} \in \mathbb{C}^{N \times M}$ is the complex channel - assumed to be slowly block-fading, $\boldsymbol{F} \in \mathbb{C}^{M \times r}$ is the analog precoder, $\boldsymbol{G} \in \mathbb{C}^{r \times d}$ the digital precoder, $\boldsymbol{y}^{(r)}$ the N-dimensional signal at the MS antennas, $\boldsymbol{x}^{(t)}$ is the d-dimensional transmit signal (at the BS) with covariance matrix $E[\boldsymbol{x}^{(t)}\boldsymbol{x}^{(t)^{\dagger}}] = \boldsymbol{I}_d$ and $\boldsymbol{n}^{(r)}$ is the AWGN noise at the

MS, with $E[\boldsymbol{n}^{(r)}\boldsymbol{n}^{(r)^\dagger}] = \sigma^2_{(r)}\boldsymbol{I}_N$. Note that $^{(t)}$ and $^{(r)}$ subscripts/superscripts denote quantities at the BS and MS, respectively. Both the analog precoder and combiner are constrained to have constant modulus elements (since the latter represent phase shifters), i.e., $\boldsymbol{F} \in \mathcal{S}_{M,r}$ and $\boldsymbol{W} \in \mathcal{S}_{N,r}$ (also referred to as the *constant-modulus* or *constant-envelope* constraint). We adopt a *total power constraint* on the effective precoder, i.e., $\|\boldsymbol{FG}\|_F^2 \leq d$, a widespread one in the hybrid analog-digital precoding literature [EARAS+14, AEALH14].

With that in mind, the received signal after filtering in the DL is given as,

$$\tilde{\boldsymbol{x}}^{(r)} = \boldsymbol{U}^\dagger\boldsymbol{W}^\dagger\boldsymbol{y}^{(r)} = \boldsymbol{U}^\dagger\boldsymbol{W}^\dagger\boldsymbol{HFG}\boldsymbol{x}^{(t)} + \boldsymbol{U}^\dagger\boldsymbol{W}^\dagger\boldsymbol{n}^{(r)} \tag{2.1.4}$$

where $\boldsymbol{W} \in \mathbb{C}^{N\times r}$ and $\boldsymbol{U} \in \mathbb{C}^{r\times d}$ are the analog and digital combiners, respectively. Similarly, exploiting channel reciprocity, the uplink received signal is given by

$$\tilde{\boldsymbol{y}}^{(t)} = \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{H}^\dagger\boldsymbol{WU}\boldsymbol{x}^{(r)} + \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{n}^{(t)} \tag{2.1.5}$$

where $\tilde{\boldsymbol{y}}^{(t)}$ is the d-dimensional signal at the BS after linear filter, and $\boldsymbol{n}^{(t)}$ is the AWGN noise at the BS, such that $E[\boldsymbol{n}^{(t)}\boldsymbol{n}^{(t)^\dagger}] = \sigma^2_{(t)}\boldsymbol{I}_M$.

We highlight the main assumptions for this part of the thesis.

Assumption 2.1.1 (No CSI)**.** No a priori channel information is assumed. Rather, the subspaces in question, have to be estimated first. $\tilde{\boldsymbol{\Phi}}_1 \approx \boldsymbol{\Phi}_1$ at the MS, and $\tilde{\boldsymbol{\Gamma}}_1 \approx \boldsymbol{\Gamma}_1$ at the BS.

Assumption 2.1.2 (Slow block-fading channel)**.** The channel coherence time is assumed to be large enough, to make the estimation possible, e.g., low-mobility scenarios.

Assumption 2.1.3 (Channel Reciprocity)**.** We assume a TDD system where the hardware between transmitter and receiver is accurately calibrated, such that channel reciprocity holds

Assumption 2.1.4 (Decoding)**.** Joint encoding and decoding of each user's desired streams in assumed, and interference is treated as noise.

Problem Statement: The main goal for this part of the thesis is to firstly estimate the dominant left / right subspaces, i.e., obtain $\tilde{\boldsymbol{\Phi}}_1 \approx \boldsymbol{\Phi}_1$ at the MS, and $\tilde{\boldsymbol{\Gamma}}_1 \approx \boldsymbol{\Gamma}_1$ at the BS. We then wish to find the analog/digital precoder that best approximates $\tilde{\boldsymbol{\Gamma}}_1$, as well as analog/digital combiner that best approximates $\tilde{\boldsymbol{\Phi}}_1$, i.e., by solving (2.2.3).

2.2 Performance Metrics

We use the following expression as a performance metric (i.e., the "user-rate" corresponding to a given choice of precoders and combiners),

$$R = \log_2 \left| \boldsymbol{I}_d + \boldsymbol{H}_e\boldsymbol{H}_e^\dagger(\sigma^2_{(r)}\boldsymbol{U}^\dagger\boldsymbol{W}^\dagger\boldsymbol{WU})^{-1} \right| \tag{2.2.1}$$

where $\boldsymbol{H}_e = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H}\boldsymbol{F}\boldsymbol{G}$, $\frac{1}{\sigma^2_{(r)}} \triangleq$ SNR is the signal-to-noise ratio. Moreover we assume, for simplicity, that uniform power allocation is performed (no waterfilling), keeping in mind that a power allocation matrix $\boldsymbol{\Lambda}$ can be easily incorporated in the expression. Although not directly optimized, the above expression was used in [EARAS$^+$14], within the context of hybrid analog-digital precoding.

Remark 2.1 (Achievable rates)**.** Note that the value of the expression in (2.2.1) is related to achievable rates over the considered hybrid analog-digital MIMO link; in particular R becomes an *achievable rate* in the scenario that both the BS and MS are provided perfect knowledge of $\boldsymbol{H}$. Moreover, to the best of our knowledge, there is no work that attempted to investigate achievable rates for hybrid analog-digital MIMO systems, namely due to the lack of a prevalent statistical channel model for such channels. With that in mind, the aim is to present an approach to maximizing the metric R defined in (2.2.1). However, the value of the objective function is not necessarily *an achievable rate* for our system. That being said, optimizing similar expressions related to achievable rates has been proved to give good results in previous work on transmission with partial CSI [BB11]. Since any rate achievable with partial CSI, cannot be larger than the corresponding rate achievable with perfect CSI, this criterion always provides an upper bound on the achievable rates in our system. Hence, in our approach, if the proposed algorithms result in values for R that are closing in on the perfect CSI upper bound, then the scheme is performing optimally (in the sense of achievable rates). Thus, the optimal precoding characterization that we provide is contingent on R being achievable.

Using Hadamard's inequality, it can be easily verified that the optimal $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}$ that maximize R in (2.2.1), are the ones that diagonalize the effective channel $\boldsymbol{H}_e$. This is formalized below.

Proposition 2.2.1. *Assuming uniform power allocation, the optimal* $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}$ *that maximize* R *in* (2.2.1)*, diagonalize the effective channel* $\boldsymbol{H}_e$*, and satisfy* $\boldsymbol{F}\boldsymbol{G} = \boldsymbol{\Gamma}_1$ *and* $\boldsymbol{W}\boldsymbol{U} = \boldsymbol{\Phi}_1$*. Moreover, the resulting maximum rate is given by,*

$$R^\star \triangleq \max \; R(\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}) = \log_2 \left| \boldsymbol{I}_d + SNR \, \boldsymbol{\Sigma}_1^2 \right| \tag{2.2.2}$$

Proof. Refer to Appendix 2.3.1 □

With that in mind, we propose to tackle the following problem,

$$(\boldsymbol{F}^\star, \boldsymbol{G}^\star) = \begin{cases} \min\limits_{\boldsymbol{F},\, \boldsymbol{G}} \; \|\boldsymbol{\Gamma}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2 \\ \text{s. t. } \|\boldsymbol{F}\boldsymbol{G}\|_F^2 \leq d, \quad \boldsymbol{F} \in \mathcal{S}_{M,d} \end{cases}$$

$$(\boldsymbol{W}^\star, \boldsymbol{U}^\star) = \begin{cases} \min\limits_{\boldsymbol{W},\, \boldsymbol{U}} \; \|\boldsymbol{\Phi}_1 - \boldsymbol{W}\boldsymbol{U}\|_F^2 \\ \text{s. t.} \quad \boldsymbol{W} \in \mathcal{S}_{N,d} \end{cases} \tag{2.2.3}$$

The above design criterion has been quite prevalent in earlier works relating to the hybrid analog-digital architecture, and applied rather successfully in [AHAS+12, EARAS+14, MRRGPH15, AEALH14]. After a series of approximations to the mutual information in [EARAS+14], it was shown that the optimal precoders, $\boldsymbol{F}, \boldsymbol{G}$, are formulated in exactly the same fashion as above (though their formulation did not include receive combining).

In a nutshell, (2.2.3) boils down to finding $\boldsymbol{FG}$ (resp. $\boldsymbol{WU}$) that "best" approximate $\boldsymbol{\Gamma}_1$ (resp. $\boldsymbol{\Phi}_1$). Moreover, if there exists optimal precoders and combiners that make the distances in (2.2.3) zero, then they must satisfy

$$\boldsymbol{F}^\star \boldsymbol{G}^\star = \boldsymbol{\Gamma}_1, \quad \boldsymbol{W}^\star \boldsymbol{U}^\star = \boldsymbol{\Phi}_1.$$

We denote by $R^\star$ the resulting "user-rate" that is obtained by plugging in the above precoders/combiners in (2.2.1). Then $R^\star$ can be expressed as,

$$R^\star \triangleq R(\boldsymbol{F}^\star, \boldsymbol{G}^\star, \boldsymbol{W}^\star, \boldsymbol{U}^\star) = \log_2 \left| \boldsymbol{I}_d + \mathrm{SNR}\, \boldsymbol{\Sigma}_1^2 \right| \tag{2.2.4}$$

Following the above discussion on the achievability of R, $R^\star$ is the *maximum achievable rate* over the precoders and combiners, when $\boldsymbol{H}$ is known to both BS and MS. We underline the fact that R in (2.2.1) depends on the subspace spanned by the precoders / combiners, rather than the Euclidean distance between the right/left dominant subspace and the precoder/combiner, i.e., (2.2.3). However, optimizing metrics that involve span or chordal distances, is not straightforward. We thus emphasize that attempts at directly maximizing R in (2.2.1) are outside the scope of this work: rather, the focus is put on proposing mechanisms for subspace estimation and decomposition, and analyzing their performance.

2.3 Appendix

2.3.1 Proof of Proposition 2.2.1

Letting $\boldsymbol{Z} = \boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W} \boldsymbol{U})^{-1}$ and applying Hadamard's Inequality to R in (2.2.1) yields,

$$\log_2 |\boldsymbol{I}_d + \boldsymbol{Z}| \leq \log_2 \prod_{i=1}^{d} (1 + [\boldsymbol{Z}]_{ii}), \tag{2.3.1}$$

where the bound is achieved for $\boldsymbol{Z}$ diagonal. Moreover,

$$\begin{aligned}
\boldsymbol{Z} \text{ diagonal} &\Leftarrow \begin{cases} \boldsymbol{H}_e \boldsymbol{H}_e^\dagger \text{ diagonal} \\ \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W} \boldsymbol{U} \text{ diagonal} \end{cases} \\
&\Leftarrow \begin{cases} \boldsymbol{H}_e = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H} \boldsymbol{F} \boldsymbol{G} \text{ diagonal} \\ \boldsymbol{W} \boldsymbol{U} \text{ has orthonormal columns} \end{cases} \\
&\Leftarrow \begin{cases} \boldsymbol{F} \boldsymbol{G} = \boldsymbol{\Gamma}_1 \\ \boldsymbol{W} \boldsymbol{U} = \boldsymbol{\Phi}_1 \end{cases}
\end{aligned}$$

Plugging in the above choice of precoders / combiners in the upper bound in (2.3.1) yields,

$$\begin{aligned}
R^\star &= \log_2 \prod_{i=1}^{d} \left(1 + \lambda_i [\boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W} \boldsymbol{U})^{-1}]\right) \\
&= \log_2 \prod_{i=1}^{d} \left(1 + \lambda_i [\boldsymbol{H}_e \boldsymbol{H}_e^\dagger] / \sigma_{(r)}^2\right) \\
&= \log_2 \prod_{i=1}^{d} \left(1 + \sigma_i^2 [\boldsymbol{H}] / \sigma_{(r)}^2\right) = \log_2 |\boldsymbol{I}_d + \text{SNR } \boldsymbol{\Sigma}_1^2|
\end{aligned}$$

Chapter 3

Subspace Estimation and Decomposition

In this chapter we propose a method (based on the well-known Arnoldi iteration) exploiting channel reciprocity in TDD systems and the sparsity of the channel's eigenmodes, to estimate the right (resp. left) singular subspaces of the channel, at the BS (resp. MS), i.e., $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}$. We first describe the algorithm in the context of conventional MIMO systems, and derive bounds on the estimation error in the presence of distortions at both BS and MS. We later identify obstacles that hinder the application of such an algorithm to the hybrid analog-digital architecture, and address them individually. In view of fulfilling the constraints imposed by the hybrid analog-digital architecture, we further propose an iterative algorithm for subspace decomposition, whereby the above estimated subspaces, are approximated by a cascade of analog and digital precoder/combiner. Finally, we evaluate the performance of our scheme against the perfect CSI, fully digital case (i.e., an equivalent conventional MIMO system), and conclude that similar performance can be achieved, especially at medium-to-high SNR (where the performance gap is less than 5%), however, with a drastically lower number of RF chains ($\sim$ 4 to 8 times less). Moreover, note that our proposed technique encompasses both beamforming and precoding, i.e., it does not depend on the number of streams.

In addition to the notation defined in Chapter 1, we let $\hat{\boldsymbol{U}} = \mathrm{qr}(\boldsymbol{U})$ denote the semi-unitary matrix returned by the QR algorithm, with $\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{I}$. and

$$\mathcal{S}_{p,q} = \left\{ \boldsymbol{X} \in \mathbb{C}^{p\times q} \mid \ |\boldsymbol{X}_{ij}| = 1/\sqrt{p} \ , \ \forall (i,k) \in \{p\} \times \{q\} \right\}.$$

3.1 Motivation

In the previous chapter, we have identified the optimal transmission for the considered hybrid MIMO link, as the one along the dominant singular subspaces of the channel, i.e. $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}_1$. However, since we assume that no channel information is available at neither the BS, nor the MS, our aim in this chapter is to firstly obtain an *estimate of the subspaces in question*, i.e. $\tilde{\boldsymbol{\Phi}}_1 \approx \boldsymbol{\Phi}_1$ at the MS, and $\tilde{\boldsymbol{\Gamma}}_1 \approx \boldsymbol{\Gamma}_1$ at

the BS. We then propose methods that optimize the precoders and combiners to *accurately approximate the estimated subspaces*, by providing means to solve problems such as $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{FG}\|_F^2$ and $\|\tilde{\boldsymbol{\Phi}}_1 - \boldsymbol{WU}\|_F^2$ (while taking into consideration the constraints inherent to the hybrid analog-digital architecture).

3.2 Eigenvalue Algorithms and Subspace Estimation

3.2.1 Subspace Estimation vs. Channel Estimation

The aim of subspace estimation (SE) methods in MIMO systems is to estimate a predetermined *low-dimensional subspace of the channel*, required for transmission. We illustrate this in the context of conventional MIMO systems, i.e., where precoders/combiners are fully digital. For the sake of exposition, we start with a simple toy example, where noiseless single-stream transmission is assumed (and ignoring any physical constraints). The BS selects a random unit-norm beamforming vector, $\boldsymbol{p}_1$, and then sends $\boldsymbol{p}_1 x^{(t)}$, where $x^{(t)} = 1$. The received signal, $\boldsymbol{q}_1 = \boldsymbol{H}\boldsymbol{p}_1$, is echoed back to the BS (in effect, this implies that the signal is complex conjugated before being sent), in an Amplify-and-Forward (A-F) like fashion.[1] Then, exploiting channel reciprocity, the received signal at the BS is first normalized, i.e., $\boldsymbol{p}_2 = \boldsymbol{H}^\dagger\boldsymbol{q}_1/\|\boldsymbol{H}^\dagger\boldsymbol{q}_1\|_2 = \boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_1/\|\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_1\|_2$, and then echoed back to the MS. This simple procedure is done iteratively, and the resulting sequences $\{\boldsymbol{p}_l\}$ at the BS, and $\{\boldsymbol{q}_l\}$ at the MS, are defined as follows,

$$\boldsymbol{p}_{l+1} = \boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_l/\|\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_l\|_2; \quad \boldsymbol{q}_{l+1} = \boldsymbol{H}\boldsymbol{p}_l \tag{3.2.1}$$

It was noted in [DCG04] that using the Power Method (PM), one can show that as $l \to \infty$, $\boldsymbol{p}_l \to \boldsymbol{\gamma}_1$ and $\boldsymbol{q}_l \to \sigma_1\boldsymbol{\phi}_1$, implying that this seemingly simple "ad-hoc" procedure will converge to the *maximum eigenmode transmission*. The authors of [DCG04] also generalized the latter method to multistream transmission, i.e., by estimating $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}_1$, using the Orthogonal/Subspace Iteration (which was dubbed Two-way QR (TQR) in [DCG04, DPCG07]).

We note that SE schemes such as the ones described above, offer the following distinct advantage over classical *pilot-based channel estimation*: in spite of the large number of transmit and receive antennas, SE methods can estimate the dominant left/right singular subspaces with a relatively low communication overhead, when the latter have small dimension (relative to the channel dimensions). Consequently, subspace estimation is much more efficient than channel estimation, especially in large low-rank MIMO systems such as mmWave channels (because the latter estimates the dominant low-dimensional subspace instead of the whole channel). For the reason above, our proposed algorithm falls under the umbrella of SE methods. We first describe this algorithm in the context of "classical" MIMO systems, and later adapt it to the hybrid analog-digital architecture.

[1] This mechanism for MIMO subspace estimation, where the MS echoes back the transmitted signal using A-F, was first reported in [DCG04].

Set m ($m \leq M$); $\boldsymbol{q}_1$ = random unit-norm ; $\boldsymbol{Q} = [\boldsymbol{q}_1]$
for $l = 1, 2, ..., m$ **do**
1.a $\boldsymbol{p}_l = \boldsymbol{A}\boldsymbol{q}_l$
1.b $t_{k,l} = \boldsymbol{q}_k^\dagger \boldsymbol{p}_l,\ k = 1, \ldots, l$
2. $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{k=1}^{l} t_{k,l}\boldsymbol{q}_k$
3. $t_{l+1,l} = \|\boldsymbol{r}_l\|_2$; **if** $(t_{l+1,l} = 0)$ **stop**
4. $\boldsymbol{Q} = [\boldsymbol{Q},\ \boldsymbol{q}_{l+1} = \boldsymbol{r}_l / t_{l+1,l}]$
end for

Table 3.1: Arnoldi Procedure

3.2.2 Arnoldi Iteration for Subspace Estimation

Despite the fact that Krylov subspace methods (such as the Arnoldi and Lanczos Iterations for symmetric matrices) are among the most common methods for eigenvalue problems [Wat07], their use in the area of channel/subspace estimation is limited to equalization for doubly selective OFDM channels [HDMF10], and channel estimation in CDMA systems [TO02]. Algorithms falling into that category iteratively build a *basis for the Krylov subspace*, $\mathcal{K}^m = span\{\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, ..., \boldsymbol{A}^{m-1}\boldsymbol{x}\}$, one vector at a time. We use one of many variants of the so-called *Arnoldi Iteration/Procedure*, and a simplified version of the latter is shown in Table 3.1 (as presented in [Saa11]). The algorithm returns $\boldsymbol{Q}_m = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m] \in \mathbb{C}^{M \times m}$ and an upper Hessenberg matrix $\boldsymbol{T}_m \in \mathbb{C}^{m \times m}$, such that

$$\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m = \boldsymbol{T}_m,\ \boldsymbol{Q}_m^\dagger \boldsymbol{Q}_m = \boldsymbol{I}_m.$$

It can be shown that the algorithm iteratively builds $\boldsymbol{Q}_m$, an orthonormal basis for $\mathcal{K}^m$ (when roundoff errors are neglected), and that $\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m = \boldsymbol{T}_m$. We then say that the eigenvalues/eigenvectors of $\boldsymbol{T}_m$ are called *Ritz eigenvalues/eigenvectors*, and approximate the eigenvalues/eigenvectors of $\boldsymbol{A}$. The main idea behind processes such as the Arnoldi (and Lanczos) is to find the dominant eigenpairs of $\boldsymbol{A}$, by finding the eigenpairs of $\boldsymbol{T}_m$.

We note that the Arnoldi algorithm is a generalization of the Lanczos algorithm for the non-symmetric case, i.e., the latter is specifically tailored for cases where $\boldsymbol{A} \succeq \boldsymbol{0}$ (this is clearly the case in this work, since $\boldsymbol{A} = \boldsymbol{H}^\dagger \boldsymbol{H}$). This being said, the reason for not using the Lanczos iteration is that in practice, noise that is inherent to the echoing process, makes the Lanczos algorithm not applicable: namely, the requirement that $\boldsymbol{T}_m$ is tridiagonal, is violated.

Our goal in this section is to first apply the above algorithm to estimate the d largest eigenvectors of $\boldsymbol{A} = \boldsymbol{H}^\dagger \boldsymbol{H}$ at the BS (which are exactly $\boldsymbol{\Gamma}_1$), by implementing a *distributed version of the Arnoldi process*, that exploits the channel reciprocity inherent to TDD systems. Moreover, we extend the original formulation of the algorithm to incorporate a *distortion variable* (representing noise, or other distortions, as will be done later).

It becomes clear at this stage, that the BS requires knowledge of the sequence $\{\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l\}_{l=1}^m$, needed for the matrix-vector product in step 1 (Table 3.1): the latter can be accomplished by obtaining an estimate $\boldsymbol{p}_l$, of $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l,\ l \in \{m\}$. Without any explicit CSI at neither the BS nor the MS, we exploit the reciprocity of the medium to obtain such an estimate, via *BS-initiated echoing*: the BS sends $\boldsymbol{q}_l$ over the DL channel, the MS echoes back the received signal in an A-F like fashion, over the uplink (UL) channel (following the process proposed in [WTW08], and detailed in Sect. 3.2.1), i.e.,

$$\begin{aligned} DL: \quad & \boldsymbol{s}_l = \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{w}_l^{(r)} \\ UL: \quad & \boldsymbol{p}_l = \boldsymbol{H}^\dagger \boldsymbol{s}_l + \boldsymbol{w}_l^{(t)} = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{H}^\dagger \boldsymbol{w}_l^{(r)} + \boldsymbol{w}_l^{(t)} \\ & = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \tilde{\boldsymbol{w}}_l \end{aligned} \tag{3.2.2}$$

where $\boldsymbol{s}_l$ is the received signal in the DL, $\boldsymbol{w}_l^{(t)}$ and $\boldsymbol{w}_l^{(r)}$ are distortions at the BS and MS, respectively (representing noise for example).

After the echoing phase, the BS has an estimate, $\boldsymbol{p}_l$, of $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l$, as seen from (3.2.2). The remainder of the algorithm follows the conventional Arnoldi Iteration, and is shown in the Subspace Estimation using Arnoldi (SE-ARN) procedure (Table 3.2). In addition to $\boldsymbol{T}_m$ at the output of the algorithm, we define the matrices, $\tilde{\boldsymbol{T}}_m$, $\tilde{\boldsymbol{W}}_m$ and $\tilde{\boldsymbol{E}}_m$, as follows,

$$\begin{aligned} [\tilde{\boldsymbol{T}}_m]_{i,l} &= \begin{cases} \boldsymbol{q}_i^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l, & \text{if } l \le m,\ \forall i \le l \\ \|\boldsymbol{r}_l\|_2, & \text{if } l < m,\ i = l+1 \\ 0, & \text{otherwise} \end{cases} \\ \tilde{\boldsymbol{W}}_m &= [\tilde{\boldsymbol{w}}_1, ..., \tilde{\boldsymbol{w}}_m],\ \tilde{\boldsymbol{E}}_m = [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_{SL} \end{aligned} \tag{3.2.3}$$

where $\boldsymbol{r}_l$ is given in Step 2.b (Table 3.2). Note that similarly to the conventional Arnoldi Iteration, $\tilde{\boldsymbol{T}}_m$ is an the upper Hessenberg matrix. It then follows from the above definitions that

$$\boldsymbol{T}_m = \tilde{\boldsymbol{T}}_m + [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_U. \tag{3.2.4}$$

This can be easily verified by plugging in Step 1.b into 2.a in Table 3.2.

At the output of the SE-ARN procedure, the dominant eigenpairs of $\boldsymbol{H}^\dagger \boldsymbol{H}$ are approximated by those of $\boldsymbol{T}_m$ as follows. Let $\boldsymbol{T}_m = \tilde{\boldsymbol{\Theta}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Theta}}^{-1}$ be eigenvalue decomposition of $\boldsymbol{T}_m$, where $\tilde{\boldsymbol{\Theta}}$ is the (possibly non-orthonormal) set of eigenvectors. Then, it can easily be shown that $\hat{\boldsymbol{\Gamma}}_1 = \mathrm{qr}(\boldsymbol{Q}_m[\tilde{\boldsymbol{\Theta}}]_{1:d})$ are the Ritz eigenvectors of $\boldsymbol{H}^\dagger \boldsymbol{H}$, where $[\tilde{\boldsymbol{\Theta}}]_{1:d}$ has as columns the eigenvectors of $\boldsymbol{T}_m$ associated with the d largest eigenvalues (in magnitude).

Remark 3.1. To be exact, the Ritz eigenvectors do not contain any estimation noise. That being said, we stick to this nomenclature, with a slight abuse of definition. Moreover, $\tilde{\boldsymbol{\Sigma}}_1$, the Ritz eigenvalues of $\boldsymbol{H}^\dagger \boldsymbol{H}$, come for free once the Ritz eigenvectors are obtained (Table 3.2).

procedure $\tilde{\boldsymbol{\Gamma}}_1, \tilde{\boldsymbol{\Sigma}}_1 =$ SE-ARN $(\boldsymbol{H}, d)$
Set m $(m \leq M)$; Random unit-norm $\boldsymbol{q}$; $\boldsymbol{Q} = [\boldsymbol{q}_1]$
for $l = 1, 2, ..., m$ **do**
// BS-initiated echoing: estimate $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l$
1.a $\boldsymbol{s}_l = \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{w}_l^{(r)}$
1.b $\boldsymbol{p}_l = \boldsymbol{H}^\dagger \boldsymbol{s}_l + \boldsymbol{w}_l^{(t)}$
// Gram-Schmidt orthogonalization
2.a $t_{k,l} = \boldsymbol{q}_k^\dagger \boldsymbol{p}_l \quad , \forall\ k = 1, \ldots, l$
2.b $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{k=1}^{l} \boldsymbol{q}_k t_{k,l}$
2.c $t_{l+1,l} = \|\boldsymbol{r}_l\|_2$
// Update $\boldsymbol{Q}$
3.a $\boldsymbol{Q} = [\boldsymbol{Q},\ \boldsymbol{q}_{l+1} = \boldsymbol{r}_l / t_{l+1,l}]$
end for
// Compute $\tilde{\boldsymbol{\Gamma}}_1$
$\boldsymbol{T}_m = \tilde{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\Theta}}^{-1}$
$\tilde{\boldsymbol{\Gamma}}_1 = \mathrm{qr}(\boldsymbol{Q}_m \tilde{\boldsymbol{\Theta}}_{1:d})$
$[\tilde{\boldsymbol{\Sigma}}_1]_{i,i} = \sqrt{|[\tilde{\boldsymbol{\Lambda}}]_{i,i}|}, \forall\ i$
end procedure

Table 3.2: Subspace Estimation using Arnoldi Iteration (SE-ARN)

Note that the latter procedure results in the BS obtaining $\tilde{\boldsymbol{\Gamma}}_1$, and consequently $\tilde{\boldsymbol{\Sigma}}_1$, using the so-called BS-initiated echoing. This same procedure can be applied using MS-initiated echoing, to estimate $\tilde{\boldsymbol{\Phi}}_1$ (i.e., the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^\dagger$), at the MS (Fig.3.1).

3.2.3 Perturbation Analysis

In what follows, we extend some of the known properties of the conventional Arnoldi iteration, to account for the estimation error, emanating from the distortion variable.

Lemma 3.2.1. *For the output of the Arnoldi process the following holds,*
$(P1)$:

$$\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m = \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m \triangleq \boldsymbol{C}_m, \tag{3.2.5}$$

where $\boldsymbol{C}_m = \boldsymbol{S}_m \boldsymbol{\Lambda}_m \boldsymbol{S}_m^{-1}$ *is such that* $[\boldsymbol{\Lambda}]_{i,i} \geq 0$ *and* $\boldsymbol{S}_m^{-1} = \boldsymbol{S}_m^\dagger$

$(P2)$: *Let* $(\lambda_i^{(m)}, \boldsymbol{s}_i^{(m)})$ *be any eigenpair of* $\boldsymbol{C}_m$. *Then* $(\lambda_i^{(m)},\ \boldsymbol{\Theta}_i^{(m)} \triangleq \boldsymbol{Q}_m \boldsymbol{s}_i^{(m)})$ *is an approximate Ritz eigenpair for* $\boldsymbol{A}$. *Furthermore, the approximation error is such*

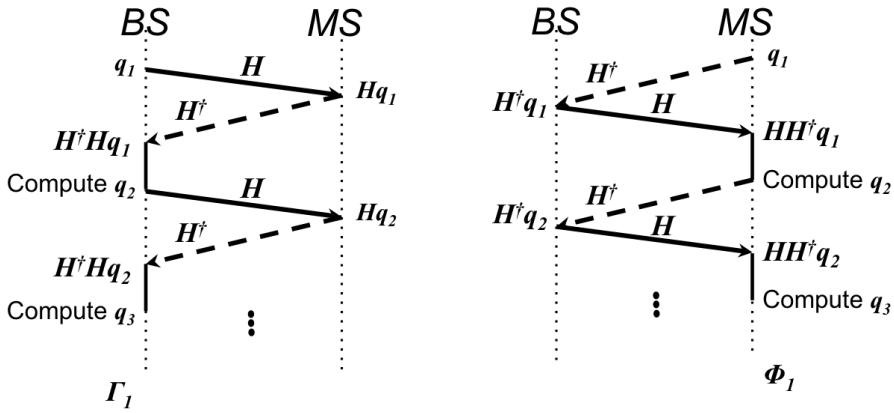

Figure 3.1: Proposed algorithm for Subspace Estimation using Arnoldi Iteration, and its resulting communication structure at BS and MS

that,

$$\|\boldsymbol{A}\boldsymbol{\Theta}_i^{(m)} - \lambda_i^{(m)}\boldsymbol{\Theta}_i^{(m)}\|_2^2 \leq c_m^{(i)} + \|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2\|\tilde{\boldsymbol{W}}_m\|_F^2, \tag{3.2.6}$$

where $c_m^{(i)} = ([\tilde{\boldsymbol{T}}_m]_{m+1,m}|[\boldsymbol{s}_i^{(m)}]_m|)^2$.

$(P3)$: *As* $m \to M$, $\|\boldsymbol{A}\boldsymbol{\Theta}_i^{(m)} - \lambda_i^{(m)}\boldsymbol{\Theta}_i^{(m)}\|_2^2 \to 0$, *implying that the eigenpairs of* $\boldsymbol{C}_m$ *perfectly approximate the eigenpairs of* $\boldsymbol{A}$*(up to round-off errors).*

Proof. The proof is shown in Appendix 3.6.1. □

We underline the fact that if the distortion variable $\tilde{\boldsymbol{W}}_m$ is zero, the above derivations reduce to the well-known results on the Arnoldi process [Saa11, Sect. 6.2]. Lemma 3.2.1 establishes the fact that each eigenpair $(\lambda_i^{(m)}, \boldsymbol{s}_i^{(m)})$ of $\boldsymbol{C}_m$, is associated with one eigenpair $(\lambda_i^{(m)}, \boldsymbol{\Theta}_i^{(m)})$ of $\boldsymbol{A}$.Though $(P3)$ in Lemma 3.2.1 implies that the error in approximating the eigenpairs of $\boldsymbol{A}$ with those of $\boldsymbol{C}_m$ vanishes as $m \to M$, our simulations will later show that very good approximations can be obtained, even for $m \ll M$.

Thus, one might be tempted to conclude at this point, that by computing the eigenpairs of $\boldsymbol{C}_m$, one can *perfectly estimate* the eigenpairs of $\boldsymbol{A}$, despite the presence of the distortion variable $\tilde{\boldsymbol{W}}_m$. However, the fact remains that $\boldsymbol{C}_m \triangleq \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m$ *cannot be computed*, mainly because $\tilde{\boldsymbol{E}}_m$ is not known to the BS. As a result, $\boldsymbol{T}_m$ *at the output of the Arnoldi process will be used instead to approximate the eigenpairs of* $\boldsymbol{A}$. Now that we established that the eigenpairs of $\boldsymbol{C}_m$ approximate that of $\boldsymbol{A}$, the natural question is *how close are the eigenpairs of* $\boldsymbol{T}_m$*, to that of* $\boldsymbol{C}_m$.

For that purpose, we first show the following,

$$\begin{aligned}\boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m &= (\tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m) + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m \\ &= \tilde{\boldsymbol{T}}_m + (\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m - [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_{SL}) \\ &= \tilde{\boldsymbol{T}}_m + [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_U \triangleq \boldsymbol{T}_m \end{aligned} \tag{3.2.7}$$

where the first equality follows from the definition of $\boldsymbol{C}_m$, and the last one from (3.2.4). Thus $\boldsymbol{C}_m$ can be viewed as the matrix in question, and $\boldsymbol{P}_m \triangleq \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m$ a perturbation matrix. We then apply the Bauer-Fike Theorem [GVL96, Th. 7.2.2] to bound the difference in eigenvalues.

Lemma 3.2.2. *Every eigenvalue $\tilde{\lambda}$ of $\boldsymbol{T}_m = \boldsymbol{C}_m + \boldsymbol{P}_m$ satisfies*

$$|\tilde{\lambda} - \lambda| \leq \sqrt{m}\, \|\tilde{\boldsymbol{W}}_m\|_F,$$

where λ is an eigenvalue of $\boldsymbol{C}_m$.

Proof. Refer to Appendix 3.6.2 □

Summarizing thus far, Lemma 3.2.1 showed that the eigenpairs of $\boldsymbol{A}$ can be approximated by the eigenvalues of $\boldsymbol{C}_m$, with arbitrarily small error. However, since the latter is not available, we approximate the eigenpairs of $\boldsymbol{C}_m$ (and consequently of $\boldsymbol{A}$) by those of $\boldsymbol{T}_m$, the upper Hessenberg matrix at the output of the Arnoldi process. Finally, Lemma 3.2.2 established the fact that this approximation error, for the eigenvalues, is upper bounded by the magnitude of the perturbation itself. We note that the relevant "error-metric" here is the distance between the true subspace $\boldsymbol{\Gamma}_1$, and estimated subspace $\tilde{\boldsymbol{\Gamma}}_1 \propto \boldsymbol{Q}_m \tilde{\boldsymbol{\Theta}}_{1:d}$ (Table 3.2). This does suggest that the estimation error is dependent on $\tilde{\boldsymbol{\Theta}}_{1:d}$, the eigenvectors of $\boldsymbol{T}_m$. However, performing a similar sensitivity analysis on the eigenvectors is much more involved, since the sensitivity of eigenvectors generally depends on the clustering of eigenvalues.

3.3 Hybrid Analog-Digital Precoding for mmWaveMIMO systems

In this section we turn our attention to applying the above framework for subspace estimation and precoding, to the hybrid analog-digital architecture. As this section will gradually reveal, several obstacles have to be overcome for that matter. We start by presenting some preliminaries that will be used throughout this section.

3.3.1 Preliminaries: Subspace Decomposition

We will limit our discussion to the digital and analog precoder, keeping in mind that the same applies to the digital and analog combiner. In conventional MIMO systems, the estimates of the right and left singular subspace, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\boldsymbol{\Phi}}_1$, obtained

using SE-ARN, can directly be used to diagonalize the channel. However, the hybrid analog-digital architecture entails a cascade of analog and digital precoder. Thus, $\tilde{\boldsymbol{\Gamma}}_1$ has to be decomposed into $\boldsymbol{FG}$ (hence the term *Subspace Decomposition (SD)*), as follows,

$$\begin{cases} \min\limits_{\boldsymbol{F},\,\boldsymbol{G}} & h_0(\boldsymbol{F},\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{FG}\|_F^2 \\ \text{s. t.} & h_1(\boldsymbol{F},\boldsymbol{G}) = \|\boldsymbol{FG}\|_F^2 \leq d \\ & \boldsymbol{F} \in \mathcal{S}_{M,d} \end{cases} \tag{3.3.1}$$

We underline the fact that the authors in [EARAS+14] arrived to the same formulation as (3.3.1), and proposed a variation on the well-known Orthogonal Matching Pursuit (OMP), to tackle it. The same framework was recently extended in [MRRGPH15] to relax the need for dictionaries based on the array response matrix. An alternate decomposition was proposed by [SY15b], where the optimization metric is the user rate. Both works were published after the initial submission of our paper.

Within the context of hybrid precoding, the authors in [ZMK05] showed that there exists (non-unique) $\boldsymbol{F} \in \mathcal{S}_{M,r}, \boldsymbol{g} \in \mathbb{C}^{r\times 1}$ such that $\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{Fg}$, if and only if $r \geq 2$. This was extended in [MRRGPH15] where it was shown that there exists $\boldsymbol{F} \in \mathcal{S}_{M,r}, \boldsymbol{G} \in \mathbb{C}^{r\times d}$ such that $\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{FG}$, if $r \geq 2d$. We note that for such cases, the cost function in (3.3.1) is zero, and we refer to such cases as *optimal decomposition* -whose performance we evaluate in the numerical results section: although the aforementioned schemes use all the available RF chains for the decomposition (and our decomposition uses a subset of the RF chains), the sum-rate performance is actually the same.

To a certain extent, (3.3.1) is reminiscent of formulations arising from areas such as blind source separation, (sparse) dictionary learning, and vector quantization [XY13, AEB06]. Though there is a battery of algorithms and techniques that have been developed to tackle such problems, the additional hardware constraint on $\boldsymbol{F}$, i.e. $\boldsymbol{F} \in \mathcal{S}_{M,r}$ makes the use of such tools not possible. As a result, we will resort to developing our own algorithm. In spite of the non-convex and non-separable nature of the above quadratically-constrained quadratic program, we propose an iterative method that attempts to determine an approximate solution.

Block Coordinate Descent for Subspace Decomposition

In this part, we further assume that only d of the r available RF chains are used, i.e., $\boldsymbol{F} \in \mathbb{C}^{M\times d}$ and $\boldsymbol{G} \in \mathbb{C}^{d\times d}$ (the reason for that will become clear later in this section). The coupled nature of the objective and constraints in (3.3.1) suggests a Block Coordinate Descent (BCD) approach. The main challenges arise from the coupled nature of the variables in the constraint (since the latter makes convergence claims of BCD, not possible [RHL12]), and from the hardware constraint on $\boldsymbol{F}$. We will show that a BCD approach implicitly enforces the power constraint in (3.3.1), and consequently the latter can be dropped without changing the problem.

Our approach consists in relaxing the hardware constraint on $\boldsymbol{F}$, and then applying a Block Coordinate Descent (BCD) approach to alternately optimize $\boldsymbol{F}$ and $\boldsymbol{G}$ (while projecting each of the obtained solutions for $\boldsymbol{F}$ on $\mathcal{S}_{M,d}$). For that matter, we first define the *Euclidean projection* on the set $\mathcal{S}_{M,d}$ in the following proposition.

Proposition 3.3.1. *Let* $\boldsymbol{X} \in \mathbb{C}^{M\times d}$ *be defined as* $[\boldsymbol{X}]_{i,k} = |x_{i,k}|\ e^{j\phi_{i,k}},\ \forall(i,k)$, *and*

$$\boldsymbol{Y} = \Pi_{\mathcal{S}}[\boldsymbol{X}] \triangleq \underset{\boldsymbol{U}\in\mathcal{S}_{M,d}}{\operatorname{argmin}} \|\boldsymbol{U}-\boldsymbol{X}\|_F^2$$

denote its (unique) Euclidean projection on the set $\mathcal{S}_{M,d}$.
Then $[\boldsymbol{Y}]_{i,k} = (1/\sqrt{M})\ e^{j\phi_{i,k}}, \forall(i,k)$.

Proof. Refer to Appendix 3.6.4 □

The latter result implies that given an arbitrary $\boldsymbol{F}$, finding the closest point to $\boldsymbol{F}$, lying in $\mathcal{S}_{M,d}$ simply reduces to *setting the magnitude of each element in* $\boldsymbol{F}$, *to* $1/\sqrt{M}$.

Neglecting the constraint on $\boldsymbol{F}$ in (3.3.1), one can indeed show that for fixed $\boldsymbol{G}$ (resp. $\boldsymbol{F}$), the resulting subproblem is convex in $\boldsymbol{F}$ (resp. $\boldsymbol{G}$). With this in mind, our aim is to produce a *sequence of updates,* $\{\boldsymbol{F}_k, \boldsymbol{G}_k\}_k$ *such that the sequence* $\{h_0(\boldsymbol{F}_k, \boldsymbol{G}_k)\}_k$ *is non-increasing* (keeping in mind that monotonicity cannot be shown due to the coupling in the power constraint). Thus, given $\boldsymbol{G}_k$, each of the updates, $\boldsymbol{F}_{k+1}$ and $\boldsymbol{G}_{k+1}$, are defined as as follows,

$$(J1)\quad \boldsymbol{F}_{k+1} \triangleq \min_{\boldsymbol{F}}\ \ h_0(\boldsymbol{F}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}_k\|_F^2$$
$$(J2)\quad \boldsymbol{G}_{k+1} \triangleq \min_{\boldsymbol{G}}\ \ h_0(\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}_{k+1}\boldsymbol{G}\|_F^2$$

Both $(J1)$ and $(J2)$ are instances of a non-homogeneous (unconstrained) convex quadratically-constrained quadratic programming (QCQP) that can easily be solved (globally) by finding stationary points of their respective cost functions, to yield,

$$\boldsymbol{F}_{k+1} = \tilde{\boldsymbol{\Gamma}}_1 \boldsymbol{G}_k^\dagger (\boldsymbol{G}_k \boldsymbol{G}_k^\dagger)^{-1} \tag{3.3.2}$$

$$\boldsymbol{G}_{k+1} = (\boldsymbol{F}_{k+1}^\dagger \boldsymbol{F}_{k+1})^{-1} \boldsymbol{F}_{k+1}^\dagger \tilde{\boldsymbol{\Gamma}}_1 \tag{3.3.3}$$

We note that our earlier assumption that only d of the RF chains are used here (i.e. $\boldsymbol{G}$ is square), guarantees that, $(\boldsymbol{G}_l \boldsymbol{G}_l^\dagger)$ in (3.3.3) is invertible, almost surely: in fact, as our numerical results will later show, the incurred performance loss is quite negligible.

Moreover, note that the solution in (3.3.2) does not necessarily satisfy the hardware constraint on $\boldsymbol{F}$. Thus, the result of Proposition 3.3.1 can be used to compute

procedure $[\boldsymbol{F}, \boldsymbol{G}] = \text{BCD-SD}\ (\tilde{\boldsymbol{\Gamma}}_1)$
  Start with arbitrary $\boldsymbol{F}_0$
  for $k = 0, 1, 2, \ldots$ **do**
    $\boldsymbol{G}_{k+1} \leftarrow (\boldsymbol{F}_k^\dagger \boldsymbol{F}_k)^{-1} \boldsymbol{F}_k^\dagger \tilde{\boldsymbol{\Gamma}}_1$
    $\boldsymbol{F}_{k+1} \leftarrow \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\Gamma}}_1 \boldsymbol{G}_{k+1}^\dagger (\boldsymbol{G}_{k+1} \boldsymbol{G}_{k+1}^\dagger)^{-1}]$
  end for
end procedure

Table 3.3: Block Coordinate Descent for Subspace Decomposition (BCD-SD)

the projection of $\boldsymbol{F}$ on $\mathcal{S}_{M,d}$. To prove our earlier observation that the optimal updates $\boldsymbol{F}_{k+1}$ and $\boldsymbol{G}_{k+1}$ satisfy the power constraint in (3.3.1), we plug (3.3.3) into the following (dropping all subscripts for simplicity),

$$\begin{aligned} \|\boldsymbol{F}\boldsymbol{G}\|_F^2 &= \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1^\dagger \boldsymbol{F} \underbrace{(\boldsymbol{F}^\dagger \boldsymbol{F})^{-1} \boldsymbol{F}^\dagger \boldsymbol{F}}_{=\boldsymbol{I}_d} (\boldsymbol{F}^\dagger \boldsymbol{F})^{-1} \boldsymbol{F}^\dagger \tilde{\boldsymbol{\Gamma}}_1 \right) \\ &\leq \operatorname{tr}\left((\boldsymbol{F}^\dagger \boldsymbol{F})^{-1} \boldsymbol{F}^\dagger \boldsymbol{F} \right) \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1 \tilde{\boldsymbol{\Gamma}}_1^\dagger \right) = d \end{aligned} \tag{3.3.4}$$

where we assumed that $\|\tilde{\boldsymbol{\Gamma}}_1\|_F^2 = 1$ w.l.o.g., and used the fact that $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \leq \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \succeq \boldsymbol{0}$. Note that the above relation holds for any arbitrary full-rank $\boldsymbol{F}$, and thus, the power constraint is satisfied even after applying the projection step. The above shows that if BCD is used, then the power constraint in (3.3.1) is always enforced. The corresponding method is termed Block Coordinate Descent for Subspace Decomposition (BCD-SD), and is shown in Table 3.3.

Remark 3.2. We underline the fact that due to the projection step, one cannot show that the sequence $\{h_o(\boldsymbol{F}_k, \boldsymbol{G}_k)\}_k$ is non-increasing. Nevertheless, despite the fact that monotonic convergence of BCD-SD cannot be showed analytically, our simulations indicate that the latter is indeed the case, under normal operating conditions.

Remark 3.3. It can be easily verified that the optimal $\boldsymbol{F}^\star, \boldsymbol{G}^\star$ that maximize R in (2.2.1) are such that $\|\boldsymbol{F}^\star \boldsymbol{G}^\star\| = d$. Though the optimal solution to (3.3.1) is not invariant to scaling, as far as the performance metric in (2.2.1) is concerned, there in no loss in optimality in scaling the solution given by BCD-SD, to fulfill the power constraint with equality.

One-dimensional case

Note that echoing (e.g., our proposed mechanism in Table 3.2) relies on the BS being able to send any vector $\boldsymbol{q}_l$, to be echoed back by the MS. For the hybrid

analog-digital architecture, this translates into the BS being able to (accurately) approximate $\boldsymbol{q}_l$ by $\boldsymbol{f}_l g_l$, where $\boldsymbol{f}_l$ is a vector, g_l is a scalar. As a result, subspace decomposition for the one-dimensional case is of great interest here. When $d = 1$, (3.3.1) reduces to the problem below,

Lemma 3.3.1. *Consider the single dimension SD problem,*

$$\begin{cases} \min\limits_{\boldsymbol{f},\ g} \ h_o(\boldsymbol{f},\ g) = \|\boldsymbol{f}\|_2^2\ g^2 - 2g\Re(\boldsymbol{f}^\dagger \tilde{\boldsymbol{\gamma}}_1) \\ \text{s. t. } [\boldsymbol{f}]_i = 1/\sqrt{M}\ e^{j\phi_i}, \forall i \end{cases} \tag{3.3.5}$$

where $g \in \mathbb{R}_+$ and $[\tilde{\boldsymbol{\gamma}}_1]_i = r_i e^{j\theta_i}$. Then the problem admits a globally optimum solution given by, $[\boldsymbol{f}^\star]_i = 1/\sqrt{M}\ e^{j\theta_i}, \forall i$ and $g^\star = \|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M}$

Proof. Refer to Appendix 3.6.3 □

Similarly to (3.3.4), it can be verified that a power constraint is indeed implicitly verified. Moreover, the approximation error $\boldsymbol{e} \triangleq \tilde{\boldsymbol{\gamma}}_1 - \boldsymbol{f}g$ is such that,

$$[\boldsymbol{e}]_i = |r_i - \|\tilde{\boldsymbol{\gamma}}_1\|_1/M| e^{j\theta_i},\ \forall i \in \{M\}. \tag{3.3.6}$$

We note that when considering the effective beamformer, i.e., $\boldsymbol{f}g$, the solution given by Lemma 3.3.1 is to some extent reminiscent of equal gain transmission in [LH03, ZXLS07], in terms of the optimal phases.

We recall that a similar hybrid beamforming setup was considered in [ZMK05] where the authors optimize $u, \boldsymbol{w}, \boldsymbol{f}, g$, to maximize the SNR as well as the spectral efficiency. Although our formulation optimizes the same quantities, the optimization metric we consider, the subspace distance, is different.

Column-wise Decomposition Note that the decomposition can be written in a simple form. Given a vector $\tilde{\boldsymbol{\gamma}}_1$, its globally optimal decomposition (from the perspective of (3.3.1)) is given as,

$$\tilde{\boldsymbol{\gamma}}_1 \approx g_1^\star \boldsymbol{f}_1^\star \triangleq (\|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M})\ \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_1].$$

This can be generalized to obtain an alternate method to BCD-SD, by decomposing $\tilde{\boldsymbol{\Gamma}}_1$, in a *column-wise* fashion,

$$\begin{aligned} \tilde{\boldsymbol{\Gamma}}_1 &= [\tilde{\boldsymbol{\gamma}}_1, \cdots, \tilde{\boldsymbol{\gamma}}_d] \approx [g_1^\star \boldsymbol{f}_1^\star, \cdots, g_d^\star \boldsymbol{f}_d^\star] \\ &\triangleq (1/\sqrt{M})\ [\Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_1], \cdots, \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_d]]\ \operatorname{diag}(\|\tilde{\boldsymbol{\gamma}}_1\|_1, \cdots, \|\tilde{\boldsymbol{\gamma}}_d\|_1) \end{aligned} \tag{3.3.7}$$

Numerical Results

As mentioned earlier, (3.3.1) was formulated and solved in [EARAS+14], using a variation on the well-known Orthogonal Matching Pursuit (OMP), by recovering $\boldsymbol{F}$ in a greedy manner, then updating the estimate of $\boldsymbol{G}$ in a least squares sense.

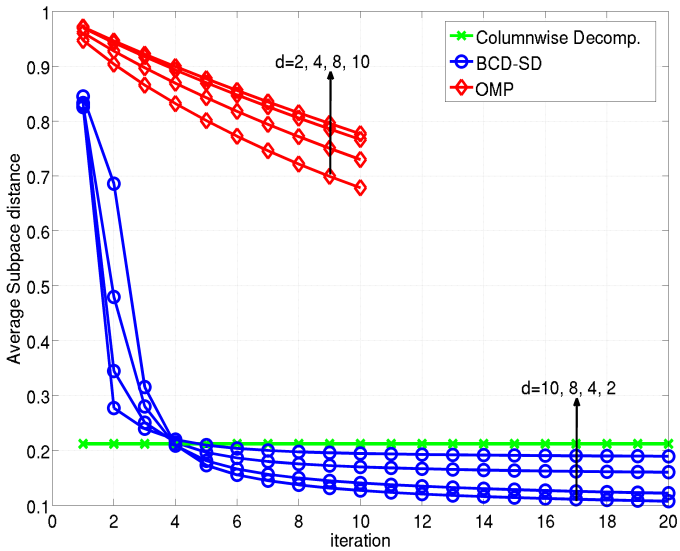

Figure 3.2: Average subspace distance $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{FG}\|_F^2$, for our proposed method and OMP

We thus compare its average performance with our proposed method, for a case where $\tilde{\boldsymbol{\Gamma}}_1 \in \mathbb{C}^{M \times d}$ is such that $M = 64, r = 10$ (for several values of d). The curves are averaged over 500 random realizations of $\boldsymbol{\Gamma}_1$ (the latter are random unitary matrices). Moreover, we follow the same setup for OMP as that of [EARAS$^+$14], namely, that the dictionary is designed based on the array response vectors (of size 256). The reason for the large performance gap in Fig. 3.2 is that BCD-SD attempts to find a locally optimal solution to (3.3.1) (though this cannot be shown due to the coupled variables). Moreover, OMP is halted after r iterations, since it recovers the columns of $\boldsymbol{F}$ one at a time, whereas our proposed method runs until reaching a stable point. With that in mind, although OMP might perform better in terms of approximating the span of $\boldsymbol{\Gamma}_1$, it is challenging to measure and optimize such metrics in practice. Moreover, we recall that in its original formulation in [EARAS$^+$14] OMP is indeed formulated to solve the problem at hand (i.e. (3.3.1)), and thus the comparison seems fair. Interestingly, despite its extreme simplicity, the column-wise decomposition in (3.3.7) offers a surprisingly good performance (as seen in Fig. 3.2).

3.3.2 Echoing in the Hybrid Analog-Digital Architecture

It is clear by now that the gist behind the schemes described in this work, is to obtain an estimate of $\{\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l\}_{l=1}^m$ at the BS, by exploiting channel reciprocity, using BS-initiated echoing described in (3.2.2). However, in the case of the hybrid analog-digital architecture, there are several issues that prevent the application of the latter procedure. Firstly, the digital beamforming vector $\boldsymbol{q}_l$ needs to be approximated by a cascade of analog and digital beamformer, using the decomposition in Sect. 3.3.1, i.e., $\boldsymbol{q}_l = \tilde{\boldsymbol{f}}_l \tilde{g}_l + \boldsymbol{e}_l$, where $\boldsymbol{e}_l$ is the approximation error given in (3.3.6). Moreover, the BS-initiated echoing relies on the MS being able to amplify-and-forward its received signal: this is clearly *not possible* using the hybrid analog-digital architecture. In addition, neither the BS nor MS can digitally process the received signal at the antennas: only after the application the analog precoder/combiner (and possibly the digital precoder/combiner) can the baseband signal be digitally manipulated [WLP+09, EARAS+14].

With this in mind, we *emulate* the A-F step in BS-initiated echoing, (3.2.2), as follows. $\boldsymbol{q}_l$ is decomposed into $\tilde{\boldsymbol{f}}_l \tilde{g}_l$ at the BS and sent over the DL. The MS linearly processes the received signal in the downlink, with the analog combiner, i.e., $\boldsymbol{s}_l = \boldsymbol{W}_l^\dagger (\boldsymbol{H}\tilde{\boldsymbol{f}}_l \tilde{g}_l)$, and same filter is used as the analog precoder, to process the transmit signal in the UL, i.e., $\boldsymbol{W}_l \boldsymbol{s}_l$. Finally, the received signal at the BS is processed with the analog precoder, $\boldsymbol{F}_l$. The resulting estimate, $\boldsymbol{p}_l$, at the BS is,

$$\boldsymbol{p}_l = \boldsymbol{F}_l^\dagger \boldsymbol{H}^\dagger \boldsymbol{W}_l \boldsymbol{W}_l^\dagger \boldsymbol{H}(\boldsymbol{q}_l - \boldsymbol{e}_l) \tag{3.3.8}$$

Note that the above process is possible using the hybrid analog-digital architecture. Since noise is present in any uplink/downlink transmission, for clarity in what follows, we drop the noise-related terms from all equations. Needless to say, their effect is accounted for in the numerical results. It is clear from (3.3.8) that $\boldsymbol{p}_l$ is no longer a "good" estimate of $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l$, for the reasons stated below.

1. *Analog-Processing Impairments (API):* Processing the signal at the MS with the analog combiner/precoder $\boldsymbol{W}_l$ greatly distorts the singular values/vectors of the effective channel. Moreover, processing the received signal at the BS with the analog combiner $\boldsymbol{F}_l \in \mathbb{C}^{M \times r}$ implies that $\boldsymbol{p}_l$ is now a low-dimensional observation of the desired M-dimensional quantity $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l$ (since $r < M$).

2. *Decomposition-Induced Distortions (DID):* The error from decomposing $\boldsymbol{q}_l$ at the BS, $\boldsymbol{e}_l$, further distorts the estimate (as seen in (3.3.8)).

The above impairments are a byproduct of shifting the burden of digital precoding, to the analog domain. In what follows, these impairments will individually be investigated and addressed.

Cancellation of Analog-Processing Impairments

Our proposed method for mitigating analog-processing impairments (API) relies on the simple idea of taking multiple measurements at both the BS and MS, and lin-

early combining them, such that $\boldsymbol{W}_l\boldsymbol{W}_l^\dagger$ and $\boldsymbol{F}_l\boldsymbol{F}_l^\dagger$ approximate an identity matrix.

In the DL, $\boldsymbol{q}_l$ is approximated by $\tilde{\boldsymbol{f}}_l\tilde{g}_l$, and $\tilde{\boldsymbol{f}}_l\tilde{g}_l$ is sent over the DL channel[2], K_r times (where $K_r = N/r$), each linearly processed with an analog combiner $\{\boldsymbol{W}_{l,k} \in \mathbb{C}^{N\times r}\}_{k=1}^{K_r}$, to obtain the digital samples $\{\boldsymbol{s}_{l,k}\}_{k=1}^{K_r}$ (this process is shown in Table (3.4)). Moreover, the analog combiners are taken from the columns of a Discrete Fourier Transform (DFT) matrix, i.e,

$$[\boldsymbol{W}_{l,1}, ..., \boldsymbol{W}_{l,K_r}] = \boldsymbol{D}_r, \tag{3.3.9}$$

$$\text{where } \boldsymbol{D}_r = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \cdots & e^{-j\frac{2(N-1)\pi}{N}} \\ \vdots & \vdots & & \vdots \\ 1 & e^{-j\frac{2(N-1)\pi}{N}} & \cdots & e^{-j\frac{2(N-1)^2\pi}{N}} \end{bmatrix}. \tag{3.3.10}$$

is a normalized $N \times N$ DFT matrix (i.e., where each column has unit norm and satisfies the unit-modulus constraint). The same analog combiners, $\{\boldsymbol{W}_{l,k}\}_k$, are used to linearly combine $\{\boldsymbol{s}_{l,k}\}_k$, to form $\tilde{\boldsymbol{s}}_l$. We dub this procedure Repetition-Aided (RAID) Echoing, and the aforementioned DL phase, is shown in Table 3.4. The resulting signal at the MS, $\tilde{\boldsymbol{s}}_l$, can be rewritten as,

$$\tilde{\boldsymbol{s}}_l = \left(\sum_{k=1}^{K_r} \boldsymbol{W}_{l,k}\boldsymbol{W}_{l,k}^\dagger\right) \boldsymbol{H}(d\tilde{\boldsymbol{f}}_l\tilde{g}_l) = d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l, \tag{3.3.11}$$

where equality follows from our earlier definition of $\{\boldsymbol{W}_{l,k}\}_k$ in (3.3.9). Note that *the effect of processing the received signal with the analog combiner has been completely suppressed.* Now, $\tilde{\boldsymbol{s}}_l$ is normalized, and echoed back in the UL direction.

A quite similar process is used in the UL: $\tilde{\boldsymbol{s}}_l$ is first decomposed into $\tilde{\boldsymbol{w}}_l\tilde{u}_l$, d RF chains are used to send it over the UL, K_t times (where $K_t = M/r$), and each observation is linearly processed with an analog combiner $\{\boldsymbol{F}_{l,m} \in \mathbb{C}^{M\times r}\}_{m=1}^{K_t}$. The resulting digital samples $\{\boldsymbol{z}_{l,m}\}_{m=1}^{K_t}$ are again linearly combined with the same $\{\boldsymbol{F}_{l,m}\}_m$, to obtain the desired estimate $\boldsymbol{p}_l$. Similar to the DL case, the analog combiners are taken from the columns of a Discrete Fourier Transform (DFT) matrix, i.e, $[\boldsymbol{F}_{l,1}, ..., \boldsymbol{F}_{l,K_t}] = \boldsymbol{D}_t$. The process for the UL is also shown in Table 3.4. We combine its steps to rewrite $\boldsymbol{p}_l$ as,

$$\boldsymbol{p}_l = \left(\sum_{m=1}^{K_t} \boldsymbol{F}_{l,m}\boldsymbol{F}_{l,m}^\dagger\right) \boldsymbol{H}^\dagger(d\tilde{\boldsymbol{w}}_l\tilde{u}_l) = d\boldsymbol{H}^\dagger\tilde{\boldsymbol{w}}_l\tilde{u}_l \tag{3.3.12}$$

[2]When sending $\tilde{\boldsymbol{f}}_l\tilde{g}_l$ over the DL, we can use d RF chains, i.e.,

$$\boldsymbol{F}_l\boldsymbol{G}_l\ \mathbf{1}_d = [\tilde{\boldsymbol{f}}_l, \cdots, \tilde{\boldsymbol{f}}_l]\ \text{diag}(\tilde{g}_l, \cdots, \tilde{g}_l)\ \mathbf{1}_d = d\tilde{\boldsymbol{f}}_l\tilde{g}_l$$

thereby resulting in an array gain factor of d. Moreover, since we know from (3.3.4) that $\|\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2^2 \leq 1$, indeed this transmission scheme satisfies the power constraint. We also make use of this observation in the UL sounding.

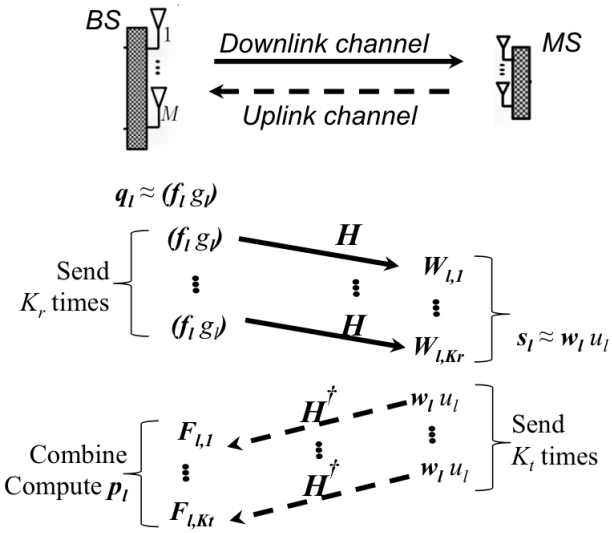

Figure 3.3: Repetition-aided (RAID) echoing for the hybrid analog-digital architecture

At the output of the RAID procedure, the BS has the following $\boldsymbol{p}_l$,

$$\begin{aligned}\boldsymbol{p}_l &= d\boldsymbol{H}^\dagger \tilde{\boldsymbol{w}}_l \tilde{u}_l = d\boldsymbol{H}^\dagger(\tilde{\boldsymbol{s}}_l - \boldsymbol{e}_l^{(r)}) = d\boldsymbol{H}^\dagger(d\boldsymbol{H}\tilde{\boldsymbol{f}}_l \tilde{g}_l - \boldsymbol{e}_l^{(r)}) \\ &= d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l - d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{e}_l^{(t)} - d\boldsymbol{H}^\dagger\boldsymbol{e}_l^{(r)} \end{aligned} \tag{3.3.13}$$

Note that $\boldsymbol{e}_l^{(t)} = \boldsymbol{q}_l - \tilde{\boldsymbol{f}}_l\tilde{g}_l$ (resp. $\boldsymbol{e}_l^{(r)} = \tilde{\boldsymbol{s}}_l - \tilde{\boldsymbol{w}}_l\tilde{u}_l$) is the error emanating from approximating $\boldsymbol{q}_l$ (resp. $\tilde{\boldsymbol{s}}_l$) at the BS (resp. MS), that we dub *BS-side* (resp. MS-side) *decomposition-induced distortion (DID)*. It is quite insightful to compare $\boldsymbol{p}_l$ in the latter equation with (3.3.8). We can clearly see that impairments originating from processing the received signals with both $\boldsymbol{W}_l$ and $\boldsymbol{F}_l$, have completely been suppressed. In (3.3.13), $\boldsymbol{p}_l$ indeed is the desired estimate, i.e., $\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l$, corrupted by distortions emanating from the BS-side decomposition, $\boldsymbol{e}_l^{(t)}$, and the MS side decomposition, $\boldsymbol{e}_l^{(r)}$ (both investigated later in the next subsection). Both UL and DL phases of he process are illustrated in Fig. 3.3, and detailed in Table 3.4.

Remark 3.4. Note that employing this process reduces the hybrid analog-digital architecture into a conventional MIMO channel: any transmitted vector in the DL, $(\tilde{\boldsymbol{f}}_l\tilde{g}_l)$, can be received in a "MIMO-like" fashion, as seen from (3.3.11), at a cost of K_r channel uses (the same holds for the UL, as seen from (3.3.12)).

It can be seen from the above, that in the DL (resp. UL), d RF chains are active at the BS (resp. MS), while all r RF chains are used at the MS (resp. BS), to minimize the overhead. With this in mind, it can be seen that the associated

// *DL phase*
$\boldsymbol{q}_l = \tilde{\boldsymbol{f}}_l \tilde{g}_l + \boldsymbol{e}_l^{(t)}$
$\boldsymbol{s}_{l,k} = \boldsymbol{W}_{l,k}^{\dagger} \boldsymbol{H}(d\tilde{\boldsymbol{f}}_l \tilde{g}_l),\ \forall k \in \{K_r\}$
$\tilde{\boldsymbol{s}}_l = \sum_{k=1}^{K_r} \boldsymbol{W}_{l,k} \boldsymbol{s}_{l,k}$
// *UL phase*
$\tilde{\boldsymbol{s}}_l = \tilde{\boldsymbol{w}}_l \tilde{u}_l + \boldsymbol{e}_l^{(r)}$
$\boldsymbol{z}_{l,m} = \boldsymbol{F}_{l,m}^{\dagger} \boldsymbol{H}^{\dagger}(d\tilde{\boldsymbol{w}}_l \tilde{u}_l),\ \forall m \in \{K_t\}$
$\boldsymbol{p}_l = \sum_{m=1}^{K_t} \boldsymbol{F}_{l,m} \boldsymbol{z}_{l,m}$

Table 3.4: Repetition-Aided (RAID) echoing

overhead with each echoing, $\Omega = (M+N)/r$ (channel uses), will decrease as more RF chains are used.

Imperfect Compensation of Analog-Processing Impairments

Though the above method perfectly removes all artifacts of analog processing, the overhead is proportional to $(M+N)/r$. A natural question is whether a similar result can still be achieved when $\boldsymbol{D}_r$ and $\boldsymbol{D}_t$ are truncated matrices i.e. when $K_r < N/r$ and $K_t < M/r$. Perfect cancellation of API relies on a careful choice of the analog precoder/combiner for each measurement, by picking $\{\boldsymbol{W}_{l,k}\}_{k=1}^{K_r}$ and $\{\boldsymbol{F}_{l,m}\}_{m=1}^{K_t}$ to span all the columns of (square) DFT matrices. We investigate the effect of picking $\boldsymbol{D}_r$ and $\boldsymbol{D}_t$ as truncated matrices, i.e. when $K_r < N/r$ and $K_t < M/r$. Focusing our discussion on just analog precoders for brevity, we seek to find a (tall) matrix $\tilde{\boldsymbol{D}}_t \in \mathbb{C}^{M \times (\eta M)}$, $\eta < 1$, such that,

$$\begin{cases} \min\limits_{\tilde{\boldsymbol{D}}_t} \| \frac{1}{M} \boldsymbol{I}_M - \tilde{\boldsymbol{D}}_t \tilde{\boldsymbol{D}}_t^{\dagger} \|_F^2 \\ \text{s. t.} \quad \tilde{\boldsymbol{D}}_t \in \mathcal{S}_{M,\ \eta M}\ . \end{cases} \tag{3.3.14}$$

Due to the apparent difficulty of the problem, one can resort to *stochastic optimization* tools, e.g. simulated annealing: this approach is ideal for the design of $\tilde{\boldsymbol{D}}_t$ (and $\tilde{\boldsymbol{D}}_r$ as well), since it is completely independent of all parameters (except M, N and η), and can thus be computed off-line and stored for later use. Then, the resulting overhead would be reduced to $\Omega = \eta \frac{M+N}{r}$.

Let $\boldsymbol{\Theta}_l$ be the phase of $\tilde{\boldsymbol{D}}_t$ at iteration l. N_c candidates for the phase update are generated, by randomly perturbing each element in $\boldsymbol{\Theta}_l$, i.e.

$$[\boldsymbol{B}_n]_{i,k} = [\boldsymbol{\Theta}_l]_{i,k} + \alpha_l\ u_{i,k}, \forall (i,k),\ n = 1, ..., N_c,$$

where $u_{i,k}$ is a uniformly distributed random variable over $[-\pi,\ \pi]$, and $\lim\limits_{l \to \infty} \alpha_l = 0$. The candidate solution that yields the best value is selected. The algorithm shown in Table 3.5 can be used to find $\tilde{\boldsymbol{D}}_t$. Then, using exactly the same method, one

Set $\alpha = 1$, $\delta < 1$, $\boldsymbol{\Theta}_0$ random
for $l = 0, 2, ..., I-1$ **do**
  $\alpha = \delta\alpha$
  for $n = 1, ..., N_c$ **do**
    $[\boldsymbol{B}_n]_{i,k} = [\boldsymbol{\Theta}_l]_{i,k} + \alpha\ u_{i,k}, \quad \forall(i,k)$
    $\boldsymbol{C}_n = (1/\sqrt{M}).e^{j\boldsymbol{B}_n}$
  end for
  $n^\star = \underset{n}{\text{argmax}}\ \|\frac{1}{M}\boldsymbol{I}_M - C_n C_n^\dagger\|_F^2$
  $\boldsymbol{\Theta}_{l+1} = \boldsymbol{B}_{n^\star}$
end for
$\tilde{\boldsymbol{D}}_t = (1/\sqrt{M}).e^{j\boldsymbol{\Theta}_I}$

Table 3.5: Random Phase Search

can also design $\tilde{\boldsymbol{D}}_r \in \mathbb{C}^{N\times(\eta N)}$ by formulating a similar problem as (3.3.14).
Further investigations along this line are outside the scope of this work, but we opted to include them briefly, for completeness.

3.3.3 Proposed Algorithms

Combining the results of the previous subsections, we can now formulate our algorithm for Subspace Estimation and Decomposition (SED) for the hybrid analog-digital architecture (shown in Algorithm 1): estimates of the right / left singular subspaces, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\boldsymbol{\Phi}}_1$, can be obtained by using the SE-ARN procedure (Sect. 3.2), keeping in mind that the *echoing phase (Steps 1.a and 1.b) is now replaced by the RAID echoing procedure* (Table 3.4. Then, the multi-dimensional subspace decomposition procedure, BCD-SD in Sect. 3.3.1, is then used to approximate each of the estimated singular spaces, by a cascade of analog and digital precoder/combiner. We highlight a desirable feature for the SED algorithm: the subspace estimation mechanism is totally decoupled from the subspace decomposition part, and thus any of the latter parts can be substituted, if desired.

Algorithm 1 Subspace Estimation and Decomposition (SED) for Hybrid Analog-Digital Architecture

// *Estimate* $\tilde{\boldsymbol{\Gamma}}_1$ *and* $\tilde{\boldsymbol{\Phi}}_1$
$\tilde{\boldsymbol{\Gamma}}_1, \tilde{\boldsymbol{\Sigma}}_1$ = SE-ARN ($\boldsymbol{H}$, d)
$\tilde{\boldsymbol{\Phi}}_1$ = SE-ARN ($\boldsymbol{H}^\dagger$, d)
// *Decompose* $\tilde{\boldsymbol{\Gamma}}_1$ *and* $\tilde{\boldsymbol{\Phi}}_1$
[$\boldsymbol{F}$, $\boldsymbol{G}$] = BCD-SD ($\tilde{\boldsymbol{\Gamma}}_1$, ρ)
[$\boldsymbol{W}$, $\boldsymbol{U}$] = BCD-SD ($\tilde{\boldsymbol{\Phi}}_1$, ρ)
Perform waterfilling on $\tilde{\boldsymbol{\Sigma}}_1$

Note that previously proposed algorithms within this context such as the PM and TQR in [DCG04], are no longer applicable here: indeed both rely on the MS being able to amplify-and-forward its received signal at the antennas - clearly this modus operandi cannot be supported by the hybrid analog-digital architecture. Interestingly, it is possible to apply elements from the RAID echoing structure that we developed, effectively modifying the original echoing structure of the latter schemes, and adapting them to the hybrid analog-digital architecture (as shown in Algorithm 2).

Algorithm 2 Modified Two-way QR (MTQR) for Hybrid Analog-Digital Architecture

for $l = 1, 2, ..., I$ **do**
  // Decompose each column of $\boldsymbol{X}$
    $[\boldsymbol{X}]_n \approx \tilde{\boldsymbol{f}}_n \tilde{g}_n, \ \forall n \in \{d\}$ (using Lemma 3.3.1)
    $\tilde{\boldsymbol{X}} = [\ \tilde{\boldsymbol{f}}_1 \tilde{g}_1\ , \cdots,\ \tilde{\boldsymbol{f}}_d \tilde{g}_d\ ]$
  // Send $\tilde{\boldsymbol{X}}$ *in DL, one column at a time*
    $\boldsymbol{T}_k = \boldsymbol{W}_k^{\dagger} \boldsymbol{H} \tilde{\boldsymbol{X}}, \ \forall k \in \{K_r\}$
    $\boldsymbol{Y} = \sum_{k=1}^{K_r} \boldsymbol{W}_k \boldsymbol{T}_k\ ; \boldsymbol{Y} = \mathrm{qr}(\boldsymbol{Y})$
  // Decompose of $\boldsymbol{Y}$
    $[\boldsymbol{Y}]_n \approx \tilde{\boldsymbol{w}}_n \tilde{u}_n, \ \forall n \in \{d\}$ (using Lemma 3.3.1)
    $\tilde{\boldsymbol{Y}} = [\ \tilde{\boldsymbol{w}}_1 \tilde{u}_1\ , \cdots,\ \tilde{\boldsymbol{w}}_d \tilde{u}_d\ ]$
  // Send $\tilde{\boldsymbol{Y}}$ *in UL, one column at a time*
    $\boldsymbol{S}_k = \boldsymbol{F}_k^{\dagger} \boldsymbol{H}^{\dagger} \tilde{\boldsymbol{Y}}, \ \forall k \in \{K_t\}$
    $\boldsymbol{Z} = \sum_{k=1}^{K_t} \boldsymbol{F}_k \boldsymbol{S}_k\ ; \boldsymbol{X} = \mathrm{qr}(\boldsymbol{Z})$
end for

Operationally, the proposed MTQR algorithm is the same as the Two-way QR (TQR) in [DCG04], whereby $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}_1$ are obtained iteratively: as $I \to \infty$, $\boldsymbol{X} \to \boldsymbol{\Gamma}_1$ (at BS) and $\boldsymbol{Y} \to \boldsymbol{\Phi}_1$ (at MS). At each iteration of the algorithm, the BS sends $\boldsymbol{X}$ in the downlink, and the QR algorithm is applied to the received signal. Then, the resulting signal is sent by the MS in the uplink, and the QR algorithm is applied at the BS to form $\boldsymbol{Y}$. While TQR assumes fully digital MIMO transmission, our contribution is to apply the RAID scheme, to make the transmission compatible with the hybrid analog-digital systems.

3.3.4 Bounds on Eigenvalue Perturbation

It can be clearly seen that the iterative nature of Algorithm 2 makes the application of Lemma 3.2.2, to quantify the impact of decomposition and approximation errors, not possible. On the other hand, for Algorithm 1, the fact that each $\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{q}_l$ is only corrupted by two sources of DID, $\boldsymbol{e}_l^{(r)}$ and $\boldsymbol{e}_l^{(r)}$, makes the latter possible. With that in mind, we specialize the result of Sect. 3.2.2 and Lemma 3.2.2 (developed for generic MIMO systems) to the case of Algorithm 1 in the hybrid analog-digital

architecture. We thus relate the eigenvalues of $\boldsymbol{T}_m$ at the output of SE-ARN, to the dominant eigenvalues of $\boldsymbol{C}_m$, and consequently of $\boldsymbol{A}$ (Sect.3.2.2).

Corollary 3.3.1. *Every eigenvalue* $\tilde{\lambda}$ *of* $\boldsymbol{T}_m$ *satisfies*

$$|\tilde{\lambda} - \lambda| \leq m\|\boldsymbol{H}\|_F^2 \ (3 + \frac{1}{d\|\boldsymbol{H}\|_F})$$

where λ *in an eigenvalue of* $\boldsymbol{C}_m$.

Proof. Refer to Appendix 3.6.5 □

Moreover, recall that as $m \to M$, λ is an eigenvalue of $\boldsymbol{A}$ (Lemma 3.2.1 - P3). Thus, this result directly relates the eigenvalues of $\boldsymbol{T}_m$, to that of $\boldsymbol{A}$: though this holds asymptotically in m, our simulations will show that good approximations can still be obtained, even for $m \ll M$. Note that we have ignored the effect of DID compensation, within the RAID echoing process, for convenience. As a result, the above bound is a "pessimistic" performance measure.

3.3.5 Practical Implementation Aspects

We evaluate the *communication overhead* of both schemes, in number of channel uses, keeping in mind that the actual overhead will be dominated by the latter. Algorithm 1 requires K_t+K_r channel uses per iteration, to estimate $\tilde{\boldsymbol{\Gamma}}_1$, and K_t+K_r to estimate $\tilde{\boldsymbol{\Phi}}_1$, for a total of

$$\Omega_{SED} = 2m \ \frac{M+N}{r}, \tag{3.3.15}$$

m being the number of iterations for the Arnoldi process. Letting I denote the number of iterations for MTQR, the number of channel uses required for Algorithm 2 is,

$$\Omega_{MTQR} = dI \ \frac{M+N}{r} \tag{3.3.16}$$

It should be emphasized here that our main focus in this work is to investigate the principle of subspace estimation employing numerical techniques, and through simulations describe the performance gain that can be expected by taking on such an approach. Hence, our major concern is not to investigate a stable and low-complexity technique that can be readily implemented in practice. We will, however, provide suggestions on what can be done to enhance the stability of the devised schemes, while admitting that many of the problems connected with practical implementation of the proposed method are subject to further study. Generally, it is known that the Arnoldi (and Lanczos) algorithm may suffer from numerical stability issues. Though analytically speaking, the basis $\boldsymbol{Q}_m$ is easily shown to be orthonormal, in practice, however, errors resulting from floating-point operations

lead to a loss in orthogonality (the extent to which it happens is dependent on the application) [Saa11, Sec. 7.3]. Moreover, for our algorithm, noise inherent to the echoing process will further amplify this effect. One of the widely adopted fixes for this matter is the Implicitly Restarted Arnoldi algorithm [Saa11, Sec. 7.3]. We did experiment with such an algorithm, and though it does enhance the numerical stability of the algorithm, the resulting overhead is increased by a large factor. This issue is critical for the SED algorithm (that employs the RAID echoing), since it renders real-world implementation quite impractical. Moreover, there are many problems connected with practical implementations of the Restarted Arnoldi method, that are subject to further study. Other methods that might enhance the stability the Arnoldi Iteration, such as deflation techniques, have been reported in [Sor96].

3.4 Numerical Results

3.4.1 Simulation Setup

In this section, we numerically evaluate the performance of our algorithms, in the context of a single-user MIMO link, using the above channel model. In what follows, we also assume that $M/r = 8$ and $N/r = 4$, i.e., as M, N increase, so does the number of RF chains. As per our discussion on the achievability of R (Remark 2.1), we use the following, as our performance metric in the simulations,

$$\tilde{R} = \log_2 \left| \boldsymbol{I}_d + \frac{1}{\sigma^2_{(r)}} \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H} \boldsymbol{F} \boldsymbol{G} \boldsymbol{G}^\dagger \boldsymbol{F}^\dagger \boldsymbol{H}^\dagger \boldsymbol{W} \boldsymbol{U} (\boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W} \boldsymbol{U})^{-1} \right| . \tag{3.4.1}$$

In that sense, $\tilde{R}$ is the 'user rate' that is based on the actual channel $\boldsymbol{H}$, and the precoders / combiners that are in turn designed based on the estimated channel.

Benchmarks/Upper bounds

We use the Adaptive Channel Estimation (ACE) method (Algorithm 2 in [AEALH14]) as a benchmark, to estimate the mmWave channel. It is based on sounding of *hierarchical codebooks* at the BS, feedback of the best codebook indexes by the MS, and finding the analog/digital precoders and combiners using OMP [EARAS$^+$14]. Moreover, the authors characterized the resulting communication overhead Ω_{ACE}, as a function of the codebook resolution. We used the corresponding MATLAB implementation that was provided by the authors. We adjust the number of iterations for both our proposed schemes and the codebook resolution of benchmark scheme, such that $\Omega_{SED} = \Omega_{MTQR} \triangleq \Omega_o \approx \Omega_{ACE}$. Note that we do not assume any quantization for phases of the RF filters. We also compare the performance of the algorithms against the "optimal performance", $R^\star$ in (2.2.2), where full CSIT/CSIR is assumed, fully digital precoding is employed, and the optimal precoders are used. All curves are averaged over 500 channel realizations.

Remark 3.5. Note that if one want to use "classical" pilot-based channel estimation to estimate the DL channel, i.e., a pilot sequence of minimum length M, then the same repetition-based framework that was used in RAID echoing, has to be used to cancel the effect of $\boldsymbol{W}$ from the effective channel estimate: it can be easily seen that the resulting total (both DL and UL) number of pilots slots would be $2MN/r^2$, thereby making the latter method infeasible.

3.4.2 Performance Evaluation

We start by investigating the performance of our schemes against the above benchmarks, for the case where $M = 128, N = 64, L = 3$, and $m = 3d$, for two cases: $d = 1$ and $d = 2$ where the resulting overhead is $\Omega_o = 72$ and $\Omega_o = 144$ channel uses, respectively. It can be seen from Fig. 3.4 that both *proposed schemes exhibit relatively similar performances*, that are in turn very *close to the optimal performance bound* $R^\star$ (especially above -10 dB). This indeed suggests that the multiplexing gain achieved by conventional MIMO systems can still be maintained in the hybrid analog-digital architecture, albeit at a much lower cost: the number of *required RF chains can be drastically decreased*, resulting in savings in terms of cost and power consumption. Moreover, we observe a sharp and significant performance gap between both our schemes and the benchmark from [AEALH14], over all SNR ranges (the gap being more significant in the low-SNR regime). We also evaluate the so-called optimal decomposition schemes [ZMK05, MRRGPH15] that can exactly decompose $\boldsymbol{\Gamma}_1$ into $\boldsymbol{FG}$ (discussed in Sec. IV). Thus, the curves labeled 'Optimal Decomp.' refer to the case where the optimal decomposition is used in conjunction with SED. Fig 3.4 clearly reveals that the ability to optimally decompose the estimated subspaces does not bring about additional gains. We note that the tiny mismatch between 'Optimal Decomp.' and Algorithm 1 is due to simulation resolution.

We attempt to shed light on the stability of the proposed algorithms, as the number of paths in the mmWave channel, L, increases (where we set $M = 64, N = 32, d = 2, m = 6$). For clarity we restrict the result to the low SNR regime. Though a degradation in the performance of both algorithms is expected, as L increases, Fig. 3.5 clearly indicates that the latter degradation is not quite significant. Though not visible here, our simulations show that this degradation is not present in the medium-to-high SNR region. As expected, this technique is best used for channels with a few paths, e.g., mmWave channels.

We investigate the performance of both SED and MTQR in terms of average subspace angle, $\theta = \mathbb{E}[\alpha(\boldsymbol{\Gamma}_1, \tilde{\boldsymbol{\Gamma}}_1)]$ where $\alpha(\boldsymbol{\Gamma}_1, \tilde{\boldsymbol{\Gamma}}_1)$ (radians) is defined as the subspace angle between $\boldsymbol{\Gamma}_1$ and $\tilde{\boldsymbol{\Gamma}}_1$ (implemented by computing the principal angles of the latter subspaces). As shown in Fig. 3.6, both schemes exhibit a similar behavior of better estimation accuracy, as the SNR increases.

Remark 3.6. Though the performance of Algorithm 2 seems to be better, Fig. 3.4-3.6 both suggest that this gap is quite narrow. Moreover, both algorithms seem to

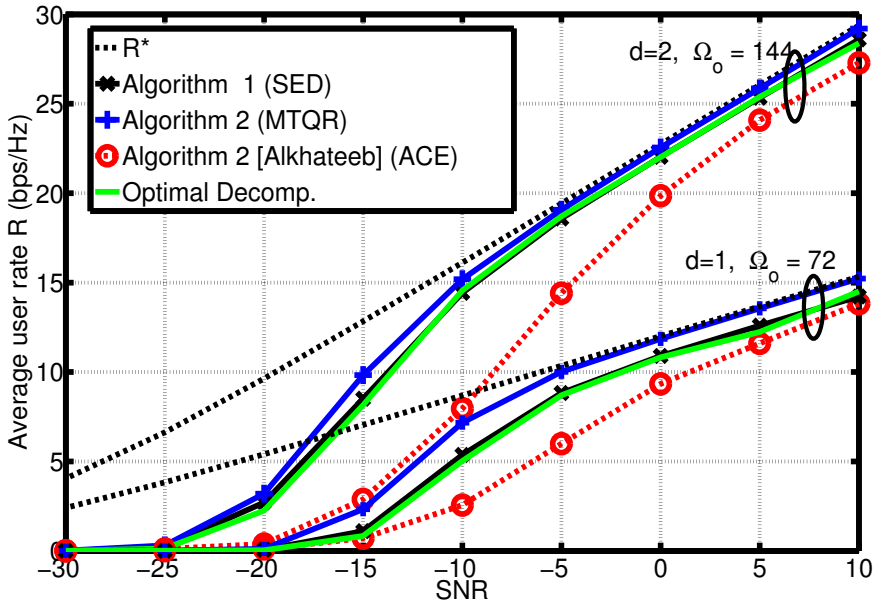

Figure 3.4: Average sum-rate of proposed schemes ($M = 128, N = 64, d = 2, L = 3, m = 6$)

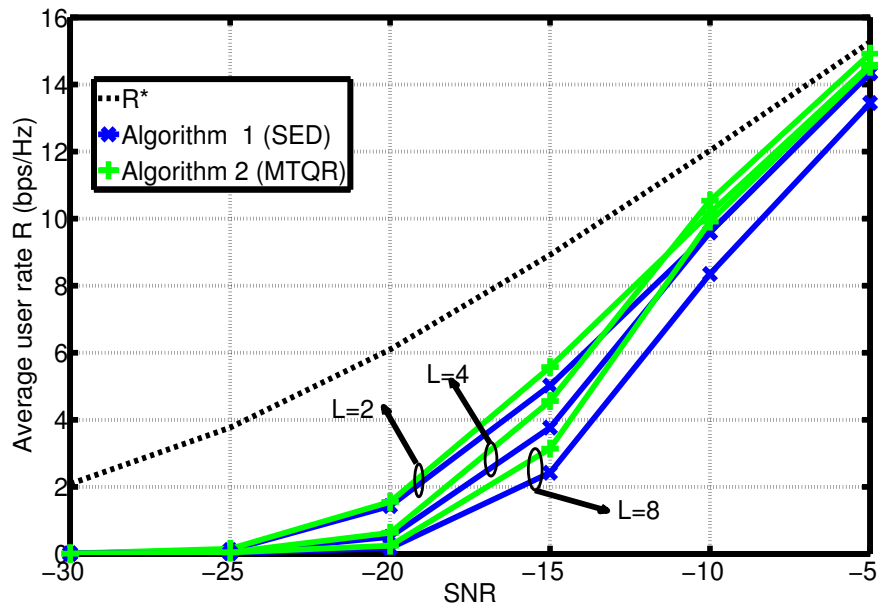

Figure 3.5: Effect of number of paths L, on the average user rate ($M = 64, N = 32, d = 2, m = 6$)

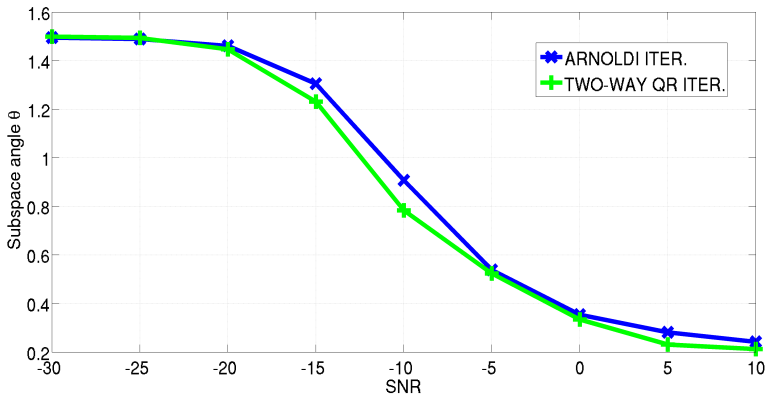

Figure 3.6: Average subspace angle ($M = 64, N = 32, d = 3, L = 4, m = 6$)

exhibit very similar behavior. With that in mind, and for the sake of clarify of our results, we opt to focus on Algorithm 1, the main object of investigation in this work.

We next investigate its scalability: we scale up M and N (assuming $N = M/2$, for simplicity), while keeping everything else fixed, i.e., $d = 2, m = 6$, and consequently $\Omega_o = 144$. In doing that, we noticed that the complexity of the benchmark scheme [AEALH14] was *prohibitively high*, thus preventing us from investigating its scalability: we were unable to get any results for systems larger than 128×64 . On the other hand, both our algorithms exhibit no such problems since all the computations that they involve are matrix-vectors/matrix-matrix operations. Consequently, the *complexity gap between Algorithm 1 and the benchmark increases drastically, as* M, N grow.

Fig 3.7 clearly shows that Algorithm 1 is able to harness the significant array gain inherent to large antenna systems (by closely following the optimal performance bound, $R^\star$, with a small constant gap), while keeping the overhead remarkably small. Though the performance might not be good enough to offset the overhead, for the 16×8 case, it surely does for the 256×128. Moreover, note that the gap between the optimal performance and Algorithm 1 is quite small (across the entire SNR range) for small systems dimensions, and quite small even for large values of M (at high SNR). The key to this result is to have M/r and N/r fixed, as M, N increase.

We also evaluate the performance of Algorithm 1 in a more realistic manner, by adopting the Spatial Channel Model (SCM) detailed in [3GP11, SDS+05], and modifying its parameters to emulate mmWave channels: the number of paths is set to 4, the carrier frequency to 60 GHz, the BS/MS array is modified to implement ULAs, and an 'urban micro' scenario is selected, where a small Ω_o is desired. Fig. 3.8

shows the average performance of such a system, with $M = 64, N = 32, m = 2d$, for several values of d (each resulting in different values for Ω_o). Though both our algorithm and the benchmark exhibit similar performances for $d = 1$, this gap increases with d, e.g. for $d = 3$ this performance gap is quite significant. Moreover, we can clearly see that Algorithm 1 yields a relatively high throughput in this realistic simulation setting (especially for $d = 3$), while still keeping the overhead at a relatively low level.

Evidently, increasing m (the number of iterations for the Arnoldi) has the effect enhancing the estimation accuracy (and increasing the communication overhead as well (3.3.15)). The marginal improvement brought about by increasing m, is decreasing, and thus our simulations indicated that setting $2d \leq m \leq 3d$ provides a good trade-off.

3.4.3 Discussions

A few remarks are in order at this stage, regarding similarities and differences between our two proposed algorithms. As discussed in Remark 3.6, when the communication overhead is normalized, both SED and MTQR exhibit a similar behavior and performance profile, across the entire SNR range (with a relatively small performance gap): indeed they can be used interchangeably with no change at all in the operational requirements. However, as this work shows, we have an accurate analytical description of the behavior of SED: the Arnoldi algorithm was adapted to the subspace estimation part (with some analytical performance guarantees), and BCD-SD to mathematically describe the decomposition algorithm. In contrast, MTQR is a (heuristic) variation on the original TQR, whose behavior we have not modeled analytically.

One of the conclusions suggested by all the above results, is the fact that the low-SNR performance of the proposed schemes is rather poor. However, interestingly, Figs. 3.4-3.8 unambiguously point out that this is the case for the benchmark scheme as well (ACE in [EARAS$^+$14]): one might be tempted to conjecture at this point that this low-SNR behavior is an inherent aspect of mmWave channel estimation. Initial investigations reveal that, if more RF chains (more than r) can be employed during the RAID echoing phase, the low-SNR performance can be greatly boosted.

3.5 Conclusion

We proposed an algorithm for blindly estimating the left and right singular subspace of a mmWave MIMO channel, by exploiting channel reciprocity that is inherent to TDD systems. Though the algorithm is a perfect match for conventional (large) MIMO systems, we extended it to operate under the constraints dictated by the hybrid analog-digital architecture, and showed via simulations that it is a good fit for large MIMO channels, with low rank, e.g., mmWave channels. Finally, our simulations showed that a similar performance to the ideal case (fully digital perfect

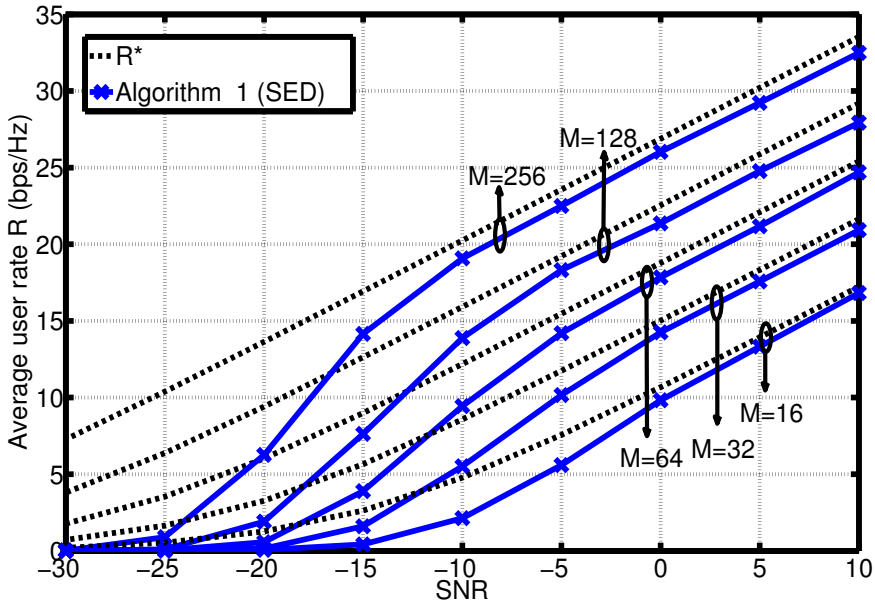

Figure 3.7: Average user-rate for different M, N ($N = M/2, d = 2, L = 4, m = 6, \Omega_o = 144$)

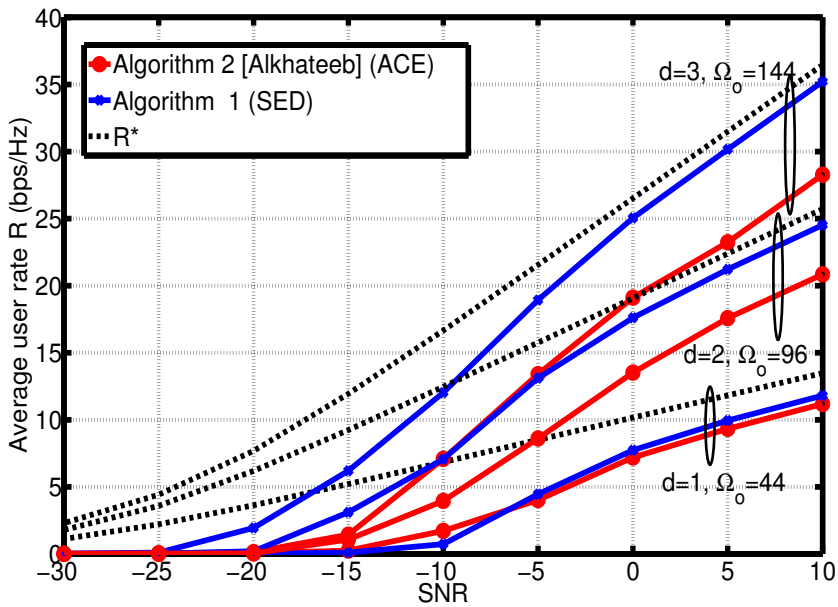

Figure 3.8: Average user-rate of proposed schemes over SCM channels ($M = 64, N = 32, m = 2d$)

CSI) can be achieved, with a only a few RF chains, thereby resulting in significant saving in energy and cost, over conventional MIMO systems.

3.6 Appendix

3.6.1 Proof of Lemma 3.2.1

$(P1)$: Combining steps (2.b) and (3.a) in the SE-ARN procedure, we write,

$$\boldsymbol{A}\boldsymbol{q}_l + \tilde{\boldsymbol{w}}_l = \sum_{i=1}^{l+1} [\tilde{\boldsymbol{T}}_m]_{i,l}\ \boldsymbol{q}_i + \sum_{i=1}^{l} [\boldsymbol{E}_m]_{i,l}\ \boldsymbol{q}_i\ ,\ \forall l \in \{m\},$$

We can rewrite the latter equation in matrix form, using the definitions of $\tilde{\boldsymbol{T}}_m, \tilde{\boldsymbol{W}}_m$ given in (3.2.3),

$$\boldsymbol{A}\boldsymbol{Q}_m + \tilde{\boldsymbol{W}}_m = \boldsymbol{Q}_m\tilde{\boldsymbol{T}}_m + [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1}\boldsymbol{b}_m^\dagger + \boldsymbol{Q}_m\boldsymbol{E}_m \tag{3.6.1}$$

where $\boldsymbol{b}_m$ is the m^{th} elementary vector, and $\boldsymbol{E}_m = [\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m]_U$. We can further simplify the above, using the fact that $\boldsymbol{Q}_m^\dagger\boldsymbol{Q}_m = \boldsymbol{I}_m$ and $\boldsymbol{Q}_m^\dagger\boldsymbol{q}_{m+1} = \boldsymbol{o}$,

$$\boldsymbol{Q}_m^\dagger\boldsymbol{A}\boldsymbol{Q}_m + \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m = \tilde{\boldsymbol{T}}_m + \boldsymbol{E}_m$$

Using the definition of $\boldsymbol{E}_m$, we write,

$$\begin{aligned}\boldsymbol{Q}_m^\dagger\boldsymbol{A}\boldsymbol{Q}_m &= \tilde{\boldsymbol{T}}_m + [\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m]_U - \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m \\ &= \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m \triangleq \boldsymbol{C}_m\end{aligned}$$

where $\tilde{\boldsymbol{E}}_m = [\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m]_{SL}$, as defined in (3.2.3).

$(P2)$: Noting that $\tilde{\boldsymbol{T}}_m + \boldsymbol{E}_m = \boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m$, we rewrite (3.6.1) as,

$$\boldsymbol{A}\boldsymbol{Q}_m - \boldsymbol{Q}_m\boldsymbol{C}_m = [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1}\boldsymbol{b}_m^\dagger - (\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger)\tilde{\boldsymbol{W}}_m$$

Multiplying the latter equation by $\boldsymbol{s}_i^{(m)}$, and using the fact that $\boldsymbol{C}_m\boldsymbol{s}_i^{(m)} = \lambda_i^{(m)}\boldsymbol{s}_i^{(m)}$, and $\boldsymbol{Q}_m\boldsymbol{s}_i^{(m)} = \boldsymbol{\Theta}_i^{(m)}$

$$\begin{aligned}\boldsymbol{A}\boldsymbol{\Theta}_i^{(m)} - &\lambda_i^{(m)}\boldsymbol{\Theta}_i^{(m)} \\ &= [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1}\boldsymbol{b}_m^\dagger\boldsymbol{s}_i^{(m)} - (\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger)\tilde{\boldsymbol{W}}_m\boldsymbol{s}_i^{(m)}\end{aligned}$$

Finally, the desired residual is upper bounded as,

$$\begin{aligned}&\|\boldsymbol{A}\boldsymbol{\Theta}_i^{(m)} - \lambda_i^{(m)}\boldsymbol{\Theta}_i^{(m)}\|_2^2 \\ &\quad\le ([\tilde{\boldsymbol{T}}_m]_{m+1,m}|\boldsymbol{b}_m^\dagger\boldsymbol{s}_i^{(m)}|)^2 + \|(\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger)\tilde{\boldsymbol{W}}_m\boldsymbol{s}_i^{(m)}\|_F^2 \\ &\quad\le ([\tilde{\boldsymbol{T}}_m]_{m+1,m}|[\boldsymbol{s}_i^{(m)}]_m|)^2 + \|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2\|\tilde{\boldsymbol{W}}_m\|_F^2\end{aligned}$$

where the last inequality follows from $\|\boldsymbol{B}_1\boldsymbol{B}_2\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{B}_1\|_F^2.\|\boldsymbol{B}_2\|_F^2.\|\boldsymbol{x}\|_2^2$

$(P3)$: The proof immediately follows by noting that $\|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2 \to 0$ and $[\tilde{\boldsymbol{T}}_m]_{m+1,m} \to 0$, as $m \to M$, thereby implying that $\|\boldsymbol{A}\boldsymbol{\Theta}_i^{(M)} - \lambda_i^{(M)}\boldsymbol{\Theta}_i^{(M)}\|_2^2 \ll 1$.

3.6.2 Proof of Lemma 3.2.2

The proof follows from a direct application of the Bauer-Fike Theorem [GVL96, Th. 7.2.2]. Let $\boldsymbol{C}_m = \boldsymbol{S}_m\boldsymbol{\Lambda}_m\boldsymbol{S}_m^{-1}$ be the diagonalizable matrix in question, and $\boldsymbol{T}_m = \boldsymbol{C}_m + \boldsymbol{P}_m$ the "perturbed" matrix. Then, every eigenvalue $\tilde{\lambda}$ of $\boldsymbol{T}_m$ satisfies,

$$|\tilde{\lambda} - \lambda|^2 \leq \|\boldsymbol{S}_m\|_2^2.\|\boldsymbol{S}_m^{-1}\|_2^2.\|\boldsymbol{P}_m\|_2^2 = \|\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m\|_2^2$$

where λ is an eigenvalue of $\boldsymbol{C}_m$, and $\|\boldsymbol{B}\|_2 \triangleq \sigma_{\max}(\boldsymbol{B})$ is the vector-induced matrix 2-norm. The last equality follows from the fact that $\boldsymbol{S}_m$ is unitary, as discussed in Lemma 3.2.1. Using the fact that $\|\boldsymbol{B}\|_2 \leq \|\boldsymbol{B}\|_F$, we rewrite the last equation,

$$|\tilde{\lambda} - \lambda|^2 \leq \|\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m\|_F^2 \leq \|\boldsymbol{Q}_m\|_F^2\|\tilde{\boldsymbol{W}}_m\|_F^2 = m\|\tilde{\boldsymbol{W}}_m\|_F^2$$

This concludes the proof.

3.6.3 Proof of Lemma 3.3.1

Note that there is not loss in optimality by assuming the $g \in \mathbb{R}_+$. Moreover, exploiting the structure of h_o, the globally optimal solution can be found by optimizing for $\boldsymbol{f}$, assuming g is fixed (and vice) versa, i.e.,

$$\begin{aligned}
\boldsymbol{f}^\star \overset{\triangle}{=} &\underset{\boldsymbol{f}}{\text{argmin}}\ g^2(\boldsymbol{f}^\dagger\boldsymbol{f}) - 2g\Re(\boldsymbol{f}^\dagger\tilde{\boldsymbol{\gamma}}_1),\ \text{s. t. } [\boldsymbol{f}]_i = 1/\sqrt{M}\ e^{j\phi_i} \\
\overset{(a)}{\Leftrightarrow}\ \{\phi_i^\star\} = &\underset{\{\phi_i\}}{\text{argmax}}\ 1/\sqrt{M}\ \Re\left(\sum_{i=1}^{M} r_i\ e^{j(\theta_i-\phi_i)}\right) \\
\{\phi_i^\star\} = &\underset{\{\phi_i\}}{\text{argmax}}\ \sum_{i=1}^{M}\Re\left(e^{j(\theta_i-\phi_i)}\right) = \{\theta_i\}
\end{aligned}$$

where (a) follows from applying the one-to-one mapping $[\boldsymbol{f}]_i \to 1/\sqrt{M}\ e^{j\phi_i}, \forall i$. Thus, $[\boldsymbol{f}^\star]_i = 1/\sqrt{M}\ e^{j\theta_i}, \forall i$. Plugging $\boldsymbol{f}^\star$ into the original problem, the optimization of g is a simple unconstrained quadratic problem,

$$g^\star \overset{\triangle}{=} \underset{g}{\text{argmin}}\ g^2 - 2g(\|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M}) = \|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M} \tag{3.6.2}$$

3.6.4 Proof of Proposition 3.3.1

Since $\boldsymbol{Y} \in \mathcal{S}_{M,d}$ by definition (i.e., $|[\boldsymbol{Y}]_{i,k}| = 1/\sqrt{M}$) the problem just reduces to finding the phase of each element in $\boldsymbol{Y}$. Thus,

$$\begin{aligned}\boldsymbol{Y} = \Pi_{\mathcal{S}}[\boldsymbol{X}] &\stackrel{\triangle}{=} \underset{\boldsymbol{U}\in\mathcal{S}_{M,d}}{\operatorname{argmin}} \|\boldsymbol{U}-\boldsymbol{X}\|_F^2 \\ &\stackrel{(a)}{\Leftrightarrow} \underset{\{\theta_{i,k}\}}{\operatorname{argmin}} \sum_{i,k} |(1/\sqrt{M})e^{j\theta_{i,k}} - x_{ik}e^{j\phi_{i,k}}|^2 \\ &\Leftrightarrow \{\theta^{\star}_{i,k}\} = \{\phi^{\star}_{i,k}\}\end{aligned}$$

where (a) follows from the fact that $\boldsymbol{U}_{i,k} = (1/\sqrt{M})e^{j\theta_{i,k}}, \forall \boldsymbol{U} \in \mathcal{S}_{M,d}$. Thus, we conclude that $[\boldsymbol{Y}]_{i,k} = (1/\sqrt{M})\ e^{j\phi_{i,k}}, \forall (i,k)$. Furthermore, it follows from this formulation that this projection is unique (despite the non-convexity of $\mathcal{S}_{M,d}$).

3.6.5 Proof of Corrollary 3.3.1

The proof consists of finding a closed-from expression for $\tilde{\boldsymbol{W}}_m$ as a function of $\boldsymbol{e}_l^{(t)}$ and $\boldsymbol{e}_l^{(r)}$, and applying the result of Lemma 3.2.2. Note that $\tilde{\boldsymbol{w}}_l$ in (3.2.2) can represent any distortion, and by comparing $\boldsymbol{p}_l$ in both (3.2.2) and (3.3.13), can infer that $\tilde{\boldsymbol{w}}_l = -\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{e}_l^{(t)} - (1/d)\boldsymbol{H}^{\dagger}\boldsymbol{e}_l^{(r)}$. Thus, $\tilde{\boldsymbol{W}}_m$ in (3.2.3) can be written as,

$$\begin{aligned}\tilde{\boldsymbol{W}}_m &= -\boldsymbol{H}^{\dagger}\boldsymbol{H}[\boldsymbol{e}_1^{(t)},\cdots,\boldsymbol{e}_m^{(t)}] - (1/d)\boldsymbol{H}^{\dagger}[\boldsymbol{e}_1^{(r)},\cdots,\boldsymbol{e}_m^{(r)}] \\ &\triangleq -\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{E}^{(t)} - (1/d)\boldsymbol{H}^{\dagger}\boldsymbol{E}^{(r)}\end{aligned}$$

Then using properties of the Frobenius norm,

$$\|\tilde{\boldsymbol{W}}_m\|_F \leq \|\boldsymbol{H}\|_F^2\|\boldsymbol{E}^{(t)}\|_F + (1/d)\|\boldsymbol{H}\|_F\|\boldsymbol{E}^{(r)}\|_F \tag{3.6.3}$$

On the other hand, recall that $\boldsymbol{e}_l^{(t)} = \boldsymbol{q}_l - \tilde{\boldsymbol{f}}_l\tilde{g}_l$ and $\boldsymbol{e}_l^{(r)} = \tilde{\boldsymbol{s}}_l - \tilde{\boldsymbol{w}}_l\tilde{u}_l$. Thus, using the results of Sec. 3.3.1,

$$\begin{aligned}\|\boldsymbol{e}_l^{(t)}\|_2 &\leq \|\boldsymbol{q}_l\|_2 + \|\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2 \leq 2 \\ \|\boldsymbol{e}_l^{(r)}\|_2 &\leq \|d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2 + \|\tilde{\boldsymbol{w}}_l\tilde{u}_l\|_2 \leq 1 + d\|\boldsymbol{H}\|_F\end{aligned}$$

and it follows that

$$\|\boldsymbol{E}^{(t)}\|_F \leq 2\sqrt{m},\ \|\boldsymbol{E}^{(r)}\|_F \leq \sqrt{m}(1+d\|\boldsymbol{H}\|_F) \tag{3.6.4}$$

The upper bound follows by combining (3.6.3) and (3.6.4).

Part II

Distributed Utility Optimization

Chapter 4

Preliminaries

4.1 Interference Management in Multiuser MIMO Networks

Going from single cell to multi-cell settings, interference has been widely recognized as the limiting factor on the sum-rate performance. Many early works characterized this using the analytical framework of *degrees-of-freedom* [CJ08]: the maximum number of interference-free signaling dimensions in a given network. It was also shown that the maximum sum degrees-of-freedom (DoFs) of MIMO interference channels are a high SNR approximation of the capacity [CJ08] - the maximum performance that can be achieved. Graphically, the sum DoFs corresponds to the high-SNR slope of the sum-rate vs SNR curve. With that in mind, it is also well-known that the presence of unsuppressed interference leads to a collapse in the DoFs of the network. The seminal works by M.A. Maddah-Ali, A.S. Motahari and A.K. Khandani on one hand, and V. Cadambe, K. Gomadam and S.A. Jafar independently, on the idea of Interference Alignment in wireless communication, highlighted the intimate relation between IA and DoF maximization.

4.1.1 Coordination in Cellular Networks

The ideas of coordinating signals from multiple BSs, in view of mitigating interference, were considered much earlier, before the advent of CoMP. Such ideas were earlier considered under the name of Virtual MIMO and Network MIMO. The so-called Network MIMO concept was investigated in [KFVY06] and [ZCA+09], where full intra-cluster coordination (to enhance the sum-rate and limited inter-cluster coordination) was considered, for reducing interference for the cluster edge users. Though basic in nature, the idea of exploiting causally known interference for multi-cellular settings, dates as far back as [SZ01], where the "writing on dirty paper" approach was employed to cancel interference that is known at the transmitter but not to the receiver. CoMP introduced the idea of cooperation among the BSs to mitigate inter-cell interference [GHH+10], and has been usually identified

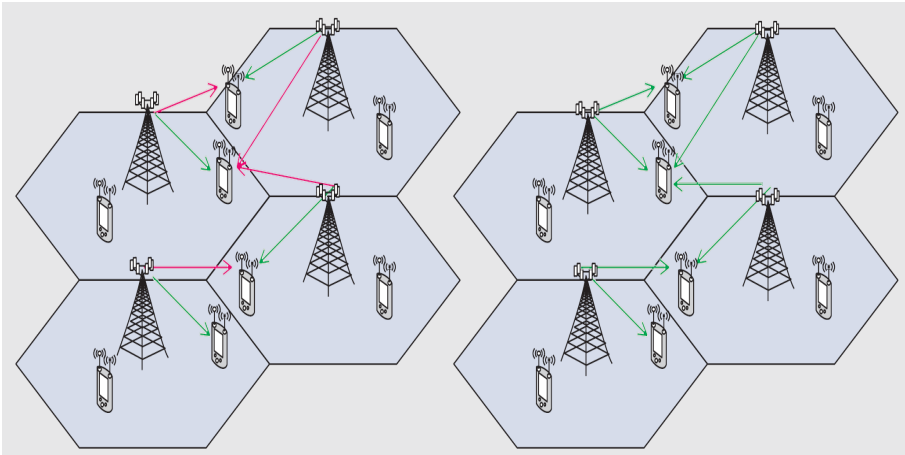

Figure 4.1: CoMP-CB (left) and CoMP-JT (right) (from [LSC+12])

with several operation modes, *Coordinated Beamforming* (CB) and *Joint Transmission* (JT), being the most relevant ones here. Though they are both precoding techniques, they are fundamentally different in their operation, and consequently, their inherent requirements and limitations [LSC+12].

CB is the setup where multiple BSs share their local CSI (i.e. channel information of each BS, to all the users), resulting in each BS having *global CSI of all the channels*. Moreover, the data of each of the users is only available at its serving BS, and *not shared* among the other BSs. One instance of CB is to design the precoding vectors at each BS in such a way that interfering signals lie in the null space of each user's desired signal subspace. This is illustrated in the left part of Fig 4.1 (where interfering transmissions are denoted in red): the precoding is done in such a way that each BS's transmission is orthogonal to all transmissions, by other BSs in the coordination area. On the other hand, JT - in addition to sharing of CSI among the BSs (such that *each BS has global CSI*), additionally requires the user data to be shared as well: each BS needs to know the *data of all the users* in the coordination area. The main idea behind JT is to design the precoding vectors at the BSs, in such a way that each user is served by *all* the BSs (as illustrated in the right part of Fig 4.1). Stated differently, JT requires signals from each BS to *align constructively* at each MS. As a result, interference in the entire coordination region is turned from destructive to constructive. Such ideas were initially put forth in [GHH+10].

Coordinated Beamforming

Coordination techniques that fall under the umbrella of CB are quite numerous and diverse. Such techniques were the focus of many works, dating as far back as [GCJ11]

and [SSB+09]. In essence, schemes falling under this category iteratively refine their respective cost functions, in a fully distributed manner, i.e., *only requiring local CSI at both users and BSs* and alleviating the need for any backhaul between transmitters, or for any centralized compute node. Despite the massive number of CB-type schemes falling under that category, they all employ the so-called framework of *forward-backward training* (detailed next, in Sect. 4.3.1).

Usually, such schemes can be categorized according to the metric that they optimize: such metrics mainly include (weighted) interference leakage [GCJ11], [GP11], (weighted) mean-squared error [SSB+09], [SRLH11], signal to interference-plus-noise ratio [GCJ11], [PH11], and (weighted) sum-rate [SGHP10a], [SRLH11] [NSGS10] (an insightful and comprehensive comparison of such schemes was done in [SSB+13]). Despite the fact that the latter methods attempt to solve a problem that is more generic than Interference Alignment (in a sense that they do not aim at suppressing interference completely), in many of the above cases, there indeed exists an intimate relation between the two: for instance, in the high-SNR sum-rate maximization problem, the precoder optimization problem reduces to finding transmit and receive filters, that satisfy the IA conditions (as formulated in [GCJ11]).

Interference Alignment

The concept of interference alignment in wireless communication was first presented in [MAMK08], for the MIMO X channel. However, the concept was clearly crystallized in [CJ08], for the K-user Interference Channel (IC). The simple approach to solve the problem of interference in the interference channel is *orthogonal* access in the time, frequency or spatial dimension (i.e. TDMA, FDMA, SDMA). As a result, the resources of the channel (time, bandwidth, antennas, etc..) are divided among the users equally, i.e. each receiver gets $1/K$ from the total resources. However with IA, every receiver can achieve $1/2$ of the channel resources, regardless of the number of users [CJ08]. This result is made possible by having every transmitter sacrifice *half* of its maximum signaling dimensions (time, frequency bands, antennas, etc...). Then, each transmitter-receiver pair can communicate over an *interference-free* channel, regardless of the number of interferers [CJ09]. The capacity of any channel (the summation of the rates achieved by all the users) can be approximated as follows [CJ08] : $C \approx d \log(1 + SNR)$ where d is referred to as the multiplexing gain or the degrees-of-freedom (DoF) of the channel.

4.2 System Model

In this section we outline several of the canonical channel models that arise in the context of multiuser communication and information theory (starting from the most generic one). We restrict our exposition to ones used throughout this part of the thesis.

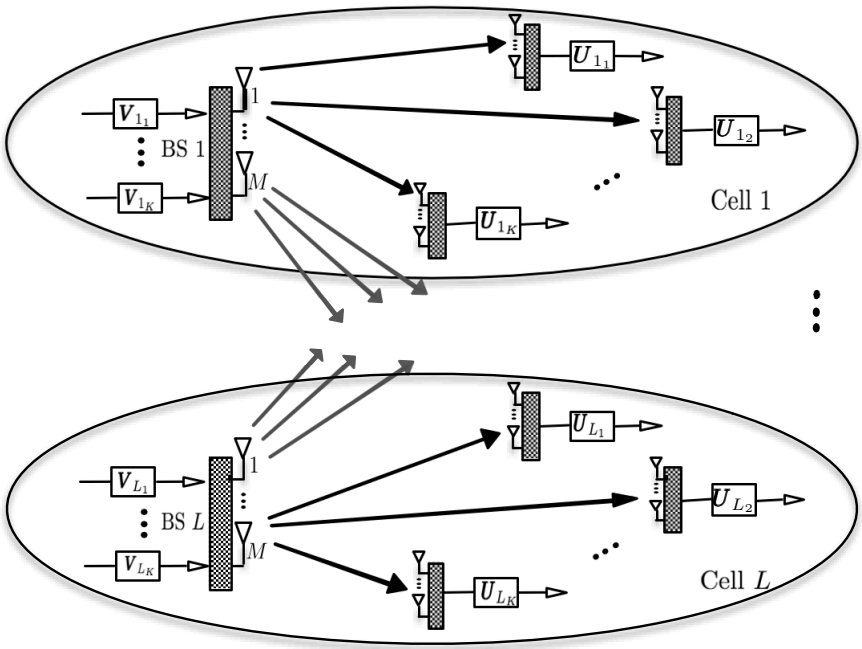

Figure 4.2: The MIMO Interring Broadcast Channel (MIMO IBC)

4.2.1 The MIMO Interfering Broadcast Channel

The MIMO Interfering Broadcast Channel (MIMO IBC) is a downlink communication scenario, consisting of a set of $\mathcal{L}$ cells (each having a BS), whereby each BS is serving multiple users, as shown in Fig. 4.2.

Let $\mathcal{L}$ be the set of BSs, and $\mathcal{K}_l$ be the set of MSs served by BS $l \in \mathcal{L}$. Moreover, denote by $\mathcal{I}$ the set of all users, i.e.,

$$\mathcal{I} = \{l_j \mid (l, j) \in \mathcal{L} \times \mathcal{K}_l\} \tag{4.2.1}$$

and user $l_j \in \mathcal{I}$ the jth user in cell l. $\boldsymbol{H}_{l,l_j}$ is the $M \times N$ MIMO channel from BS $l \in \mathcal{L}$, to user $l_j \in \mathcal{I}$ (assumed to have circularly symmetric i.i.d. complex random variables with zero mean and unit variance), $\boldsymbol{V}_{l_j} \in \mathbb{C}^{M \times d}$ is the d-dimensional transmit filter for user $l_j \in \mathcal{I}$. Then, the signal that is transmitted from BS l is $\sum_{j \in \mathcal{K}_l} \boldsymbol{V}_{l_j} \boldsymbol{x}_{l_j}$, and $\boldsymbol{x}_{l_j}$ is the d-dimensional zero-mean circularly symmetric complex Gaussian transmit signal for user $l_j \in \mathcal{I}$, such that $\mathbb{E}[\boldsymbol{x}_{l_j} \boldsymbol{x}_{l_j}^{\dagger}] = (\rho/d)\, \boldsymbol{I}_d$. Moreover, $\boldsymbol{U}_{l_j} \in \mathbb{C}^{N \times d}$ denotes the d-dimensional receive filter of user $l_j \in \mathcal{I}$ to linearly process

its received signal. The received signal, $\boldsymbol{y}_{l_j}$, for user $l_j \in \mathcal{I}$ is,

$$\boldsymbol{y}_{l_j} = \boldsymbol{H}_{l,l_j}\boldsymbol{V}_{l_j}\boldsymbol{x}_{l_j} + \sum_{\substack{i_k \in \mathcal{I} \\ i_k \neq l_j}} \boldsymbol{H}_{k,l_j}\boldsymbol{V}_{i_k}\boldsymbol{x}_{i_k} + \boldsymbol{n}_{l_j}, \tag{4.2.2}$$

and the received signal of user $l_j \in \mathcal{I}$, after linear filtering, is given by,

$$\hat{\boldsymbol{x}}_{l_j} = \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{H}_{l,l_j}\boldsymbol{V}_{l_j}\boldsymbol{x}_{l_j} + \sum_{\substack{i_k \in \mathcal{I} \\ i_k \neq l_j}} \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{H}_{k,l_j}\boldsymbol{V}_{i_k}\boldsymbol{x}_{i_k} + \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{n}_{l_j}, \tag{4.2.3}$$

where the firs terms represents that desired signal, and the second one both intra and inter-cell interference. We further define the signal and interference covariance matrices of user $l_j \in \mathcal{I}$, as follows,

$$\boldsymbol{R}_{l_j} = (\rho/d)\boldsymbol{H}_{l,l_j}\boldsymbol{V}_{l_j}\boldsymbol{V}_{l_j}^{\dagger}\boldsymbol{H}_{l,l_j}^{\dagger} \text{ and} \tag{4.2.4}$$

$$\boldsymbol{Q}_{l_j} = \sum_{i_k \in \mathcal{I}} \boldsymbol{H}_{k,l_j}\boldsymbol{V}_{i_k}\boldsymbol{V}_{i_k}^{\dagger}\boldsymbol{H}_{k,l_j}^{\dagger} - \boldsymbol{R}_{l_j}, \tag{4.2.5}$$

The MIMO IBC is the communication scenario that is investigated in Chap. 6.

Special Case: The MIMO Interference Channel The MIMO Interference Channel (MIMO IC) is known as the scenarios where a set $\mathcal{L}$ of transmit-receive parts are sharing the same resource blocks (e.g. time, frequency). Each of the receivers wishes to decode the signal originating from its own transmitter, subject to interference from the remaining transmitters. In cellular networks, this would correspond to a set of $\mathcal{L}$ BSs, each serving a single MS, over the same resource block. Then, the MIMO IFC is a special case of the MIMO IBC.

4.2.2 The MIMO Interfering Multiple-Access Channel

The MIMO Interfering Multiple-Access Channel (MIMO IMAC) represents an uplink communication scenario, in the context of cellular networks. In that sense, the MIMO IMAC is the "network dual" of the MIMO IBC. It comprises of a network of $\mathcal{L}$ cells, each cell containing one BS, where BS l is serving a set $\mathcal{K}_l$ of MSs, and each user wishes to send data to its serving BS (shown in Fig. 4.3).

We use the same notation as (4.2.1) to denote the total set of users, $\mathcal{I}$,

$$\mathcal{I} = \{l_j \mid (l,j) \in \mathcal{L} \times \mathcal{K}_l\} \tag{4.2.6}$$

where l_j denotes the index of user $j \in \mathcal{K}_l$, at BS $l \in \mathcal{L}$. $\boldsymbol{x}_{i_k} \in \mathbb{C}^d$ represents the d-dimensional vector of independently encoded symbols sent by MS $i_k \in \mathcal{I}$ (zero mean circularly symmetric), with covariance matrix $\mathbb{E}[\boldsymbol{x}_{i_k}\boldsymbol{x}_{i_k}^{\dagger}] = (\rho/d)\boldsymbol{I}_d$. $\boldsymbol{V}_{i_k} \in \mathbb{C}^{M \times d}$

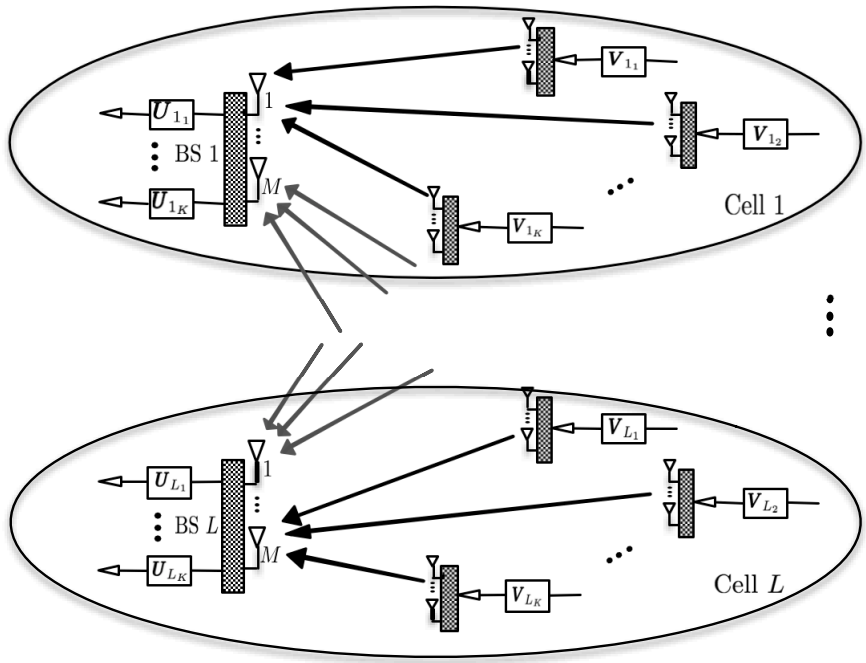

Figure 4.3: The MIMO Interfering Multiple-Access Channel (MIMO IMAC)

denotes the d-dimensional transmit filter employed by transmitter $i_k \in \mathcal{I}$, $\boldsymbol{H}_{l,i_k}$ the $N \times M$ MIMO channel from MS $i_k \in \mathcal{I}$, to BS l, assumed to have circularly symmetric i.i.d. complex random variables with zero mean and unit variance. With that in mind, the received signal at BS $l \in \mathcal{L}$ is given by,

$$\boldsymbol{y}_l = \sum_{i_k \in \mathcal{I}} \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{x}_{i_k} + \boldsymbol{n}_l, \tag{4.2.7}$$

We denote by $\boldsymbol{U}_{l_j} \in \mathbb{C}^{N \times d}$ the d-dimensional receive filter of user $l_j \in \mathcal{I}$. After linear processing with the receive filter, the recovered signal vector of user j in cell l, $\hat{\boldsymbol{x}}_{l_j}$, is given by,

$$\hat{\boldsymbol{x}}_{l_j} = \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{H}_{l,l_j} \boldsymbol{V}_{l_j} \boldsymbol{x}_{l_j} + \sum_{\substack{i_k \in \mathcal{I} \\ i_k \neq l_j}} \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{x}_{i_k} + \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{n}_l, \ \forall \ l_j \in \mathcal{I} \tag{4.2.8}$$

where the first term represents the desired signal, the second both intra and inter-cell interference, and $\boldsymbol{n}_l$ represents the N-dimensional AWGN noise, such that $\mathbb{E}[\boldsymbol{n}_l \boldsymbol{n}_l^{\dagger}] = \sigma^2 \boldsymbol{I}_N$. Moreover, $\boldsymbol{R}_{l_j}$ and $\boldsymbol{Q}_{l_j}$ are the desired signal and interference

covariance matrices of user j's signal, at BS l, respectively, and are given by,

$$\begin{aligned}\boldsymbol{R}_{l_j} &= (\rho/d)\, \boldsymbol{H}_{l,l_j} \boldsymbol{V}_{l_j} \boldsymbol{V}_{l_j}^{\dagger} \boldsymbol{H}_{l,l_j}^{\dagger} \ , \\ \boldsymbol{Q}_{l_j} &= (\rho/d) \sum_{i_k \in \mathcal{I}} \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{V}_{i_k}^{\dagger} \boldsymbol{H}_{l,i_k}^{\dagger} - \boldsymbol{R}_{l_j}\end{aligned} \tag{4.2.9}$$

The MIMO IMAC is the communication scenario that is investigated in Chap. 5.

4.2.3 General Remarks

Note that both the MIMO IBC and MIMO IMAC form the basis for deriving all downlink and uplink models, i.e., every model is a special case of either one. In that sense, the MIMO Interference Channel and the MIMO Broadcast Channel (MIMO BC) are special cases of the MIMO IBC, and the MIMO Multiple-Access Channel (MIMO MAC) is special case of the MIMO IMAC. Thus, we restrict our presentation in this chapter, to the MIMO IBC and MIMO IMAC.

Note that we use the term transmit filters / receiver filters in view of keeping nomenclature generic: this way, our framework is equally applicable to both MIMO IMAC and MIMO IBC. With that in mind, the transmit filters $\{\boldsymbol{V}_{l_j}\}$ represent the set of transmit precoders at the BSs (in the MIMO IBC context), and the transmit precoders used by the UE's (in the MIMO IMAC context). In addition, the receiver filters $\{\boldsymbol{U}_{l_j}\}$ represent the set of receive filters at the UEs (in the MIMO IBC context), and the receive filter applied at the BSs (in the MIMO IMAC context). This is shown in Fig. 4.2 and Fig. 4.3

Note that in the above models (and what follows thereafter) we assume that M, N and d are the same across users and BSs, for conciseness. However, this can easily extended by adding the relevant indexes.

4.2.4 Assumptions

We first make explicit the following definitions.

Definition 4.2.1 (Local and Global CSI)**.** Given a network of transmitters / receivers, *global CSI* refers to channel knowledge regarding all channels in the network. On the other hand, *local CSI* refers to the channels that are directly linked to a transmitter or receiver.

Definition 4.2.2 (Distributed Algorithm)**.** In the scope of this thesis, an algorithm is classified as distributed if it requires local CSI at each transmitter and receiver.

We outline the main assumptions of this part in the thesis.

Assumption 4.2.1 (Local CSI)**.** We assume that the users and BSs have local CSI, i.e., each user (resp. BS) knows the channels to its *desired and interfering* BSs (resp. users). We recall that investigating the CSI acquisition mechanism is not part of this work.

Assumption 4.2.2 (Perfect CSI)**.** CSI at each BS and user is assumed to be perfectly known, i.e., channel estimation errors are not accounted for.

Assumption 4.2.3 (Distributed Operation)**.** All schemes are required to use local CSI only, using the framework of Forward-Backward training.

Assumption 4.2.4 (Decoding)**.** Note that a common assumption in multi-user uplink communication scenarios, is that *multi-user decoding* is performed (e.g. successive interference cancellation). Due to the fact that such receivers are hard to realize in practice, we do not make such assumptions. Moreover, joint encoding and decoding of each user's desired streams in assumed, while interference is treated as noise.

Assumption 4.2.5 (Low-Overhead Regime)**.** Following the argument put forth in Sect. 4.3.3, we restrict our proposed schemes to operate in the *low-overhead regime*, where only a small number of F-B iterations is used.

Assumption 4.2.6 (Transmit and receive power constraints)**.** We underline the fact that most of the work thus far only enforces a power constraint on the precoder. The reason for that is the fact that communication in those setting is only one-directional, i.e. from transmitter to receiver, and thus no receive power constraint is needed. However, our argument is as follows: when using distributed optimization schemes employing F-B iterations, receivers are *active* in one of the phases (i.e., by sending pilots). Thus, generally, one *does* need a maximum transmit power constraint for the receiver filter, in addition to the maximum transmit power constraint. In contrast to what has been done so far, we take this fact into account in all our contributions, within this thesis: we impose a maximum power constraint for *both* the transmitter and receiver.

4.3 Distributed CSI Acquisition

Obviously, all the methods discussed in the last section require each transmitter and receiver to have local CSI. In this part, the CSI acquisition mechanism is distributed. Such a mechanism underlies all our proposed schemes in this part of the thesis, namely, the algorithms proposed in Chap. 5 and Chap. 6. And although the investigation of different CSI acquisition mechanisms are outside the scope of the thesis, we briefly summarize that process, for completeness.

4.3.1 Forward Backward Iterations

Also known as *ping-pong iterations*, *over-the-air iterations*, and *bi-directional training*. In the context of cellular networks, one of the fundamental building blocks for distributed optimization techniques are the so-called *Forward-Backward (F-B) iterations*. In brief, F-B iterations exploit the reciprocity of the network - which only holds in systems employing Time-Division Duplexing (TDD), and local CSI at each

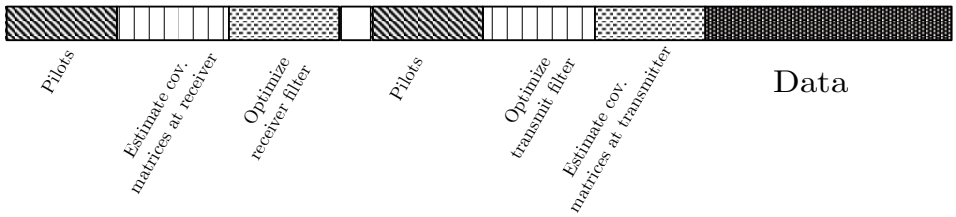

Figure 4.4: Basic structure of Forward-Backward (F-B) iteration

node, to gradually refine each of the transmit and receive filters, one at a time. In the forward phase, transmitters send precoded pilots that allow each receiver to acquire local CSI. Based on the latter, each receiver optimizes its receive filter (depending on cost function that is used). The same process is used for the backward phase, whereby receivers send precoded pilots, transmitters estimate their local CSI and update their respective filters. This constitutes one F-B iteration (represented in Fig. 4.4), the number of such iterations is a design choice. The main feature of this process is that it results in a distributed algorithm. Though not made explicit, this underlying F-B structure is ubiquitous in almost all distributed optimization techniques for cellular networks, e.g., [GCJ11, SSB+09, SRLH11, PH11, SGHP10a]. Those works are among the first to use this particular F-B structure within the context of MIMO IC and MIMO IBC, but its usage is attributed to many earlier works such as [CTRF02, Ben02].

4.3.2 Mechanism for CSI Acquisition

The operation of the aforementioned schemes is contingent upon each transmitter / receiver being able to estimate the *signal* and/or the *interference-plus-noise* covariance matrices, in a fully distributed manner. This is accomplished via the use of precoded pilots to estimate the *effective channels*. The methods developed in [BB15a] are fully applicable, and we thus summarize the basic underlying structure. In the first phase, the signal covariance matrix for receiver l_j, $\boldsymbol{R}_{l_j}$, can be computed after estimating the effective signal channel, and the interference-plus-noise covariance matrix is computed after estimating the effective interfering channels. The receive filters at the BSs are updated following any of the proposed algorithms. Then, in the second phase, the same procedure is used to estimate the signal and interference-plus-noise covariance matrices, and update the filters at the receivers. The process is summarized in Fig. 4.4. This aforementioned process constitutes one forward-backward (F-B) iteration. We let T denote the total number of such iterations that are carried out. We refer the interested reader to [BB15a] for a fully detailed description of this mechanism.

4.3.3 Communication Overhead

Thus, for such schemes to be fully distributed (i.e. requiring only local CSI), the required CSI quantities have to be obtained via uplink-downlink pilots and training. As we can see, each F-B iteration has an associated communication overheard, namely the cost of bi-directional transmission of pilots. Although many other works consider a more comprehensive definition of overhead (such as [EALH12] and [EALH11]), we adopt a more simplistic definition, keeping in mind that the actual overhead will be close to this quantity: it is the number of (minimal orthogonal) pilots symbols needed for estimating the required CSI quantities (assuming that minimal number of orthogonal pilots is used, i.e., d orthogonal pilot slots for each uplink/downlink effective channel).

The complexity, pilot requirement, and number of forward-backward iterations depend on the type of global cost function that is being optimized. Simple cost functions such as interference leakage [GCJ11] and MSE [SSB$^+$09], were initially considered, and later extended to directly optimize more complex cost functions such as the sum-rate [SGHP10b], weighted sum-rate [SRLH11]. It becomes clear at this stage that the overhead associated with such schemes is largely dominated by the number of such F-B iterations, before convergence is reached. That being said, almost all schemes falling under the category of F-B iterations, require a relatively large number of such iterations, in the order of *hundreds to thousands* [SSB$^+$13]. Moreover, this number seems to increase with the dimensions of the system. Consequently, this modus operandi is not feasible in a cellular network (since F-B iterations are carried out over-the-air, and the associated overhead would be higher than the potential gains). We thus focus on a regime where $T = 2 \sim 5$.

This major limitation became the object of recent investigations such as [KTJ13, BB15b, GKBS15, GKBS16b]. In our recent work, we also proposed algorithms with improved convergence properties, using the interference leakage as cost function [GKBS15], and (lower bounds on) the sum-rate [GKBS16b] function. The proposed schemes achieve a similar performance as their conventional counterparts, however, with a drastically lower number of forward-backward iterations. Such schemes result in *orders of magnitude reduction in overhead*, over the conventional counterparts. In fact, one of the main contributions of this part (Chap. 6 and Chap. 5) is to design algorithms that operate under a very low-overhead regime (where just a few F-B iterations are performed).

Although additional issues such as robustness and CSI error, have to considered as well, such matter are outside the scope of our work (we refer the interested reader to [BB15a]).

4.4 Problem Formulation

4.4.1 Distributed Utility Maximization

We refer to *network-utility functions* as a type of utility functions that are used in the context of communication networks, namely, cellular networks in our work. With that in mind, *distributed network-utility optimization* refers to the process of designing distributed optimization algorithms, that optimize some network-wide sum-utility function, over a set of optimization variables. In the context of cellular networks (more specifically, the MIMO IBC / MIMO IMAC), this can be written in generic from, as follows,

$$\begin{cases} \max \ u_{\Sigma} = \sum_{l_j \in \mathcal{I}} u_{l_j}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\}) \\ \text{s. t.} \ \ \boldsymbol{V}_{l_j} \in \mathcal{V}_{l_j}, \ \ \boldsymbol{U}_{l_j} \in \mathcal{U}_{l_j}, \ \forall \ l_j \in \mathcal{I} \end{cases} \tag{4.4.1}$$

where $\boldsymbol{V}_{l_j} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{U}_{l_j} \in \mathbb{C}^{N \times d}$ are transmit and receive filter for user $l_j \in \mathcal{I}$, respectively. u_{Σ} is the network-wide utility function (also called sum-utility function), and u_{l_j} the utility of user $l_j \in \mathcal{I}$, defined as

$$u_{l_j}(\{\boldsymbol{U}_{l_j}, \boldsymbol{V}_{l_j}\}) : \{\mathbb{C}^{N \times d} \times \mathbb{C}^{M \times d}\}^{KL} \rightarrow \mathbb{R}_{+} \tag{4.4.2}$$

and assumed to be smooth and twice differentiable. Moreover, u_{Σ} is assumed to be (additively) separable, i.e., $u_{\Sigma} = \sum_{l_j \in \mathcal{I}} u_{l_j}(\boldsymbol{U}_{l_j}, \{\boldsymbol{V}_{l_j}\})$

Moreover, the sets $\mathcal{V}_{l_j}$ and $\mathcal{U}_{l_j}$ representing individual constraints (possibly non-convex), are assumed to be closed. While constraints arising in the context of the wireless communication can be quite diverse, in the proposed framework, they are assumed to be per-user, e.g., $\|\boldsymbol{V}_{l_j}\|_F^2 \leq P_{l_j}, \ \forall l_j \in \mathcal{I}$ and $\|\boldsymbol{U}_{l_j}\|_F^2 \leq P_{l_j}, \ \forall l_j \in \mathcal{I}$. The argument in favor of having a maximum transmit power constraint, for the receiver as well, is discussed in Sect. 4.2.6 .

4.4.2 Block-Coordinate Descent

The framework under consideration entails tackling problems such as (4.4.1), using the well known Block-Coordinate Descent (BCD) method (described in Sect. 1.1.1): the block $\{\boldsymbol{U}_{l_j}\}$ is optimized, while the block $\{\boldsymbol{V}_{l_j}\}$ is assumed to be fixed (and vice-versa). Letting ^{n} denote the iteration index, the resulting method is formalized below,

$$\underbrace{\{\boldsymbol{V}_{l_j}^{n+1}\} \triangleq \underset{\{\boldsymbol{V}_{l_j}\}}{\operatorname{argmax}} \ u_{\Sigma} \left(\underbrace{\{\boldsymbol{U}_{l_j}^{n+1}\} \triangleq \underset{\{\boldsymbol{U}_{l_j}\}}{\operatorname{argmax}} \ u_{\Sigma}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}^{n}\})}_{J1} \ , \ \{\boldsymbol{V}_{l_j}\} \right)}_{J2}, \ n = 1, 2, ... \tag{4.4.3}$$

Moreover, the resulting subproblems are as follows,

$$(J1)\ \begin{cases} \max\limits_{\{\boldsymbol{U}_{l_j}\}} \sum_{l_j \in \mathcal{I}} u_{l_j}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}^n\}) \\ \text{s. t. } \boldsymbol{U}_{l_j} \in \mathcal{U}_{l_j}, \forall l_j \in \mathcal{I} \end{cases} \tag{4.4.4}$$

$$(J2)\ \begin{cases} \max\limits_{\{\boldsymbol{V}_{l_j}\}} \sum_{l_j \in \mathcal{I}} u_{l_j}(\{\boldsymbol{U}_{l_j}^{n+1}\}, \{\boldsymbol{V}_{l_j}\}) \\ \text{s. t. } \boldsymbol{V}_{l_j} \in \mathcal{V}_{l_j}, \forall l_j \in \mathcal{I} \end{cases} \tag{4.4.5}$$

Since u_Σ is separable, then the latter problems can be rewritten in equivalent form,

$$(J1):\ \boldsymbol{U}_{l_j}^{n+1}\ \begin{cases} \operatorname{argmax}\limits_{\boldsymbol{U}_{l_j}} u_{l_j}(\boldsymbol{U}_{l_j}, \{\boldsymbol{V}_{l_j}^n\}) \\ \text{s. t. } \boldsymbol{U}_{l_j} \in \mathcal{U}_{l_j} \end{cases}, \forall l_j \in \mathcal{I} \tag{4.4.6}$$

$$(J2):\ \boldsymbol{V}_{l_j}^{n+1}\ \begin{cases} \operatorname{argmax}\limits_{\boldsymbol{V}_{l_j}} u_{l_j}(\{\boldsymbol{U}_{l_j}^{n+1}\}, \boldsymbol{V}_{l_j}) \\ \text{s. t. } \boldsymbol{V}_{l_j} \in \mathcal{V}_{l_j}, \end{cases} \quad \forall l_j \in \mathcal{I} \tag{4.4.7}$$

where $\boldsymbol{U}_{l_j}^{n+1}$ and $\boldsymbol{V}_{l_j}^{n+1}$ are local optimizers for $(J1)$ and $(J2)$, respectively , i.e., $\boldsymbol{U}_{l_j}^{n+1}$ and $\boldsymbol{V}_{l_j}^{n+1}$ need not be globally optimal solutions to their respective problems, but only satisfy the KKT conditions.

Regarding convergence of the BCD method in(4.4.3) under the proposed framework, the latter can readily be established in a straightforward manner.

Proposition 4.4.1 (Monotonicity)**.** *Let $\psi^n \triangleq u_\Sigma(\{\boldsymbol{U}_{l_j}^n\}, \{\boldsymbol{V}_{l_j}^n\})$, $n = 1, 2, ...$ be the sequence of iterates for the objective value. Then, $\{\psi^n\}$ is non-decreasing in n, and converges to a limit point, ψ_0*

The proof is simple. The application of each of the updates, $\{\boldsymbol{U}_{l_j}^{n+1}\}$ and $\{\boldsymbol{V}_{l_j}^{n+1}\}$, cannot decrease the cost function,

$$u_\Sigma(\{\boldsymbol{U}_{l_j}^n\}, \{\boldsymbol{V}_{l_j}^n\}) \leq u_\Sigma(\{\boldsymbol{U}_{l_j}^{n+1}\}, \{\boldsymbol{V}_{l_j}^n\}) \leq u_\Sigma(\{\boldsymbol{U}_{l_j}^{n+1}\}, \{\boldsymbol{V}_{l_j}^{n+1}\})$$

where the inequalities follow from the fact that the application of the optimal updates for $(J1)$ and $(J2)$, $\boldsymbol{U}_{l_j}^{n+1}$ and $\boldsymbol{V}_{l_j}^{n+1}$, cannot increase ψ_n. It follows that ψ_n is non-decreasing, and there exists a limit point, ψ_0, such that $\psi_0 = \lim_{n\to\infty} \psi^n$ Thought monotonic convergence to a limit point is guaranteed, establishing the fact that the latter is a stationary point of u_Σ obviously requires more conditions. If strong convexity for each of the subproblems, $(J1)$ and $(J2)$, can be established, the corresponding minimizers, $\boldsymbol{U}_{l_j}^{n+1}$ and $\boldsymbol{V}_{l_j}^{n+1}$, are unique, and convergence to a stationary point immediately follows.

As it will become clear, the above formulation encompasses a wide array of network-utility optimization problems. Thus it can be verified that a significant fraction of previous works is a special case of the latter, namely, mean-squared error minimization [SSB+09, PH11], interference leakage minimization [GCJ11]. Moreover, the two major contributions of this thesis, in this part, are alternate embodiments of that framework: in Chap. 5, u_Σ represents the so-called DLT bound (a lower bound on the sum-rate), and in Chap. 6 u_Σ represents the total interference leakage.

4.4.3 Sum-rate maximization

The first utility in question is the sum-rate, defined as the sum of all individual achievable rates (across all users).

SRM for MIMO IBC

For the MIMO IBC setup in (4.2.1), the achievable rate of receiver l_j is given by,

$$r_{l_j} = \log_2 \left| \boldsymbol{I}_d + \left(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j} \right) \left(\boldsymbol{U}_{l_j}^\dagger (\boldsymbol{Q}_{l_j} + \sigma^2 \boldsymbol{I}_N) \boldsymbol{U}_{l_j} \right)^{-1} \right|, \; l_j \in \mathcal{I} \tag{4.4.8}$$

Then, the corresponding sum-rate maximization problem (for the MIMO IBC) is formulated as follows,

$$(SRM - IBC) \begin{cases} \max \; R_\Sigma(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\}) = \sum_{l_j \in \mathcal{I}} r_{l_j} \\ \text{s. t.} \; \|\boldsymbol{V}_{l_j}\|_F^2 \leq P_t, \; \|\boldsymbol{U}_{l_j}\|_F^2 \leq P_r \; \forall l_j \in \mathcal{I} \end{cases} \tag{4.4.9}$$

where P_t is the total power budget of transmitter l_j, and P_r that power constraint of receiver l_j (the argument for including a receive power constraint was discussed in Assumption 4.2.6). In contrast to previous formulations of $(SRM - IBC)$ where a sum-power constraint is used [SRLH11], we follow the main assumptions of the proposed framework (Sect. 4.4), and assume a per-user power constraint (as shown in the above problem) - keeping in mind that a sum-power constraint could potentially be handled. Note that the above problem degenerates into a the sum-rate maximization problem for MIMO IC, $(SRM - IC)$, when one user is served by each BS. And since the latter is NP-hard [RLL11], it follows that $(SRM - IBC)$ is NP-hard as well.

Special Case: SRM for MIMO IC Following the signal model for the MIMO IC in (4.2.1), the achievable sum-rate of receiver $l \in \mathcal{I}$ is given by,

$$r_l = \log_2 \left| \boldsymbol{I}_d + \left(\boldsymbol{U}_l^\dagger \boldsymbol{R}_l \boldsymbol{U}_l \right) \left(\boldsymbol{U}_l^\dagger (\boldsymbol{Q}_l + \sigma^2 \boldsymbol{I}_N) \boldsymbol{U}_l \right)^{-1} \right|, \tag{4.4.10}$$

We can then formulate the sum-rate maximization problem, $(SRM-IC)$, as follows,

$$(SRM-IC)\begin{cases}\max\ R_{\Sigma}(\{\boldsymbol{U}_l\},\{\boldsymbol{V}_l\})=\sum_{l\in\mathcal{I}}^{L} r_l\\ \text{s. t. }\ \|\boldsymbol{V}_l\|_F^2\leq P_t,\ \|\boldsymbol{U}_l\|_F^2\leq P_r,\forall\ l\in\mathcal{I}\end{cases} \tag{4.4.11}$$

SRM for MIMO IMAC

Following the signal model for the MIMO IMAC setup in (4.2.2), the achievable rate of user $l_j\in\mathcal{I}$ is given by,

$$r_{l_j}=\log_2|\boldsymbol{I}_d+(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}(\boldsymbol{Q}_{l_j}+\sigma^2\boldsymbol{I}_N)\boldsymbol{U}_{l_j})^{-1}|, \tag{4.4.12}$$

Then, the corresponding sum-rate maximization problem (for the MIMO IMAC is formulated as follows,

$$(SRM-IMAC)\begin{cases}\max\ R_{\Sigma}(\{\boldsymbol{U}_{l_j}\},\{\boldsymbol{V}_{l_j}\})=\sum_{l_j\in\mathcal{I}} r_{l_j}\\ \text{s. t. }\ \|\boldsymbol{V}_{l_j}\|_F^2\leq P_t,\ \|\boldsymbol{U}_{l_j}\|_F^2\leq P_r,\ \forall l_j\in\mathcal{I}\end{cases} \tag{4.4.13}$$

The discussion motivating the use of a receive power constraint was discussed in Assumption 4.2.6. In contract the the MIMO IBC case, trying to impose the same sum-power constraint in the MIMO IMAC leads to a sum-power constraint across all UEs, in one cell: this is clearly not applicable in practice since it would hinder the distributed nature of the algorithm. For the reasons above, a per-user constraint is the natural choice, in the sum-rate maximization problem for the MIMO IMAC, a per-user power allocation is assumed. Similarly to the MIMO IBC case, the NP-hardness of $(SRM-IMAC)$ can be easily establish.

Relevant Work Despite the fact that most SRM problems are NP-hard, several approaches still tacked the latter problem. The authors in [SGHP10a] use Block Coordinate Descent (BCD) to alternately optimize the transmit and receive filter for a MIMO IC, while moving in the direction of sum-rate gradient at each step (convergence to a local minimum could not be shown due to the projection step). The weighted SRM problem for the MIMO IC was addressed in [NSGS10] where the authors exploit the equivalence between the weighted SRM problem and the Weighted MMSE problem to design an BCD-based algorithm to alternately optimize the transmit and receive filters (convergence to a local optimum was shown). The same problem and algorithm was generalized in [SRLH11] for the MIMO IBC setting, as the well known Weighted MMSE algorithm. Note that the above SRM problems (IBC, IC and IMAC) do not fall under the framework presented in Sect. 4.4. However, the main contribution of Chap. 5 is to propose a mechanism for circumventing the latter problem, by introducing a utility that we dub Difference of log and trace (DLT).

4.4.4 Interference Leakage Minimization

The interference leakage is yet another metric that is used for optimizing the performance of multiuser cellular networks. In contrast to (SRM) problems that are quite complicated, the leakage metric yields simple tractable expressions. Within the context of Interference Alignment for MIMO IC, the interference leakage minimization (LM) problem was first formulated in [GCJ11]. We present it for the generic MIMO IBC case.

4.4.5 LM for MIMO IBC

Following the signal model introduced in Sect. 4.2.1, the Leakage Minimization problem (LM) is formulated as,

$$(LM-IBC)\begin{cases}\min\ \phi(\{\boldsymbol{U}_{l_j}\},\{\boldsymbol{V}_{l_j}\})=\sum_{l_j\in\mathcal{I}}\mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})\\ \text{s. t. } \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{U}_{l_j}=\boldsymbol{I}_d,\ \boldsymbol{V}_{l_j}^{\dagger}\boldsymbol{V}_{l_j}=\boldsymbol{I}_d,\ \forall l_j\in\mathcal{I}\end{cases} \tag{4.4.14}$$

It can be easily checked that $(LM-IBC)$ belongs to the framework presented in Sect. 4.4: ϕ is separable, and convex in each block of variables. Moreover, the resulting subproblem, has a unique solution (given by the eigenvectors of the covariance matrix in question). The framework presented in Chap. 4.4 is fully applicable. This is the basis of the well-known Distributed Interference Alignment algorithm [GCJ11].

Motivation It can be seen that $(LM-IBC)$ is a "surrogate problem" for interference alignment, in the sense that if $\phi=0$, then the IA conditions in Sect. 4.1.1 are satisfied. Moreover, another motivation for leakage minimization problem is its connection to sum-rate maximization problem [GKBS15]. Referring to (4.4.9), as $\sigma^2\to 0$ (high-SNR regime), the achievable rate r_{l_j} can be approximated by,

$$\tilde{r}_{l_j}=\log_2\left|\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}\right|-\log_2\left|\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}\right|$$

Then, one can approximate $(SRM-IBC)$ as follows,

$$(SRM-IBC)\quad\max_{\{\boldsymbol{U}_{l_j}\},\{\boldsymbol{V}_{l_j}\}}\quad\tilde{R}_{\Sigma}=\sum_{l=1}^{K}\sum_{j=1}^{L}\tilde{r}_{l_j}. \tag{4.4.15}$$

By construction, algorithms based on interference leakage (referred to as subspace methods) only optimize the interference subspace (as previously proposed algorithms in [GCJ11], [PH11]). Thus, by dropping the signal term in $\tilde{r}_{l_j}$, we can bound

it as follows,

$$\tilde{r}_{l_j} \geq -\log_2 \left| \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j} \right| \overset{(a)}{\geq} \sum_{i=1}^{d} -\log_2 \left([\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}]_{ii} \right)$$
$$\overset{(b)}{>} -\sum_{i=1}^{d} [\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}]_{ii} = -\mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})$$

where (a) follows directly from applying Hadamard's inequality, i.e. $|\boldsymbol{A}| \leq \prod [\boldsymbol{A}]_{ii}$ for $\boldsymbol{A} \succeq \boldsymbol{0}$, and (b) from the fact that $x > \log_2(x),\ \forall\ x > 0$. Although this result is expected, it proves that *minimizing the interference leakage at each user, results in optimizing a lower bound on the user's high-SNR rate.*

Related Work Several later works opted to use the interference leakage as metric for optimization due to its inherent simplicity (e.g., [PH11] and [GCJ11]). The authors in [MNM11] used the interference leakage as a metric, however, their formulation entails a constraint on the desired signal space. The main contribution of the thesis for the LM problem, is that of Chap. 6, where we investigate relaxing the $(LM - IBC)$ problem (replacing the unitary constraints with maximum power constraint), thereby speeding up the convergence of their proposed algorithm.
Problem Statement: The goal for this part of the thesis is to develop algorithms under the umbrella of the above framework. Moreover, such algorithms should be fast-converging (i.e., $T = 2 \sim 5$ F-B iterations), with convergence that is shown analytically.

Chapter 5

Sum-Rate Maximization Algorithms

We address the problem of sum-rate maximization in MIMO Interfering Multiple-Access Channels (MIMO IMAC) in this chapter. Due to the NP-hard nature of the problem (discussed in Chap. 4.2.2), we propose to lower bound the problem using a so-called *DLT bound* (i.e., a difference of log and trace). We show that it is a lower bound on the sum-rate, shed light on its tightness, and underline a major advantage of using such a bound: The resulting problem is an instance of the distributed network-utility optimization, and it thus leads to separable subproblems that decouple at both the transmitters and receivers. Moreover, we derive the solution to the latter subproblem, that we dub *non-homogeneous waterfilling* (a variation on the MIMO waterfilling solution), and underline an inherent desirable feature: its ability to turn-off streams exhibiting low-SINR, thereby greatly speeding up the convergence of the proposed algorithm. We then show the convergence of the resulting algorithm, max-DLT, to a stationary point of the DLT bound (a lower bound on the sum-rate).

We also propose a distributed algorithm dubbed Alternating Iterative Maximal Separation (AIMS), that is a generalization of max-SINR [GCJ11]. Furthermore, we argue (and later verify via simulations) that this generalization offers superior performance over max-SINR. Finally, we rely on extensive simulation of various network configurations, to establish the superior performance of our proposed schemes, with respect to other state-of-the-art methods.

Remark 5.1. Though the paper addresses the problem at hand for a MIMO IMAC, it can be verified that the latter framework and methods are applicable to the network-dual problem, the MIMO IBC, without modifications. Needless to say, it also applies to all ensuing special cases, such as the MIMO IFC, and the MIMO Multiple-Access Channel. This will be done in the numerical results section of this chapter.

Notation: In addition to the notation defined in Chap. 1, we define the following: for a given matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-\dagger}$ denotes $(\boldsymbol{A}^{\dagger})^{-1}$.

5.1 Maximizing DLT bounds

In this section we propose another approach to tackle the sum-rate optimization problem. The central idea behind this approach is to use a lower bound on the sum-rate, that results in separable sub-problems. Following the MIMO IMAC model presented in Chap. 4.2.2, the rate of user l_j is given by,

$$r_{l_j} = \log_2 |\boldsymbol{I}_d + (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}|, \tag{5.1.1}$$

where $\boldsymbol{R}_{l_j}$ and $\boldsymbol{Q}_{l_j}$ are the desired signal and interference-plus-noise (IPN) covariance matrices for user j, at BS l, respectively, and are given by,

$$\begin{aligned}
\boldsymbol{R}_{l_j} &= \boldsymbol{H}_{l,l_j} \boldsymbol{V}_{l_j} \boldsymbol{V}_{l_j}^\dagger \boldsymbol{H}_{l,l_j}^\dagger, l_j \in \mathcal{I} \\
\boldsymbol{Q}_{l_j} &= \sum_{i=1}^{L} \sum_{k=1}^{K} \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{V}_{i_k}^\dagger \boldsymbol{H}_{l,i_k}^\dagger + \sigma_l^2 \boldsymbol{I}_N - \boldsymbol{R}_{l_j}, l_j \in \mathcal{I}.
\end{aligned}$$

and σ_l^2 is the noise variance as BS $l \in \mathcal{L}$. Moreover, we define,

$$\begin{aligned}
\bar{\boldsymbol{R}}_{i_k} &= \boldsymbol{H}_{i,i_k}^\dagger \boldsymbol{U}_{i_k} \boldsymbol{U}_{i_k}^\dagger \boldsymbol{H}_{i,i_k}, \ i_k \in \mathcal{I} \\
\bar{\boldsymbol{Q}}_{i_k} &= \sum_{l=1}^{L} \sum_{j=1}^{K} \boldsymbol{H}_{l,i_k}^\dagger \boldsymbol{U}_{l_j} \boldsymbol{U}_{l_j}^\dagger \boldsymbol{H}_{l,i_k} + \bar{\sigma}_{i_k}^2 \boldsymbol{I}_M - \bar{\boldsymbol{R}}_{i_k}, i_k \in \mathcal{I}
\end{aligned}$$

as the signal and IPN covariance matrices of user i_k, in the reverse network (where $\bar{\sigma}_{i_k}^2$ is the noise variance at user i_k). Finally, we henceforth denote $\boldsymbol{L}_{l_j} \boldsymbol{L}_{l_j}^\dagger$ as the Cholesky Decomposition of $\boldsymbol{Q}_{l_j}$, and $\boldsymbol{K}_{i_k} \boldsymbol{K}_{i_k}^\dagger$ as that of $\bar{\boldsymbol{Q}}_{i_k}$

We restate the resulting sum-rate maximization problem for the MIMO IMAC (4.4.13), for convenience.

$$(SRM - IMAC) \begin{cases} \max \ R_\Sigma(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\}) = \sum_{l_j \in \mathcal{I}} r_{l_j} \\ \text{s. t. } \|\boldsymbol{V}_{l_j}\|_F^2 = P_t, \ \|\boldsymbol{U}_{l_j}\|_F^2 = P_r, \ \forall l_j \in \mathcal{I} \end{cases} \tag{5.1.2}$$

Note that while the original formulation entails inequality constraints on the transmit/receive filters, we will adopt in this chapter, and equality (as seen above). The reason for this choice will be discussed in details, in Chap 5.1.4. Our proposal is to lower bound r_{l_j}, using the so-called DLT bound, as follows,

$$r_{l_j}^{(LB)} \triangleq \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}). \tag{5.1.3}$$

5.1.1 Problem Formulation

The derivations leading up to $r_{l_j}^{(LB)}$ are detailed in this section, focusing on interference-limited case where the following holds,

$$\begin{aligned}&\lambda_i[\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}] \to \infty,\ \forall i \in \{d\},\\ &\Leftrightarrow \begin{cases} A1)\ \lambda_i[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}] \to 0,\ \forall i \in \{d\} \\ A2)\ \boldsymbol{I}_d \succeq (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1} \end{cases}\end{aligned} \tag{5.1.4}$$

Proposition 5.1.1. *When* $(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) \succeq \boldsymbol{I}_d$*, the user-rate* r_{l_j} *in* (4.4.12) *is lower bounded by,*

$$\begin{aligned} r_{l_j} &\geq \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|, \quad (b.1) \\ &\geq \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) \triangleq r_{l_j}^{(LB)}, \end{aligned} \tag{5.1.5}$$

where $r_{l_j}^{(LB)}$ *is such that,*

$$\begin{aligned} \Delta_{l_j} &\triangleq r_{l_j} - r_{l_j}^{(LB)} \\ &= \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}| \\ &\quad + \mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}]),\ \forall l_j \in \mathcal{I}\ . \end{aligned} \tag{5.1.6}$$

Proof. Refer to Appendix 5.6.2. □

The DLT bound, $r_{l_j}^{(LB)}$, shall be used as basis for the optimization algorithm. With that in mind, the sum-rate R_Σ, can be lower bounded by $R_\Sigma^{(LB)}$,

$$\begin{aligned} &R_\Sigma^{(LB)} \\ &= \sum_{l_j \in \mathcal{I}} \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) \end{aligned} \tag{5.1.7}$$

$$= \sum_{i_k \in \mathcal{I}} \log_2 |\boldsymbol{I}_d + \boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{R}}_{i_k} \boldsymbol{V}_{i_k}| - \mathrm{tr}(\boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{Q}}_{i_k} \boldsymbol{V}_{i_k}), \tag{5.1.8}$$

where the last equality is due to $\log|\boldsymbol{I} + \boldsymbol{AB}| = \log|\boldsymbol{I} + \boldsymbol{BA}|$, and the linearity of tr(.). Then, the MIMO IMAC sum-rate optimization problem in (5.1.2), can be bounded below by solving the following,

$$\begin{cases} \max\limits_{\{\boldsymbol{V}_{l_j}, \boldsymbol{U}_{l_j}\}} R_\Sigma^{(LB)} \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r,\ \|\boldsymbol{V}_{l_j}\|_F^2 = P_t,\ \forall l_j \in \mathcal{I} \end{cases} \tag{5.1.9}$$

Note that the above problem is not jointly convex in all the optimization variables, due to the coupling between the transmit and receive filters.

5.1.2 Proposed Algorithm

The formulation in (5.1.9) is an alternate embodiment of the distributed sum-utility optimization framework (Sect. 4.4): we employ it to tackle 5.1.9. We use the superscript $^{(n)}$ to denote the iteration number: at the nth iteration, the transmit filters, $\{\boldsymbol{V}_{l_j}^{(n)}\}$, are fixed, and the update for the receive filters, $\{\boldsymbol{U}_{l_j}^{(n+1)}\}$, is the one that maximizes the objective (and vice versa). This is formalized in (5.1.10), and each of the resulting subproblems are detailed below.

$$\underbrace{\{\boldsymbol{V}_{l_j}^{(n+1)}\} \triangleq \underset{\{\boldsymbol{V}_{l_j}\}}{\operatorname{argmax}}\ R_{\Sigma}^{(LB)} \left(\underbrace{\{\boldsymbol{U}_{l_j}^{(n+1)}\} \triangleq \underset{\{\boldsymbol{U}_{l_j}\}}{\operatorname{argmax}}\ R_{\Sigma}^{(LB)}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}^{(n)}\})}_{J1}\ ,\ \{\boldsymbol{V}_{l_j}\} \right)}_{J2},\ n = 1, 2, ... \tag{5.1.10}$$

Essentially, in each of the two stages, BCD decomposes the original coupled problem (5.1.9), into a set of parallel subproblems, that can solved in distributed fashion. When the transmit filters are fixed, the problem decouples in the receive filters $\{\boldsymbol{U}_{l_j}\}$(as seen from (5.1.9)), and the resulting subproblems are given by,

$$(J1)\ \begin{cases} \min\limits_{\boldsymbol{U}_{l_j}}\ \sum_{l_j \in \mathcal{I}} \operatorname{tr}(\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r,\ \forall l_j \in \mathcal{I} \end{cases} \tag{5.1.11}$$

By recalling that $R_{\Sigma}^{(LB)}$ can be written in both (5.1.9) and (5.1.8), $(J2)$ can be written as,

$$(J2)\ \begin{cases} \min\limits_{\boldsymbol{V}_{i_k}}\ \sum_{i_k \in \mathcal{I}} \operatorname{tr}(\boldsymbol{V}_{i_k}^{\dagger} \bar{\boldsymbol{Q}}_{i_k} \boldsymbol{V}_{i_k}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{V}_{i_k}^{\dagger} \bar{\boldsymbol{R}}_{i_k} \boldsymbol{V}_{i_k}| \\ \text{s. t. } \|\boldsymbol{V}_{i_k}\|_F^2 = P_t,\ \forall i_k \in \mathcal{I} \end{cases} \tag{5.1.12}$$

Using the fact that $R_{\Sigma}^{(LB)}$ is separable, the resulting sub-problem at each receiver, and transmitter are given as,

$$(J1)\ \begin{cases} \min\limits_{\boldsymbol{U}_{l_j}} \operatorname{tr}(\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r, \end{cases} ,\forall l_j \in \mathcal{I}, \tag{5.1.13}$$

$$(J2)\ \begin{cases} \min\limits_{\boldsymbol{V}_{i_k}} \operatorname{tr}(\boldsymbol{V}_{i_k}^{\dagger} \bar{\boldsymbol{Q}}_{i_k} \boldsymbol{V}_{i_k}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{V}_{i_k}^{\dagger} \bar{\boldsymbol{R}}_{i_k} \boldsymbol{V}_{i_k}| \\ \text{s. t. } \|\boldsymbol{V}_{i_k}\|_F^2 = P_t, \end{cases} ,\forall i_k \in \mathcal{I}, \tag{5.1.14}$$

respectively. Thus, choosing DLT expressions is rather advantageous, since they lead to subproblems that decouple in both $\{\boldsymbol{U}_{l_j}\}$ and $\{\boldsymbol{V}_{l_j}\}$

Thus, choosing DLT expressions is rather advantageous, since they lead to subproblems that decouple in both $\{\boldsymbol{U}_{l_j}\}$ and $\{\boldsymbol{V}_{l_j}\}$. Note that the equality constraints in $(J1)$ and $(J2)$, do not affect the convexity of the problems, as they are already non-convex. Indeed, expressions such as $-\log_2|\boldsymbol{I}_d+\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}|$ are not convex in $\boldsymbol{U}_{l_j}$. [1] However, this does not make BCD less applicable, as long as $(J1)$ and $(J2)$ are solved *globally*. The solution to each of the subproblems is given by the following result.

Lemma 5.1.1. *Non-homogeneous Waterfilling.*
Consider the following problem,

$$(P)\ \begin{cases}\min\limits_{\boldsymbol{X}\in\mathbb{C}^{n\times r}} f(\boldsymbol{X})\triangleq \mathrm{tr}(\boldsymbol{X}^{\dagger}\boldsymbol{Q}\boldsymbol{X})-\log_2|\boldsymbol{I}_d+\boldsymbol{X}^{\dagger}\boldsymbol{R}\boldsymbol{X}|\\ \text{s. t. } \|\boldsymbol{X}\|_F^2=\zeta.\end{cases} \tag{5.1.15}$$

where $\boldsymbol{Q}\succ\boldsymbol{0}$ and $\boldsymbol{R}\succeq\boldsymbol{0}$, $r<n$. Let $\boldsymbol{Q}\triangleq\boldsymbol{L}\boldsymbol{L}^{\dagger}$ be the Cholesky factorization of $\boldsymbol{Q}$, and $\boldsymbol{M}\triangleq\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$, $\boldsymbol{M}\succeq\boldsymbol{0}$, and define the following, $\{\alpha_i\triangleq\lambda_i[\boldsymbol{M}]\}_{i=1}^{r}$, $\boldsymbol{\Psi}\triangleq v_{1:r}[\boldsymbol{M}]$, $\{\beta_i\triangleq\boldsymbol{\Psi}_{(i)}^{\dagger}(\boldsymbol{L}^{\dagger}\boldsymbol{L})^{-1}\boldsymbol{\Psi}_{(i)}\}_{i=1}^{r}$.
Then the globally optimal solution for the above problem is given by,

$$\boldsymbol{X}^{\star}=\boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{\star}, \tag{5.1.16}$$

where $\boldsymbol{\Sigma}^{\star}$ (diagonal) is the optimal power allocation,

$$(P4)\ \begin{cases}\min\limits_{\{x_i\}}\sum_{i=1}^{r}\left(x_i-\log_2(1+\alpha_i x_i)\right)\\ \text{s. t. }\ \sum_{i=1}^{r}\beta_i x_i=\zeta,\ x_i\geq 0,\forall i\end{cases} \tag{5.1.17}$$

Proof. Refer to Appendix 5.6.3 □

We underline that a similar problem was solved in [KG11]. However, a closer look reveals that their problem formulation concerns covariance optimization - a convex problem, as opposed to the non-convex precoder optimization in (P), and does not entail a power constraint. Hence, their results are not applicable to (P). With that in mind, $(P4)$ has a closed-form solution can be obtained using standard Lagrangian techniques.

Lemma 5.1.2. *The solution to the optimal power allocation in $(P4)$ is given by,*

$$\boldsymbol{\Sigma}_{(i,i)}^{\star}=\sqrt{\left(1/(1+\mu^{\star}\beta_i)-1/\alpha_i\right)^{+}},\forall i, \tag{5.1.18}$$

$\mu^{\star}$ is the unique root to,

$$g(\mu)\triangleq\sum_{i=1}^{r}\beta_i\left(1/(1+\mu\beta_i)-1/\alpha_i\right)^{+}-\zeta,$$

[1] To see this, consider the (degenerate) scalar case. It can be verified that $-\log_2(1+ru^2)$, $r>0$ is concave for $u\ll 1$, and convex for $u\gg 1$.

on the interval $]-1/(\max_i \beta_i),\ \infty[$*, and* $g(\mu)$ *is monotonically decreasing on that interval.*

Proof. Refer to Appendix 5.6.3 □

With that result, the optimal receive and transmit filter updates - the solution to $(J1)$ and $(J2)$ respectively, can be written as,

$$\begin{aligned}
\boldsymbol{U}_{l_j}^{\star} &= \boldsymbol{L}_{l_j}^{-\dagger}\boldsymbol{\Psi}_{l_j}\ \boldsymbol{\Sigma}_{l_j}^{\star},\quad \boldsymbol{\Psi}_{l_j} \triangleq v_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}],\quad \boldsymbol{\Psi}_{l_j} \in \mathbb{C}^{N\times d},\quad \forall\ l_j\ , \\
\boldsymbol{V}_{i_k}^{\star} &= \boldsymbol{K}_{i_k}^{-\dagger}\boldsymbol{\Theta}_{i_k}\ \boldsymbol{\Lambda}_{i_k}^{\star},\quad \boldsymbol{\Theta}_{i_k} \triangleq v_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}],\quad \boldsymbol{\Theta}_{i_k} \in \mathbb{C}^{M\times d},\quad \forall\ i_k\ ,
\end{aligned} \tag{5.1.19}$$

where $\boldsymbol{\Sigma}_{l_j}^{\star} \in \mathbb{R}_{+}^{d\times d}$ and $\boldsymbol{\Lambda}_{i_k}^{\star} \in \mathbb{R}_{+}^{d\times d}$ are the optimal diagonal power allocation, for the receive and transmit filter updates, respectively (given in Lemma 5.1.1). Moreover, $\boldsymbol{L}_{l_j}\boldsymbol{L}_{l_j}^{\dagger} \triangleq \boldsymbol{Q}_{l_j}, \boldsymbol{L}_{l_j} \in \mathbb{C}^{N\times N}$ is the Cholesky factorization of $\boldsymbol{Q}_{l_j}$, and $\boldsymbol{K}_{i_k}\boldsymbol{K}_{i_k}^{\dagger} \triangleq \bar{\boldsymbol{Q}}_{i_k}, \boldsymbol{K}_{i_k} \in \mathbb{C}^{M\times M}$ is the Cholesky factorization of $\bar{\boldsymbol{Q}}_{i_k}$, respectively.

The resulting algorithm, max-DLT, is detailed in Algorithm 3 (where T is the number of F-B iterations). Moreover, due to the monotone nature of $g(\mu)$, $\mu^{\star}$ can be found using simple 1D search methods, such as bisection.

Remark 5.2. We note that $\boldsymbol{\Sigma}_{l_j}^{\star}$ and $\boldsymbol{\Lambda}_{i_k}^{\star}$ are both required to ensure the monotonically increasing nature of the updates. And despite the fact that the actual user rate in (4.4.12) is invariant to $\boldsymbol{\Sigma}_{l_j}$, the latter is indeed heavily dependent on the choice of $\boldsymbol{\Lambda}_{i_k}$.

Algorithm 3 Maximal DLT (max-DLT)

for $t = 1, 2, ..., T$ **do**
  // forward network optimization: receive filter update
    Estimate $\boldsymbol{R}_{l_j}, \boldsymbol{Q}_{l_j}$, and compute $\boldsymbol{L}_{l_j}$, $\forall l_j$
    $\boldsymbol{U}_{l_j} \leftarrow \boldsymbol{L}_{l_j}^{-\dagger} v_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}]\boldsymbol{\Sigma}_{l_j},\quad \forall l_j$
  // reverse network optimization: transmit filter update
    Estimate $\bar{\boldsymbol{R}}_{i_k}, \bar{\boldsymbol{Q}}_{i_k}$, and compute $\boldsymbol{K}_{i_k}$, $\forall i_k$
    $\boldsymbol{V}_{i_k} \leftarrow \boldsymbol{K}_{i_k}^{-\dagger} v_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}]\boldsymbol{\Lambda}_{i_k},\quad \forall i_k$
end for

5.1.3 Relation to Other Methods

The fact that the proposed approach seems close to other heuristics such as successive convex programming (SCP) and the convex-concave procedure (CCP), is misleading. Those methods start with expressions such as $(b.1)$ (Proposition 5.1.1), and approximate $\log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|$ with a linear function (in the case of CCP [LB15]),

or lower bound it with a quadratic one (in the case of SCP [SZ95]). The approximation is iteratively updated until convergence. We will derive updates for the case of CCP, to illustrate our argument.

Starting with expressions such as (b.1) (Proposition 5.1.1),

$$\min_{\boldsymbol{U}_{l_j}} \underbrace{\log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|}_{h(\boldsymbol{U}_{l_j})} - \underbrace{\log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|}_{g(\boldsymbol{U}_{l_j})}, \tag{5.1.20}$$

CCP [LB15] generates a sequence of iterates $\{\boldsymbol{U}_{l_j}^{(n)}\}_n$, where at iteration n, h is approximated using its Taylor expansion at $\boldsymbol{U}_{l_j}^{(n)}$, as follows.

$$\begin{aligned}
\boldsymbol{U}_{l_j}^{(n+1)} &= \operatorname*{argmin}_{\boldsymbol{U}_{l_j}} \left(h(\boldsymbol{U}_{l_j}^{(n)}) + \operatorname{tr}\left\{ \nabla h(\boldsymbol{U}_{l_j}^{(n)})^\dagger (\boldsymbol{U}_{l_j} - \boldsymbol{U}_{l_j}^{(n)}) \right\} \right) - g(\boldsymbol{U}_{l_j}) \\
&= \operatorname*{argmin}_{\boldsymbol{U}_{l_j}} \operatorname{tr}(\nabla h(\boldsymbol{U}_{l_j}^{(n)})^\dagger \boldsymbol{U}_{l_j}) - g(\boldsymbol{U}_{l_j}) \\
&= \operatorname*{argmin}_{\boldsymbol{U}_{l_j}} \operatorname{tr}((\boldsymbol{A}_{l_j}^{(n)})^\dagger \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|
\end{aligned} \tag{5.1.21}$$

As the resulting problem in (5.1.21) is not convex, the expressions do yield closed form solutions for the updates $\boldsymbol{U}_{l_j}^{(n)} \to \boldsymbol{U}_{l_j}^{(n+1)}$. For those reasons, classical approaches such as the CCP are not of great use for the problem at hand. A similar argument can be made about the unsuitability of SCP as well. While the derivations in (5.1.21) apply CCP to transmit/receive filter optimization, they are better suited for covariance optimization problems. In such instances, their application yields convex problems (in contrast to (5.1.21)). Despite the fact that some earlier works successfully applied such methods to weighted sum-rate maximization problems ([NLN14, YCL14]), this was in done in the context of transmit covariance optimization: We will benchmark against such a CCP scheme, where transmit covariance optimization was considered in the MIMO IMAC setting [NLN14].

Note that a comparison between the CCP updates in (5.1.21) and those resulting from our proposed approach, e.g., (J1), reveals that indeed our approach is different from CCP. With that in mind, iteratively updating the DLT bound around the operating point (in a similar fashion to CCP or SCP), is not applicable: this is not of interest in this work, as the resulting bound would not be *separable* and decouple at transmit/receive filter. We also note that such approaches will inevitably lead to additional communication overhead and complexity; this goes against the main motivation of the work (communication overhead is detailed in Sec. 5.3.4). While this design choice might not lead to the best bound, the tightness of the DLT bound is shown in Proposition 5.1.1. In addition, the choice of our particular DLT expression, follows from the fact that few choices of bounds result in separable subproblems at *both* the transmitters and receivers.

5.1.4 Analysis and Discussions

Interpretation We provide an intuitive interpretation of the problem in Lemma 5.1.1 and its solution. It can be easily verified that $\{\alpha_i \triangleq \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}]\}_{i=1}^{r}$ are also the eigenvalues of $\boldsymbol{Q}^{-1}\boldsymbol{R}$ (where $\boldsymbol{R}$ and $\boldsymbol{Q}$ represent the signal and IPN covariance matrix, respectively). Thus, $\{\alpha_i\}$ acts as a (quasi)-SINR measure, for each of the data streams. Moreover, it can be seen that the optimal power allocated to stream i, $\boldsymbol{\Sigma}^{\star}_{(i,i)}$ in (5.1.18), tends to zero as $\alpha_i \to 0$, i.e., no power is allocated to streams that have low-SINR.[2] Moreover, note that $\{\beta_i\}$ represents the cost of allocating power to each of the streams (this can be seen in $(P4)$). Thus, the non-homogeneous waterfilling solution in (5.1.16) simply allocates power to each of the streams, based on the SINR and cost of each (possibly not allocating power to some streams).

Discussion We now discuss the reason for adopting the equality power constraints for the problem at hand (i.e., $(J1)$ and $(J2)$), by showing the limitation of using an inequality constraint. Note that in the noise-limited regime, $\sigma_l \gg 1,\ \forall l \in \mathcal{L}$, and consequently $\alpha_i \triangleq \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}] \to 0,\ \forall i \in \{r\}$. Using the fact that $\log(1+y) \approx y,\ y << 1$, the optimal power allocation in $(P4)$ is approximated as,

$$\sum_{i=1}^{r} x_i - \log_2(1+\alpha_i x_i) \approx \sum_{i=1}^{r} x_i - \alpha_i x_i = \sum_{i=1}^{r}(1-\alpha_i)x_i \overset{\alpha_i \to 0}{\approx} \sum_{i=1}^{r} x_i \tag{5.1.22}$$

When inequality constrains are considered, $(P4)$ takes the following form,

$$\min \sum_{i} x_i \ \text{ s. t.} \sum_{i=1}^{r} \beta_i x_i \leq \zeta,\ x_i \geq 0. \tag{5.1.23}$$

One can see that the optimal solution is $x_i^{\star} = 0, \forall i$, and the optimal transmit/receive filter in $(J1)$ and $(J2)$ is zero. Thus, operating with an inequality power constraint leads to degenerate solution, in the noise-limited regime. Though it might seem that an equality power constraint makes $(J1)$ and $(J2)$ harder to solve, this is not the case as both have non-convex cost functions already. Moreover, the convergence of BCD is unaffected since the globally optimal solution is found for each subproblem (formalized in the next subsection).

Convergence of max-DLT Regarding convergence of the proposed algorithm, max-DLT, it is established using standard BCD convergence results.

Proposition 5.1.2. *Let* $\psi_n \triangleq R_{\Sigma}^{(LB)}(\{\boldsymbol{U}_{l_j}^{(n)}\}, \{\boldsymbol{V}_{l_j}^{(n)}\}),\ n = 1, 2, \ldots$ *be the sequence of iterates for the objective value. Then,* $\{\psi_n\}$ *is non-decreasing in* n*, and converges to a stationary point of* $R_{\Sigma}^{(LB)}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\})$

[2] Although the optimal power allocation to stream i is zero for some streams, i.e., $\boldsymbol{\Sigma}_{(i,i)} = 0$, in the actual implementation of the algorithm, $\boldsymbol{\Sigma}_{(i,i)} = \delta$ where $\delta \ll 1$.

Proof. The sequence ψ_n converges to a stationary point of the objective, since a unique minimizer is found at each step. This follows form BCD convergence results in [Tse01] and [LY73, Chap 7.8]. □

5.2 Generalizing max-SINR

In in part, we generalize the well-known max-SINR algorithm. This is not fully aligned with this part of the thesis, as convergence of this algorithm cannot be shown (similarly to max-SINR). However, we included it for completeness, and due to its superior performance over max-SINR.

5.2.1 Problem Formulation

The intuition is also to lower bound the $(SRM - IMAC)$ in (5.1.2). We make use of the fact that $\log|\boldsymbol{X}|$ is monotonically increasing on the positive-definite cone, i.e.,

$$\log|\boldsymbol{X}_2| \geq \log|\boldsymbol{X}_1|, \text{ for } \boldsymbol{X}_2 \succeq \boldsymbol{X}_1 \succ \boldsymbol{0} \tag{5.2.1}$$

Applying the above property, we lower bound r_{l_j} in (4.4.12) as,

$$\begin{aligned} r_{l_j} &> \log_2|(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}| \\ &= \log_2 \frac{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|} \triangleq \tilde{r}_{l_j}, \ \forall l_j \in \mathcal{I} \end{aligned} \tag{5.2.2}$$

Note that $\tilde{r}_{l_j}$ is an approximation of the actual user rate r_{l_j}, where the approximation error is $\mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}])$ (refer to Appendix 5.6.2). Moreover, bounds such as (5.2.2) are already prevalent in the MIMO literature. Thus, the sum-rate R_Σ can be bounded below, as follows,

$$R_\Sigma > \sum_{l_j \in \mathcal{I}} \tilde{r}_{l_j} = \log_2(\prod_{l_j} q_{l_j}), \text{ where } q_{l_j} \triangleq \frac{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|}$$

Since the log function is monotonic, the sum-rate maximization problem in (5.1.2) is lower bounded as,

$$(SRM) \begin{cases} \max\limits_{\{\boldsymbol{U}_{l_j}, \boldsymbol{V}_{l_j}\}} \prod_{l_j \in \mathcal{I}} \ q_{l_j} \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r \ , \|\boldsymbol{V}_{l_j}\|_F^2 = P_t \ , \forall \ l_j \in \mathcal{I} \end{cases} \tag{5.2.3}$$

Referring to (SRM), q_{l_j} is the so-called Generalized Multi-dimensional Rayleigh Quotient (GMRQ. It is a well-know *separability metric* that is extensively in the study of linear discriminant analysis [Bis06, Chap. 4.1]. In a nutshell, it measure the separation between the signal and IPN subspace. Consequently, given the signal and IPN covariance matrices, $\boldsymbol{R}_{l_j}$ and $\boldsymbol{Q}_{l_j}$, each receiver chooses its filter such to maximize the separation between signal and IPN subspace.

5.2.2 Maximization of Generalized Multi-dimensional Rayleigh Quotient

The main limitation of solving problems such (SRM) is the fact it is not jointly convex in all the optimization variables. Though Block Coordinate Decent (BCD) stands out as a strong candidate, one major obstacle persists: while the problem decouples in the receive filters (as shown in (SRM)), attempting to write a similar expression by factoring out the transmit filters, leads to a coupled problem. Therefore, we propose an alternative (purely heuristic) method: the receive filters are updated as the solution to maximize the sum-rate (assuming fixed transmit filters), while the transmit filters are chosen as the solution of the reverse network sum rate maximization (this same structure is implicitly exploited in max-SINR [GCJ11]), i.e.,

$$(SRM_F)\begin{cases}\max\limits_{\{\boldsymbol{U}_{l_j}\}} \prod_{l_j\in\mathcal{I}} \quad q_{l_j}(\boldsymbol{U}_{l_j}) = \frac{|\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|}\\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r \ , \ \forall\ l_j \ ,\end{cases} \tag{5.2.4}$$

$$(SRM_B)\begin{cases}\max\limits_{\{\boldsymbol{V}_{i_k}\}} \prod_{i_k\in\mathcal{I}} \quad p_{i_k}(\boldsymbol{V}_{i_k}) = \frac{|\boldsymbol{V}_{i_k}^{\dagger}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{V}_{i_k}|}{|\boldsymbol{V}_{i_k}^{\dagger}\bar{\boldsymbol{Q}}_{i_k}\boldsymbol{V}_{i_k}|}\\ \text{s. t. } \|\boldsymbol{V}_{i_k}\|_F^2 = P_t \ , \ \forall\ i_k \ .\end{cases} \tag{5.2.5}$$

In other words, assuming transmit filters as fixed, the receive filters are updated such as to maximize the separability metric in the forward phase. Similarly, the transmit filters are chosen to maximize the separability in the backward training phase. Moreover, as seen from the above problems, the objective in each subproblem is invariant to scaling of the optimal solution. Thus, they can be solved as unconstrained problems, and optimal solutions can be scaled, without loss of optimality.

Before we proceed, we first require a solution to the GMRQ maximization. The solution to this problem was earlier proposed in [Pri03]. We provide a more generic solution to the problem (Appendix 5.6.4).

Lemma 5.2.1. *Consider the following maximization of the r-dimensional GMRQ,*

$$\boldsymbol{X}^{\star} \triangleq \underset{\boldsymbol{X}\in\mathbb{C}^{n\times r}}{\operatorname{argmax}}\ q(\boldsymbol{X}) = \frac{|\boldsymbol{X}^{\dagger}\boldsymbol{R}\boldsymbol{X}|}{|\boldsymbol{X}^{\dagger}\boldsymbol{Q}\boldsymbol{X}|} \ , \tag{5.2.6}$$

where $\boldsymbol{Q}\in\mathbb{S}_{++}^{n\times n}$, $\boldsymbol{R}\in\mathbb{S}_{+}^{n\times n}$ and $r<n$. The optimal solution to this non-convex problem is given by

$$\boldsymbol{X}^{\star} = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\hat{\boldsymbol{V}} \ , \tag{5.2.7}$$

where

$$\begin{aligned}&\boldsymbol{L}\boldsymbol{L}^{\dagger} = \boldsymbol{Q},\ \boldsymbol{L}\in\mathbb{C}^{n\times n} \ ,\\ &\boldsymbol{\Psi} = v_{1:r}[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}],\ \boldsymbol{\Psi}\in\mathbb{C}^{n\times r} \ ,\end{aligned}$$

and $\hat{\boldsymbol{V}} \in \mathbb{C}^{r\times r}$ *is an arbitrary non-singular square matrix.*

Proof. Refer to Appendix 5.6.4 □

It is worth mentioning that the above solution is a generalized formulation of the well-known generalized eigenvalues solution: this result was also obtained in [Pri03].

Corollary 5.2.1. *Consider a special case of* (5.2.7) *where* $\hat{\boldsymbol{V}} = \boldsymbol{I}_r$. *Then, this corresponds to the generalized eigenvalues solution, i.e.,*

$$\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi} \Leftrightarrow \boldsymbol{R}\boldsymbol{X}^\star = \boldsymbol{Q}\boldsymbol{X}^\star\boldsymbol{\Lambda}_r \tag{5.2.8}$$

where $\boldsymbol{\Lambda}_r \in \mathbb{R}^{r\times r}$ *be the (diagonal) matrix of eigenvalues for* $\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$.

Proof. Refer to [Pri03]. □

With this in mind, we can write the optimal transmit and receive filter updates, as follows,

$$\begin{aligned}\boldsymbol{U}_{l_j}^\star &= \boldsymbol{L}_{l_j}^{-\dagger}\boldsymbol{\Psi}_{l_j}, \quad \boldsymbol{\Psi}_{l_j} \triangleq v_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}] \ , \ \forall \ l_j \ , \\ \boldsymbol{V}_{i_k}^\star &= \boldsymbol{K}_{i_k}^{-\dagger}\boldsymbol{\Theta}_{i_k}, \quad \boldsymbol{\Theta}_{i_k} \triangleq v_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}] \ , \ \forall i_k \ ,\end{aligned} \tag{5.2.9}$$

where we used the fact we can set $\hat{\boldsymbol{V}} = \boldsymbol{I}_d$ in the solution of (5.2.7). We note that the optimal filter updates for the transmitter are more heuristic than the receiver ones: While the receive filter updates directly maximizes a lower bound on the sum-rate - as seen in (SRM), no such claim can be made about the transmit filter updates. The details of our algorithm, Alternating Iterative Maximal Separation (AIMS), are shown in Algorithm 4 (where T denotes the number of F-B iterations).

Algorithm 4 Alternating Iterative Maximal Separation (AIMS)

for $t = 1, 2, ..., T$ **do**
  // *forward network optimization: receive filter update*
  Estimate $\boldsymbol{R}_{l_j}, \boldsymbol{Q}_{l_j}$, and compute $\boldsymbol{L}_{l_j}$, $\forall l_j$
  $\boldsymbol{U}_{l_j} \leftarrow \boldsymbol{L}_{l_j}^{-\dagger} v_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}], \quad \forall l_j$
  $\boldsymbol{U}_{l_j} \leftarrow \sqrt{P_r}\, \boldsymbol{U}_{l_j}/\|\boldsymbol{U}_{l_j}\|_F$
  // *reverse network optimization: transmit filter update*
  Estimate $\bar{\boldsymbol{R}}_{i_k}, \bar{\boldsymbol{Q}}_{i_k}$, and compute $\boldsymbol{K}_{i_k}$, $\forall i_k$
  $\boldsymbol{V}_{i_k} \leftarrow \boldsymbol{K}_{i_k}^{-\dagger} v_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}], \quad \forall i_k$
  $\boldsymbol{V}_{i_k} \leftarrow \sqrt{P_t}\, \boldsymbol{V}_{i_k}/\|\boldsymbol{V}_{i_k}\|_F$
end for

A few comments are in order at this stage, regarding the difference between AIMS and max-SINR. Referring to (SRM_F) and (SRM_B), it is clear that our proposed algorithm reduces to max-SINR, in case of single-stream transmission,

i.e., setting $d = 1$. Moreover, an inherent property of the max-SINR solution is that it yields equal power allocation across all the streams (since the individual columns of each transmit/receive filter are normalized to unity). However, as evident from (5.2.9), our proposed solution does not normalize the individual columns of the receive filter, but rather the whole filter norm (as seen in Algorithm 4). This allows for different power allocation, across columns of the same filter. That being said, the proposed solution is expected to yield better sum-rate performance (w.r.t. max-SINR), especially in the interference-limited regime. This is due to the intuitive fact that much can be gained from allocating low power to streams that suffer from severe interference, and higher power to streams with lesser interference (this will be validated in the numerical results section). We next introduce a rank adaptation mechanism that further enhances the interference suppression capabilities of the algorithm.

5.2.3 AIMS with Rank Adaptation

We introduce one additional (heuristic) mechanism to robustify AIMS against severely interference-limited scenarios, by introducing a mechanism of Rank Adaptation (RA): in addition to the transmit / receive filter optimization (Lemma 5.2.1), the latter allows the filter rank to be optimized as well. Mathematically speaking, RA addresses the following problem,

$$r^\star \triangleq \underset{r}{\operatorname{argmax}} \left[\boldsymbol{X}^\star \triangleq \underset{\boldsymbol{X}\in\mathbb{C}^{n\times r}}{\operatorname{argmax}} \frac{|\boldsymbol{X}^\dagger \boldsymbol{R} \boldsymbol{X}|}{|\boldsymbol{X}^\dagger \boldsymbol{Q} \boldsymbol{X}|} \right], \tag{5.2.10}$$

Using the same argument as Lemma 5.2.1, one can verify that $\boldsymbol{X}^\star$ and $r^\star$ are as follows,

$$\begin{aligned} \boldsymbol{X}^\star &= [\boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}]_{1:r^\star}\ , \text{ where } \boldsymbol{\Psi} = v_{1:n}[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}] \\ r^\star &= \underset{r}{\operatorname{argmax}}\ |\boldsymbol{\Lambda}_r| = \left|\{i \mid \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger} \geq 1\}\right| \end{aligned} \tag{5.2.11}$$

where $\boldsymbol{\Lambda}_r \in \mathbb{R}^{r\times r}$ is the (diagonal) matrix consisting of the r-largest eigenvalues of $\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$. Simply put, $r^\star$ is the number of eigenvalues greater than one.

When RA is incorporated into AIMS, this mechanism will boost the performance of the algorithm (namely in interference-limited settings). However, one still needs to ensure that the filter ranks for each transmit-receive pair are the same, i.e., $\operatorname{rank}(\boldsymbol{U}_{l_j}) = \operatorname{rank}(\boldsymbol{V}_{l_j})\ \forall\ l_j$. One quick (heuristic) solution is as follows. For each transmit-receive filter pair, compute the optimal filter rank for both the transmit and receive filter, and use the minimum.[3] Needless to say, ensuring this condition requires additional signalling overhead. We thus envision RA, as potential "add-on" for AIMS, when one can afford the resulting overhead increase. In fact, rank-reduction offers a trade-off between reducing interference and diversity of the signal:

[3] Alternately, one can apply RA to the receive filters only, in the last iteration of the algorithm, since the transmit filter updates are more heuristic than the receive filter updates.

however, in interference-limited scenarios, the former is more critical than the latter. This will be validated in the numerical results section.

5.3 Practical Aspects

5.3.1 Comparison

A few remarks are in order at this stage, regarding the similarities and differences between AIMS and max-DLT. Referring to the optimal update equations for each algorithm, i.e., (5.2.9) and (5.1.19), we clearly see that both span the same subspace, i.e. the generalized eigenspace between the signal and IPN covariance matrices. In addition, max-DLT computes the optimal power allocation for each stream. Despite this significant similarity among the two solutions, recall that they are derived from two fundamentally different problems. While (5.2.9) is a heuristic (an extension of max-SINR) that greedily maximizes the separability at each BS and user, the updates in (5.1.19) maximize a lower bound on the sum-rate capacity (and are shown to converge to a stationary point of the DLT bound). That being said, their performance evaluation is done via numerical results.

5.3.2 Benchmarks

As mentioned earlier, we will also investigate the proposed approach in alternate scenarios such as MIMO IBC, and the MIMO Interference Channel (MIMO IFC). We benchmark our algorithms against widely adopted ones,

- *max-SINR* [GCJ11] in the MIMO IMAC, MIMO IFC and MIMO IBC
- *MMSE and Weighted-MMSE* [PH11, SRLH11] in the MIMO IFC and MIMO IBC

as well as relevant fast-converging algorithms,

- *CCP-WMMSE* [NLN14]: an accelerated version of WMMSE algorithm (using CCP), for the MIMO IMAC
- *IWU* [GKBS15]: a fast-convergent leakage minimization algorithm for the MIMO IFC

Both IWU and CCP-WMMSE reply on the use of turbo iterations, where I inner-loop iterations are carried within each main F-B iteration. While those turbo iterations are carried at the BS/UE in the case of IWU, they are run over-the-air in the case of CCP-WMMSE, thus leading to higher overhead. We note as well that earlier works applied SCA to MIMO IBC settings, e.g., [KTJ12], but their algorithms are restricted to beamforming/combining (no precoding).

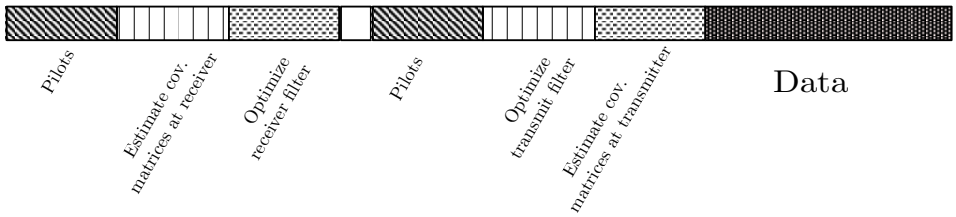

Figure 5.1: Basic structure of Forward-Backward Iteration

5.3.3 Distributed CSI Acquisition

We underline in this section some practical issues that relate to the proposed schemes, such as the mechanism for distributed CSI acquisition, and the resulting communication overhead and computational complexity. Although additional issues such as robustness have to considered as well, such matters are outside the scope of the current paper. We reiterate the fact that CSI acquisition mechanisms are outside the scope of the paper (we refer the reader to [BB15a]). We just summarize the basic operation behind F-B iterations.

Evidently, the operation of such schemes is contingent upon each transmitter / receiver being able to estimate the signal and the IPN covariance matrices, in a fully distributed manner. From the perspective of this work, this is accomplished via the use of *precoded pilots* to estimate the *effective channels*.[4] In the forward phase, the signal covariance matrix, as well as the IPN covariance matrices, can be computed after estimating the effective signal channel, and the effective interfering channels, respectively. The receive filters at the base stations are updated following any of the proposed algorithms (summarized in Fig. 5.1). Then, in the downlink phase, the same procedure is used to estimate the signal and IPN covariance matrices, and update the filters at the receivers. This aforementioned process constitutes one forward-backward (F-B) iteration. Recall that T is the total number of such iterations that are carried out.

5.3.4 Communication Overhead

Thus, for such schemes to be fully distributed, the required CSI quantities have to be obtained via uplink-downlink pilots. Each F-B iteration has an associated communication overheard, namely that of bi-directional transmission of pilots. We adopt a simplistic definition of the communication overhead, as the number of (minimal orthogonal) pilots symbols needed for estimating the required CSI quantities (recall that the actual overhead will be dominated by this quantity). We note that almost all prior algorithms that have been proposed in the context of cellular system, focus on a regime with a high enough number of F-B iterations ($T = 100 \sim 1000$). On

[4] A full investigation of the total overhead of this decentralized solution, as compared to a centralized implementation, falls outside the scope of the current paper.

the contrary, and in line with recent attempts such as [KTJ13, BB15b, GKBS15], we assume that this modus operandi is not feasible in the systems we consider (since F-B iterations are carried out over-the-air, and the associated overhead would be higher than the potential gains). We thus focus on a regime where $T = 2 \sim 5$. In addition, we assume that the minimal number of orthogonal pilots is used, i.e., d orthogonal pilot slots for each uplink/downlink effective channel. Moreover, the pilots are orthogonal across users and cells, resulting in a total of KLd orthogonal pilots for each uplink/downlink training phase. Consequently, the total overhead of both AIMS and max-DLT is approximately,

$$\Omega_{\text{prop}} = T(\underbrace{KLd}_{UL} + \underbrace{KLd}_{DL}) = 2TKLd \text{ channel uses.}$$

It can be verified that the overhead is the same for benchmarks such as max-SINR, IWU and MMSE. Moreover, using similar arguments one can approximate the overhead of CCP-WMMSE and WMMSE, as

$$\Omega_{\text{ccp-wmmse}} = T[(\underbrace{KLM}_{\substack{UL\ chann.\\ estim.}})\underbrace{\times(L-1)}_{\substack{CSI\\ sharing}} + \underbrace{I}_{turbo} \times (\underbrace{KLN}_{\substack{cov.\ mat\\ upd.}})] \text{ c. u.}$$

$$\Omega_{\text{wwmse}} = T(\underbrace{KLd}_{UL} + \underbrace{KLM}_{weights} + \underbrace{KLd}_{DL}) \text{ c.u.}$$

Though a coarse measure, we can see that the overhead associated with WMMSE and its fast-converging variant CCP-WMMSE are significantly higher than that of the proposed schemes. Moreover, CCP-WMMSE exhibits massively larger overhead than the other two, namely due to the fact that the turbo optimization is carried over-the-air (as described in Sec.5.3.2), and that the CSI for the uplink channels is shared among the BSs [NLN14]. The overhead of the aforementioned schemes will be included in the numerical results.

5.3.5 Complexity

Despite the fact that the communication overhead is the limiting resource in cellular networks, we nonetheless shed light on the complexity of the proposed approaches, for completeness. By noticing that operations such as matrix multiplication and bisection search are quite negligible compared to other ones, both AIMS and max-DLT have similar computational complexity: it is dominated by the Cholesky Decomposition of the IPN covariance matrix, and the Eigenvalue Decomposition of $\boldsymbol{M}$, both of which have similar complexity of $\mathcal{O}(N^3)$. Thus, the complexity is dominated by,

$$C_{\text{prop}} = \mathcal{O}(2KL(M^3 + N^3))$$

Note that the same holds for benchmarks such as max-SINR, IWU, and WMMSE since they are dominated by matrix inversion of the IPN covariance matrix. While

the complexity of CCP-WMMSE is also dominated by the above quantity, it also involves running a series of semi-definite programs (using interior point solvers), within each turbo iteration. This renders the algorithm quite costly.

5.4 Numerical Results

5.4.1 Simulation Methodology

We use the achievable sum-rate in the network as the performance metric, where the achievable user rate is given by (4.4.12). Because the approach here is presented in the context of MIMO IMAC, a significant fraction of the results will be under the latter. As mentioned earlier, we will also investigate the proposed approach in alternate scenarios. We specialize our results to some *MIMO IFC* scenarios (a special case of the MIMO IMAC by setting $K = 1$), where interference alignment has been shown to be feasible [YGJK10]. We also investigate the *MIMO IBC* setup, since the proposed algorithms are equally applicable to that case, with little-to-no modification.

As mentioned earlier, we also proposed another algorithm in this work, AIMS (a generalization of max SINR), whose development is not included here. We include it in in this section for completeness. In this work, we assume a block-fading channel model with static users, where channel coefficients are drawn from independent and identically distributed complex Gaussian random variables, with zero mean and unit variance, for the sake of simplicity. We note at this point that we applied our approaches to a much more realistic 5G setup. Since a description of the resulting simulation methodology is rather lengthly, we refer the interested readers to [MET15][Sect. 3.3.3]. In addition, we limit the number of F-B iterations, T, to a small number. We further assume that both the signal and IPN covariance matrices are perfectly estimated at the transmitter / receiver, i.e., we do not model channel estimation errors. Finally, we note that all curves are averaged over 500 channel realizations.

5.4.2 Results for Standard Scenarios

We first investigate the performance of such schemes in conventional canonical scenarios, for benchmarking. We distinguish among feasible, proper and improper setups [YGJK10].

MIMO IFC We start with a feasible MIMO IFC scenario, by setting $M = N = 4, d = 2$ and fixing the number of F-B iterations, $T = 4$, for all schemes. We evaluate the sum-rate of both our algorithms against other well-known algorithms such max-SINR [GCJ11], MMSE [PH11], the rank-reducing algorithm (IWU-RR) earlier proposed in [GKBS15] (since such algorithms are designed for scenarios where fast convergence is desired). We also included Weighted-MMSE with the corresponding number of F-B iterations ($T = 4$), and a large enough number of iterations (as an

upper bound). It is clear from Fig. 5.2 that while max-DLT has similar performance as W-MMSE (for $T = 4$) in the low-to-medium SNR regime, the gap increases in the high-SNR region (SNR $\geq$ 20 dB). Moreover, we note that our proposed schemes, outperform all other benchmarks, across all SNR regimes. In particular, the performance gap between max-DLT and the benchmarks, is quite significant in the medium-to-high SNR region. Moreover, despite the fact that only the rank-reducing scheme and max-DLT are able to achieve some degrees-of-freedom gain, max-DLT offers a 35% gain over the rank reducing scheme. Though max-DLT and IWU-RR are able to turn off some streams in view of reducing interference, the significant performance gap is due to the fact that max-DLT also optimizes the signal subspace as well. Finally, we note that the high-SNR performance of max-DLT is indicative of the fact it is able to achieve some to degrees-of-freedom, while the others algorithms seem to have a significant amount of residual interference, i.e., no degrees-of-freedom gain. Note that the 'optimal-performance' of WMMSE is achieved for $T = 200$, but the resulting overhead is massive. Although the performance of max-DLT is similar to WMMSE ($T = 4$) in low-to-medium SNR regime, the overhead is much lower for the former. Moreover, the gap increases in the medium-to-high SNR region.

MIMO IMAC We next evaluate MIMO IMAC setting with $L = 2, K = 2, M = 4, N = 4, d = 2$, as a function of the number of F-B iterations, T. We also benchmark against CCP-WMMSE (summarized in Sec. 5.3.2) by varying the number of turbo iterations I, and testing the resulting performance and overhead. Fig. 5.3 clearly shows the fast-converging features of both algorithms. More specifically, this is apparent in the case of max-DLT, that reaches 95% of its performance in 2 iterations. While the performance of max-DLT is slightly better than CCP-WWMMSE for $I = 1$, the overhead of the latter is twice that of the former (CCP-WMMSE becomes better than max-DLT for $I = 2$, but the resulting overhead is thrice as high). Note that the 'full' performance CCP-WMMSE is achieved for $I = 50$, but the the resulting overhead (and complexity) are orders-of-magnitude larger than the proposed schemes. Its performance is quite sensitive to solving the inner problem to optimality (i.e., until the turbo iteration converges), thus making it ill-suited for larger setups. Indeed, the running time of CCP-WMMSE (using a mosek solver in CVX) prevented us form testing its performance for larger antenna configurations.

Proper MIMO IBC As mentioned earlier, our schemes are equally applicable to MIMO IBCs. For that matter, we investigate their performance in a proper MIMO IBC setup with 8×2 (single-stream) MIMO links, with $L = 3, K = 3$ (for several SNR values). We benchmark our results against the well-known Weighed-MMSE (WMMSE) algorithm [SRLH11]. Note that for the latter, we keep the sum-power constraint that is employed by WMMSE, and adjust the per-user transmit power constraint P_t, for our algorithms, assuming equal power allocation among users.[5]

[5]If ρ is the per-BS sum-power constraint for WWMMSE, then ρ/K is the per-user transmit power constraint for our algorithms.

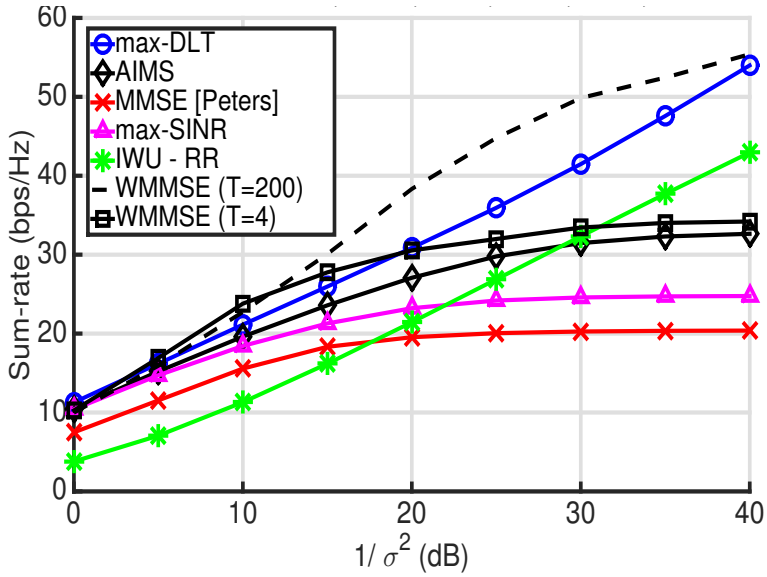

Figure 5.2: Ergodic sum-rate for $L = 3, K = 1, M = N = 4, d = 2, T = 4$ (feasible MIMO IFC)

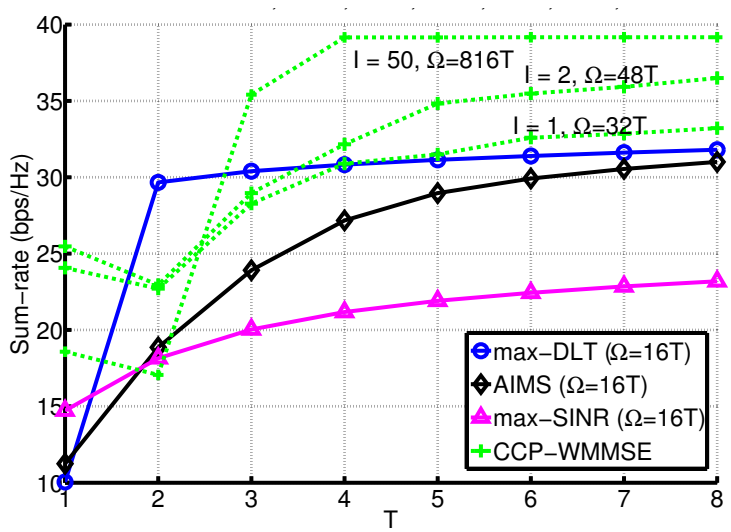

Figure 5.3: Ergodic sum-rate for $L = 2, K = 2, M = N = 4, d = 2$ (Improper MIMO IMAC)

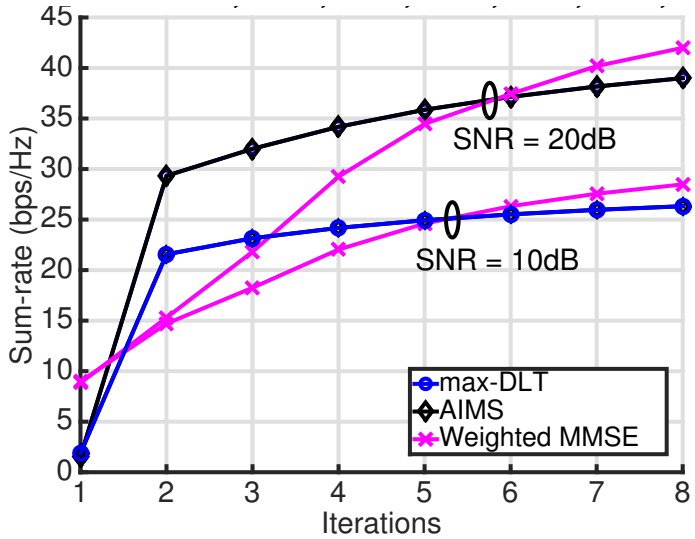

Figure 5.4: Ergodic sum-rate for $L = 3, K = 3, M = 8, N = 2, d = 1$ (proper MIMO IBC)

This implies that a more stringent constraint is placed on our schemes. Despite this unfavorable setup, as Fig. 5.4 shows, both our schemes significantly outperform the benchmark, in the low-overhead regime, i.e., for small T. We reiterate the fact that this is the regime of interest in this work. Needless to say, the full-performance that WMMSE is expected to deliver, is reached after more iterations are performed. As for the overhead, it is $\Omega = 18T$ for our schemes, while is $\Omega = 36T$ for WMMSE.

Effect of non-homogeneous waterfilling The fast-converging behavior is due to the fact that the non-homogeneous waterfilling solution in max-DLT can freely allocate different powers to different directions of the subspace spanned by the transmit / receive filter (possibly turning off some directions, when they suffer from significant interference). As a result, it transforms an improper system, into a (virtually) proper one by turning off one of the streams, for each user (for this particular case). We attempt to capture the latter effect for an improper system with $L = 3, K = 3, M = M = 10$ with $d = 2$. We simulate the average value of the smallest singular value of the transmit filter (across all users),

$$\gamma(\text{SNR}) \triangleq \mathbb{E}\{\frac{1}{|\mathcal{I}|}\sum_{l_j \in \mathcal{I}} \sigma_d^2[\boldsymbol{V}_{l_j}]\}$$

for several SNR values. As we can see from Table 5.1, max-DLT is able to arbitrarily reduce the smallest singular value of the transmit filters, especially in the very high

Table 5.1: γ(SNR) for each scheme

	−20 dB	**0 dB**	**20 dB**	**40 dB**
max-DLT	0.0522	0.002	$3.34\ 10^{-7}$	$3.31\ 10^{-7}$
AIMS	0.4859	0.3325	0.2055	0.1939
max-SINR	0.0515	0.2790	0.2262	0.2268

SNR (interference-limited) regime: this is a critical, since reducing interference is vital to increasing the sum-rate. Note that this adaptation is clearly not present in the case for AIMS and max-SINR.

5.4.3 Scaling up the system

In this section we evaluate the performance of the proposed schemes in uplink and downlink scenarios where much more antennas are available at the BS, than the users.

Large-scale Multi-user Multi-cell MIMO uplink We leverage the larger number of antenna available at the BS. We evaluate a large-scale (in the number of antennas at the BS) multi-user multi-cell uplink with $L = 5, K = 5, d = 2, M = 4, N = 32$. Fig. 5.5 shows the resulting sum-rate of the proposed schemes (and max-SINR), for $T = 2$ and $T = 4$ (we were unable to include CPP-WMMSE as the resulting simulation time was too high). Recall that for each of the simulated values of T, the overhead is the same for all schemes. We observe that both our schemes outperform max-SINR significantly. In particular, max-DLT offer twice the performance of max-SINR at 5dB (this performance gap increases with the SNR). And while both our schemes show significant performance gain by increasing T, the corresponding gain that max-SINR exhibits is negligible in comparison.

Large-scale Multi-user Multi-cell MIMO downlink We next investigate a dual communication setup of the one just above, exploiting the larger number of transmit antennas at the BS (i.e, setting $M = 32, N = 4$ and all else being the same). We benchmark our results against the well-known WMMSE algorithm [SRLH11]. Note that while WMMSE employs a sum-power constraint, our schemes have a per-user power constraint, and thus assume equal power allocation among the users.[6] This implies that a more stringent constraint is placed on our schemes. Despite this unfavorable setup, both our schemes significantly outperform WMMSE, the gap becoming quite large when SNR = 20dB (as seen in Fig. 5.6).

[6] If P_t is the per user constraint for our schemes, then KP_t is the per-BS sum-power constraint for WWMMSE.

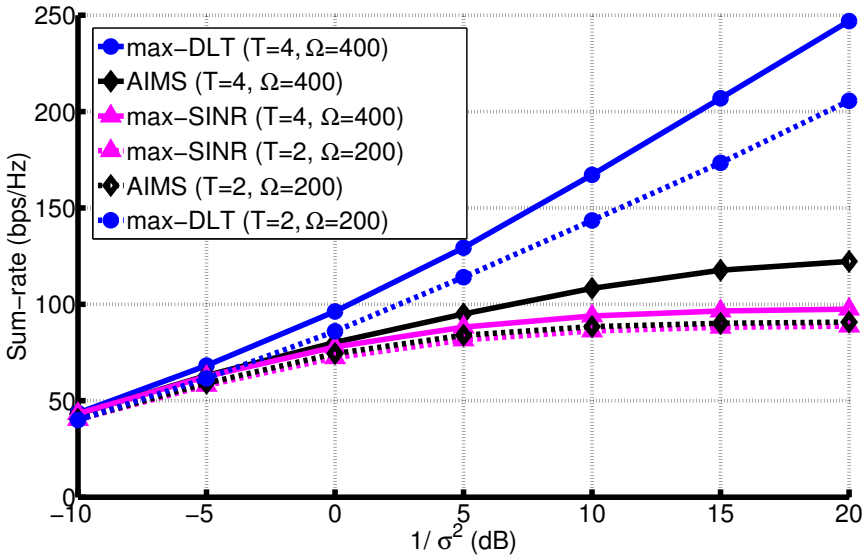

Figure 5.5: Ergodic sum-rate for $L = 5, K = 5, M = 4, N = 32, d = 2$ (Uplink)

Note as well that the overhead of our proposed schemes, is half that of WMMSE as the latter requires feedback of the weights (refer to Sec. 5.3.4 for the overhead calculations). Needless to say, the full-performance that WMMSE is expected to deliver, is reached after more iterations are performed. The reason behind this behavior is the fast-converging nature of our algorithms, allowing them to reach a good operating point, in just 2 iterations. In the case of the max-DLT, this is turn due to the stream control feature of the non-homogeneous waterfilling.

5.4.4 Discussions

As mentioned earlier, the non-homogeneous waterfilling solution clearly shows that streams that have low SINR are turned-off, and power is only allocated to the ones that exhibit relatively high SINR. This greatly speeds up the convergence of max-DLT, and allows it to achieve its required performance, with that limited number of F-B iterations (e.g., 2). On the other hand, due to the large dimensions inherent to low-band mmWave systems (i.e., more antennas, cells, users) other benchmarks will require more iterations to reach a similar performance. As for the overhead, our schemes are based on the framework of F-B iterations and result in minimal overhead (the overhead consisting of uplink/downlink pilots only). However, other schemes such as WMMSE and CCP-WMMSE require additional pilots and feedback, and result in significantly higher overhead (as detailed in Sec. 5.3.4).

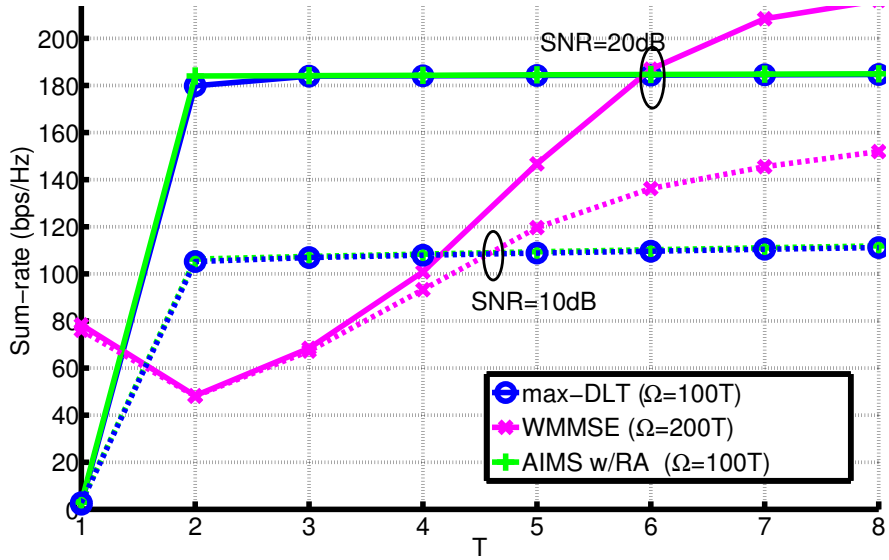

Figure 5.6: Ergodic sum-rate for $L = 5, K = 5, M = 32, N = 4, d = 2$ (Downlink)

5.5 Conclusion

In this chapter, we advocated the use of DLT bounds (as a lower bound on the sum-rate) and highlighted their significant advantage in yielding optimization problems that decouple at both the transmitters and receivers. More importantly, we provided a generic solution for the latter, the so-called non-homogeneous waterfilling: we underlined its built-in stream-control feature, and its role in speeding up the convergence. We proposed a distributed algorithm, max-DLT, that solves the latter problem in a distributed manner. We later verified through simulations that our proposed algorithms massively outperform other relevant benchmark algorithms (especially in interference-limited multi-user environments), for several communication scenarios.

5.6 Appendix

5.6.1 Uniqueness of SVD

The proofs in this work rely on the central premise of mapping the problem into a series of equivalent forms, where equivalence is ensured by the uniqueness of each mapping [BV04]. One of the steps is rewrite the problem by mapping the variable into its SVD form: for example, let $\max_{\boldsymbol{X}} f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^\dagger \boldsymbol{Q} \boldsymbol{X})$ and let $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger$ be the SVD of $\boldsymbol{X}$. Generally speaking, the SVD is unique only up to rotations of the left and right singular vectors, i.e., the actual SVD of $\boldsymbol{X}$ takes this form,

$\boldsymbol{X} = (\boldsymbol{U}\boldsymbol{\Theta})\boldsymbol{\Sigma}(\boldsymbol{\Theta}^\dagger\boldsymbol{V}^\dagger)$ where $\boldsymbol{\Theta}$ is diagonal with phase elements. Due to the quadratic nature of $f(\boldsymbol{X})$ it is easy to verify that the ambiguity brought by $\boldsymbol{\Theta}$ is lifted, i.e., it is easy to verify that $f(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger) = f((\boldsymbol{U}\boldsymbol{\Theta})\boldsymbol{\Sigma}(\boldsymbol{\Theta}^\dagger\boldsymbol{V}^\dagger))$. Note that the same arguments holds for all the objective functions that we use in this work.

5.6.2 Proof of Proposition 5.1.1

We start by lower bounding the user rate in (4.4.12), as

$$\begin{aligned} r_{l_j} &\geq \log_2 |(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1} + (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}| \\ &= \log_2 |(\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}| \\ &= \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}| \\ &\geq \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) \triangleq r_{l_j}^{(LB)} \end{aligned} \tag{5.6.1}$$

where the first inequality follows from combining (5.1.4) and the monotonically increasing nature $\log|\boldsymbol{X}|$ (i.e, $\log|\boldsymbol{X}_1| \geq \boldsymbol{X}_2$, $\forall\ \boldsymbol{X}_1 \succeq \boldsymbol{X}_2 \succ \boldsymbol{0}$). Moreover the last one follows from using $\log|\boldsymbol{A}| \leq \mathrm{tr}(\boldsymbol{A})$ for $\boldsymbol{A} \succeq \boldsymbol{0}$.

We rewrite r_{l_j} in (4.4.12) as,

$$\begin{aligned} r_{l_j} &= \log_2 |(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}[\boldsymbol{I}_d \\ &\qquad + (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}]| \\ &= \log_2 |(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})^{-1}| \\ &\qquad + \log_2 |\boldsymbol{I}_d + (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}| \\ &= \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}| \\ &\qquad + \mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}]) \end{aligned}$$

Thus, r_{l_j} is approximated by $\log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|$ (where the error is given in the above equation). Plugging this result in Δ_{l_j} yields,

$$\begin{aligned} \Delta_{l_j} &= \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}| \\ &\qquad - [\log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})] \\ &\qquad + \mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}]) \end{aligned}$$

Referring to the above, in the interference-limited regime (5.1.4), the first and third terms become negligible w.r.t. the second and fourth. Consequently,

$$\begin{aligned} \Delta_{l_j} &= \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}| \\ &\qquad + \mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}]) \end{aligned}$$

5.6.3 Proof of Lemma 5.1.1

We rewrite the problem into a series of equivalent forms. Letting $\boldsymbol{Z} = \boldsymbol{L}^{\dagger}\boldsymbol{X} \Leftrightarrow \boldsymbol{X} = \boldsymbol{L}^{-\dagger}\boldsymbol{Z}$, then (P) in (5.1.15) is equivalent to,

$$(P2) \begin{cases} \min\limits_{\boldsymbol{Z}} f(\boldsymbol{Z}) \triangleq \mathrm{tr}(\boldsymbol{Z}^{\dagger}\boldsymbol{Z}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{Z}^{\dagger}\boldsymbol{M}\boldsymbol{Z}| \\ \text{s. t. } \mathrm{tr}(\boldsymbol{Z}^{\dagger}\boldsymbol{A}\boldsymbol{Z}) = \zeta \end{cases}$$

where $\boldsymbol{A} = (\boldsymbol{L}^{\dagger}\boldsymbol{L})^{-1}$. Letting $\boldsymbol{Z} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{V}^{\dagger}$ be the SVD of $\boldsymbol{Z}$ ($\boldsymbol{T} \in \mathbb{C}^{n\times r}, \boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$) we rewrite $(P2)$ as into an equivalent form,

$$(P3) \begin{cases} \min\limits_{\boldsymbol{T},\boldsymbol{\Sigma}} \mathrm{tr}(\boldsymbol{\Sigma}^2) - \log_2 |\boldsymbol{I}_d + \boldsymbol{\Sigma}^2\boldsymbol{T}^{\dagger}\boldsymbol{M}\boldsymbol{T}| \\ \text{s. t. } \mathrm{tr}(\boldsymbol{\Sigma}^2\boldsymbol{T}^{\dagger}\boldsymbol{A}\boldsymbol{T}) = \zeta \end{cases}$$

The above problem is separable in $\boldsymbol{T}$, in the sense that the optimal $\boldsymbol{T}$ is independent of $\boldsymbol{\Sigma}$. It can be easily obtained from Hadamard's Inequality i.e., $\boldsymbol{T}^{\star} \triangleq v_{1:r}[\boldsymbol{M}] = \boldsymbol{\Psi}$. Moreover, the feasible set of $(P3)$ becomes $\mathrm{tr}(\boldsymbol{\Sigma}^2\boldsymbol{\Psi}^{\dagger}\boldsymbol{A}\boldsymbol{\Psi}) = \sum_i \sigma_i^2\beta_i$, where $\{\sigma_i\}$ are the diagonal elements of $\boldsymbol{\Sigma}$. With that in mind, $(P3)$ is equivalent to,

$$\min_{\{\sigma_i\}} \sum_{i=1}^{r} \left(\sigma_i^2 - \log_2(1+\alpha_i\sigma_i^2)\right) \text{ s. t. } \sum_{i=1}^{r} \beta_i\sigma_i^2 = \zeta$$

Letting $x_i = \sigma_i^2$, we can rewrite the problem as,

$$(P4) \begin{cases} \min\limits_{\{x_i\}} \sum_{i=1}^{r} \left(x_i - \log_2(x_i + \frac{1}{\alpha_i})\right) \\ \text{s. t. } \sum_{i=1}^{r} \beta_i x_i = \zeta,\ x_i \geq 0, \forall i \end{cases}$$

$(P4)$ is a generalization of the well-known waterfilling problem: in fact, $(P4)$ reduces to the waterfilling problem, if $\beta_i = 1, \forall i$, and by dropping the first term in the objective. We start by writing the associated KKT conditions.

$$\begin{cases} 1 - (x_i + \alpha_i^{-1})^{-1} + \mu\beta_i - \lambda_i = 0,\ \forall i \\ \sum_i \beta_i x_i = \zeta,\ x_i \geq 0 \\ \lambda_i x_i = 0,\ \lambda_i \geq 0,\ \mu \neq 0, \forall i \end{cases}$$

Firstly, note that λ_i act as slack variables and can thus easily be eliminated. Then, considering two cases, $\lambda_i = 0, \forall i$ or $\lambda_i > 0, \forall i$, the optimal solution can be found in a straightforward manner,

$$\begin{aligned} x_i^{\star} &= \begin{cases} (1+\mu\beta_i)^{-1} - \alpha_i^{-1}, \text{ if } \mu < (\alpha_i - 1)/\beta_i \\ 0, \text{ if } \mu > (\alpha_i - 1)/\beta_i \end{cases} \\ &= \left(1/(1+\mu^{\star}\beta_i) - 1/\alpha_i\right)^{+}, \forall i \end{aligned}$$

where $\mu^\star$ is the unique root to

$$g(\mu) \triangleq \sum_{i=1}^{r} \beta_i \Big(1/(1+\mu\beta_i) - 1/\alpha_i\Big)^+ - \zeta$$

Note that $g(\mu)$ is monotonically decreasing, for $\mu > -1/(\max_i \beta_i)$, and $\mu^\star$ can be found using standard 1D search methods, such as bisection.

Thus, the optimal solution for $(J1)$ is $\boldsymbol{Z}^\star = \boldsymbol{\Psi}\boldsymbol{\Sigma}^\star$(where $\boldsymbol{\Sigma}^\star_{(i,i)} = \sqrt{x_i}, \forall i$), and that of (5.1.15) is $\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\boldsymbol{\Sigma}^\star$

5.6.4 Proof of Lemma 5.1.2

It was shown in [Pri03] that the solution to (5.2.6) is given by, $\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}$. We note that it can verified that this optimal solution is invariant to scaling, i.e., $q(\boldsymbol{X}^\star\boldsymbol{\Sigma}) = q(\boldsymbol{X}^\star)$, unitary rotation, i.e., $q(\boldsymbol{X}^\star\boldsymbol{V}) = q(\boldsymbol{X}^\star)$, and $q(\boldsymbol{X}^\star\boldsymbol{S}) = q(\boldsymbol{X}^\star)$ for any $\boldsymbol{S} \in \mathbb{C}^{r\times r}$ that is non-singular. Thus the generic form of the solution is,

$$\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}(\boldsymbol{\Sigma}\boldsymbol{V}^\dagger\boldsymbol{S}) = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\hat{\boldsymbol{V}}$$

where $\hat{\boldsymbol{V}}$ is square and non-singular.

Chapter 6

Leakage Minimization Algorithms

In this chapter, we propose a low-overhead distributed schemes for transmit and receive filter optimization. In line with the main design goals of this part of the thesis (Sect. 4.3.3), the proposed schemes in this chapter only require a few forward-backward iterations, thus causing minimal communication overhead. For that purpose, we relax the well-known leakage minimization problem, and then propose two different filter update structures to solve the resulting non-convex problem: though one leads to conventional full-rank filters, the other results in rank-deficient filters, that we exploit to gradually reduce the transmit and receive filter rank, and greatly speed up the convergence. Furthermore, inspired from the decoding of turbo codes, we propose a turbo-like structure to the algorithms, where a separate inner optimization loop is run at each receiver (in addition to the main forward-backward iteration). This is illustrated in Fig. 6.1. In that sense, the introduction of this turbo-like structure converts the communication overhead required by conventional methods to computational overhead at each receiver (a cheap resource), allowing us to achieve the desired performance, under a minimal overhead constraint. Finally, we show through comprehensive simulations that both proposed schemes hugely outperform the relevant benchmarks, especially for large system dimensions. Although the algorithms and results in this Chapter are presented in the context of MIMO IC for simplicity, they are still applicable to both MIMO IBC and MIMO IMAC. The notation in this chapter deviates from the thesis notation (defined in Chap. 1), in the following aspect only $\lambda_i[\boldsymbol{Q}]$ denotes the i^{th} eigenvalue of a Hermitian matrix $\boldsymbol{Q}$ (assuming the eigenvalues are sorted in increasing order),

6.1 System Model and Problem Formulation

We hence start from the MIMO IBC signal model (presented in Chap. 4.2.1), we restate the leakage minimization problem as (Chap. 4.4.4),

$$\begin{cases} \min\ \phi(\{\boldsymbol{U}_{l_j}\},\{\boldsymbol{V}_{l_j}\}) = \sum_{l_j\in\mathcal{I}} \mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) \\ \text{s. t. } \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{U}_{l_j} = \boldsymbol{I}_d,\ \boldsymbol{V}_{l_j}^{\dagger}\boldsymbol{V}_{l_j} = \boldsymbol{I}_d,\ \forall l_j \in \mathcal{L} \end{cases} \tag{6.1.1}$$

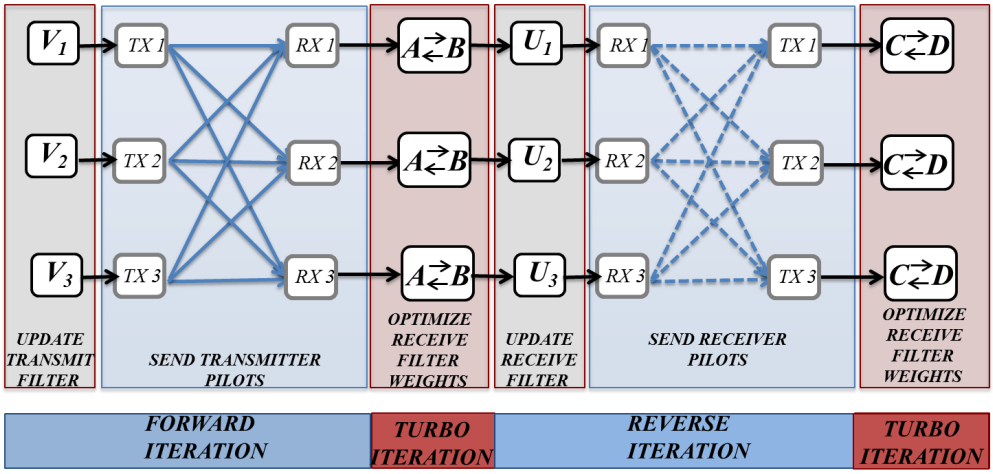

Figure 6.1: Proposed Algorithm Structure, within the framework of F-B training

where we define the interference covariance matrix at receiver $l_j \in \mathcal{I}$ as,

$$\boldsymbol{Q}_{l_j} = \sum_{\substack{i_k \in \mathcal{I} \\ i_k \neq l_j}} (\rho/d) \boldsymbol{H}_{k,l_j} \boldsymbol{V}_{i_k} \boldsymbol{V}_{i_k}^{\dagger} \boldsymbol{H}_{k,l_j}^{\dagger} \tag{6.1.2}$$

and the interference covariance matrix at transmitter l_j, as follows,

$$\bar{\boldsymbol{Q}}_{l_j} = \sum_{\substack{i_k \in \mathcal{I} \\ i_k \neq l_j}} \boldsymbol{H}_{j,i_k}^{\dagger} \boldsymbol{U}_{i_k} \boldsymbol{U}_{i_k}^{\dagger} \boldsymbol{H}_{j,i_k} \tag{6.1.3}$$

Recall that the above leakage minimization problem results in optimizing a lower bound on the sum-rate - albeit a loose one (as shown in Chap. 4.4.4).

6.1.1 Problem Formulation

Now that we have motivated the leakage minimization problem, we turn out attention to devising an iterative algorithm for that purpose. As mentioned earlier, the schemes that we study in this work, fall under the category of distributed schemes, where each receiver / transmitter optimizes its filter, based on the estimated interference covariance matrix. In other words, at the l^{th} F-B iteration, after estimating and updating its interference covariance matrix, $\boldsymbol{Q}_{l_j}^n \leftarrow \boldsymbol{Q}_{l_j}^{n+1}$, receiver l_j aims to update its filter, $\boldsymbol{U}_{l_j}^n \leftarrow \boldsymbol{U}_{l_j}^{n+1}$, such as to optimize some predetermined metric (interference leakage, mean-squared error, sum-rate, etc...). The F-B iteration structure was first applied within the context of IA, in the distributed IA algorithm

(proposed in [GCJ11] and re-written below for later reference), where each receive filter update is such that,

$$\begin{aligned} &\min_{\boldsymbol{U}_{l_j}^{n+1}} f_{l_j}(\boldsymbol{U}_{l_j}^{n+1}) = \mathrm{tr}(\boldsymbol{U}_{l_j}^{n+1^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{U}_{l_j}^{n+1}) \\ &\text{s.t. } \boldsymbol{U}_{l_j}^{n+1^\dagger}\boldsymbol{U}_{l_j}^{n+1} = (P_r/d)\boldsymbol{I}_d \ , \end{aligned} \tag{6.1.4}$$

where P_r is the receive filter power constraint. In other words, in the forward phase each receiver estimates its interference covariance matrix and updates its filter such as to minimize the interference leakage. Then, in the backward phase, exploiting channel reciprocity, transmitters estimate their respective interference covariance matrices, and use the same update rule of minimizing the leakage. It can be shown that this iteration process, will converge to stationary points of the leakage function. Thus, for the interference leakage cost function, F-B iterations can be used to gradually refine the transmit and receive filters, thereby ultimately creating a d-dimensional interference-free subspace at every receiver. Ideally, as $n \to \infty$, the transmit and receive filters that the algorithm yields should satisfy the IA conditions outlined in [GCJ11]. The existence of transmit and receive filters that fulfill this condition is guaranteed, if the system is feasible (as described in [YGJK10]). The distributed IA algorithm has been extensively used and experimentally observed to closely match the theoretical predictions of IA, in small to moderate network configurations. However, one can see that as the dimensions of the problem grow (more antennas and streams), better performance can be achieved be relaxing the unitary constraint.

This sub-optimal performance in multi-stream settings, is partly attributed to the fact that all the streams are allocated the same power - an inherent property of the unitary constraint in (6.1.4). It is evident at this point that much could be gained from allocating different powers to different streams, especially as the number of such streams grow, i.e., as d increases. Consequently, we propose to relax the unitary constraint in (6.1.4), and allow the transmit / receive filter columns to have unequal norms, i.e., the receive filter update $\boldsymbol{U}_{l_j}^{n} \leftarrow \boldsymbol{U}_{l_j}^{n+1}$, is as follows,

$$\begin{aligned} &\min_{\boldsymbol{U}_{l_j}^{n+1}} f_{l_j}(\boldsymbol{U}_{l_j}^{n+1}) = \mathrm{tr}(\boldsymbol{U}_{l_j}^{n+1^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{U}_{l_j}^{n+1}) \\ &\text{s.t. } \|\boldsymbol{U}_{l_j}^{n+1}\|_F^2 = P_r. \end{aligned} \tag{6.1.5}$$

Note that the factor (P_r/d) in (6.1.4) ensures that the receive power constraint, $\|\boldsymbol{U}_{l_j}^{n+1}\|_F^2$, is the same for both (6.1.4) and (6.1.5). Let $\mathcal{R}$ and $\mathcal{S}$ be the feasible sets of (6.1.4) and (6.1.5) respectively, i.e., $\mathcal{R} = \{\boldsymbol{U} \in \mathbb{C}^{N\times d} \mid \boldsymbol{U}^\dagger\boldsymbol{U} = (P_r/d)\boldsymbol{I}_d\}$ and $\mathcal{S} = \{\boldsymbol{U} \in \mathbb{C}^{N\times d} \mid \mathrm{tr}(\boldsymbol{U}^\dagger\boldsymbol{U}) = P_r\}$. Consequently, for any $\boldsymbol{U} \in \mathcal{R} \Rightarrow \boldsymbol{U}^\dagger\boldsymbol{U} = (P_r/d)\boldsymbol{I}_d \Rightarrow \mathrm{tr}(\boldsymbol{U}^\dagger\boldsymbol{U}) = P_r \Rightarrow \boldsymbol{U} \in \mathcal{S}$. This implies that $\mathcal{R} \subseteq \mathcal{S}$, and that indeed (6.1.5) is a relaxation of (6.1.4). In addition, note that the distributed IA problem in (6.1.4) has a simple analytical (well-known) solution. Although the reformulation

in (6.1.5) promises to deliver better performance, it does make the problem non-convex.

In spite of this non-convexity, the problem can still be tackled in many ways. Firstly, note that (6.1.5) can in fact easily be solved by writing the problem in vector form and finding the globally optimal rank-one solution spanned by the eigenvector of $\boldsymbol{Q}_{l_j}^{n+1}$ with the minimum eigenvalue. In addition, it is also known that in the case of (6.1.5), Semi-Definite Relaxation (SDR) provides the optimal solution as well [LMS+10]. However, the solution that both these methods yield is rank-one[1], and it is well-known from the interference alignment literature that the optimal filter rank in the high-SNR regime is d (assuming that d has been selected properly such that the system is feasible). On the other hand, at medium and low-SNR, the sum-rate performance will improve if the filters have reduced rank (in the limit, the waterfilling power allocation results in one stream being active, in the very low-SNR). The main idea behind our proposed algorithm is therefore to *not* solve (6.1.5) but rather to use it as a heuristic, while preventing the algorithm from always converging to the aforementioned rank-one solution of (6.1.5), either explicitly using a rank-preserving algorithm or implicitly by exploiting the transient phase of the rank-reducing algorithm and stopping after a small number of iterations (more about this in Sect. III). As a result, those algorithms should give a better performance than the optimal solution to (6.1.5) given above (simulations will show that this claim is indeed true).

Thus, imposing two different update rules on the transmit / receive filters yields the two different algorithms mentioned above: while one of the update rules do not necessarily result in full-rank transmit / receive filters (which we refer to as *rank-reducing updates*), the other one implicitly enforces full-rank transmit / receive filters (which we refer to as *rank-preserving updates*). The reason for this distinction, as well as its impact, will become clearer in Sect. 6.3.1.

6.2 Rank-reducing Updates

Within this class, we opted to use the most generic update rule (i.e., the one that represents the "widest" class of matrices), for obvious reasons. Thus, we propose the following update structure,

$$\boldsymbol{U}_{l_j}^{n+1} = \boldsymbol{\Delta}_{l_j} \boldsymbol{A}_{l_j}^{n} + \boldsymbol{\Phi}_{l_j} \boldsymbol{B}_{l_j}^{n}, \tag{6.2.1}$$

where $\boldsymbol{\Delta}_{l_j} \in \mathcal{U}(N, d)$ and $\boldsymbol{\Phi}_{l_j} \in \mathcal{U}(N, N-d)$ are such that $\boldsymbol{\Delta}_{l_j}^{\dagger} \boldsymbol{\Phi}_{l_j} = \boldsymbol{0}$. Furthermore, $\boldsymbol{A}_{l_j}^{n} \in \mathbb{C}^{d \times d}$ and $\boldsymbol{B}_{l_j}^{n} \in \mathbb{C}^{(N-d) \times d}$ are the combining weights of $\boldsymbol{\Delta}_{l_j}$ and $\boldsymbol{\Phi}_{l_j}$, respectively.[2] We underline the fact that some choices of $\boldsymbol{\Delta}_{l_j}$ and $\boldsymbol{\Phi}_{l_j}$ should be better

[1] Since the rank is a coarse measure, we use a wider definition of the rank of a matrix, throughout this paper. Let $\boldsymbol{A} \in \mathbb{C}^{n \times m} (n > m)$, then we define $rank(\boldsymbol{A}) = card(\{\sigma_i(\boldsymbol{A}) \mid \sigma_i(\boldsymbol{A}) > \delta, \ \forall \ i = 1, ..., m\})$, where $\{\sigma_1(\boldsymbol{A}), ..., \sigma_m(\boldsymbol{A})\}$ are the singular values of $\boldsymbol{A}$, and δ a predetermined tolerance.

[2] Generally speaking, there are other ways to "partition" the N-dimensional space in question, i.e., $\boldsymbol{\Delta}_{l_j} \in \mathcal{U}(N, r)$, $\boldsymbol{A}_{l_j}^{n} \in \mathbb{C}^{r \times d}$ and $\boldsymbol{\Phi}_{l_j} \in \mathcal{U}(N, N-r)$, $\boldsymbol{B}_{l_j}^{n} \in \mathbb{C}^{(N-r) \times d}$, where $1 \leq r \leq N-1$.

than others, in terms of cost function value. Although this would suggest that they should be optimized within each iteration, a quick look at the resulting optimization problem reveals that the complexity of such a scheme would be tremendously high. As a result, we opt to have the sets $\{\boldsymbol{\Delta}_{l_j}\}$ and $\{\boldsymbol{\Phi}_{l_j}\}$ fixed throughout the algorithm. In addition to the fact that the update rule in (6.2.1) is the most generic possible (i.e., it can represent any matrix), another reason for picking such a structure is that the resulting optimization problem is a relaxation (although a non-convex one) of the optimization solved by the distributed IA [GCJ11] - a result that is formalized in the next subsection.

6.2.1 Relaxation Heuristic

By incorporating the update in (6.2.1) into (6.1.5), the resulting optimization problem is given by,

$$\begin{aligned} \min_{\boldsymbol{U}_{l_j}^{n+1}} \; & f_{l_j}(\boldsymbol{U}_{l_j}^{n+1}) = \mathrm{tr}(\boldsymbol{U}_{l_j}^{n+1^\dagger} \boldsymbol{Q}_{l_j}^{n+1} \boldsymbol{U}_{l_j}^{n+1}) \\ \text{s.t. } & \|\boldsymbol{U}_{l_j}^{n+1}\|_F^2 = P_r \\ & \boldsymbol{U}_{l_j}^{n+1} = \boldsymbol{\Delta}_{l_j} \boldsymbol{A}_{l_j}^n + \boldsymbol{\Phi}_{l_j} \boldsymbol{B}_{l_j}^n \; . \end{aligned} \tag{6.2.2}$$

Since we already proved that (6.1.5) is a relaxation (6.1.4), it remains to show that (6.1.5) is equivalent to (6.2.2) (as defined in [BV04]). Note that this immediately follows from the one-to-one nature of the update in (6.2.1): indeed (6.2.1) should be seen as a one-to-one mapping G, from $\boldsymbol{U}_{l_j}^{n+1}$ to $\boldsymbol{A}_{l_j}^n, \boldsymbol{B}_{l_j}^n$ (for fixed $\boldsymbol{\Delta}_{l_j}$ and $\boldsymbol{\Phi}_{l_j}$), i.e., $G : \boldsymbol{U}_{l_j}^{n+1} \rightarrow G(\boldsymbol{A}_{l_j}^n, \boldsymbol{B}_{l_j}^n)$.

Summarizing thus far, we relaxed the distributed IA problem in (6.1.4), but made the process of solving it more complex. In view of simplifying the solution process, we imposed a structure on the variables of the problem (the update rule in (6.2.1)): generally, this has the effect of constraining the variables to have a particular structure, i.e., adding an additional constraint set $\mathcal{S}$ to the problem. Thus $\mathcal{S}$ needs to be as "wide" as possible, such that it does not alter the feasible region. This is the reason for choosing a generic update rule (that results in $\mathcal{S}$ encompassing a "wide" range of matrices, e.g., unitary).

Although the relaxation argument implies that such a scheme will yield "better" solutions than its distributed IA counterpart, two comments on the latter statement are in order. Firstly, the obvious fact that the solution of the relaxed problem, (6.2.2), will be lower than that of the original problem, (6.1.4), is contingent upon both schemes being able to find the global solutions to their respective problems. Furthermore, since both problems have to be solved at every iteration, it is rather hard to show that at any given iteration, the leakage value for one of the schemes

However, in that case, selecting the best value of r will likely depend on the particular problem instance, and thus will have to be selected based on empirical evidence. Consequently, we set $r = d$ for the sake of simplicity

will be better or worse than the other one (since the sequence $\{\boldsymbol{Q}_{l_j}^n\}_n$ is different for each of the schemes). As a result, although the relaxation argument cannot lead to a rigorous proof of the superiority of any of the schemes, it does provide a well-founded heuristic for adopting such an update rule.

6.2.2 Problem Formulation

Now that we showed that (6.2.2) is a relaxation of (6.1.4), we proceed to rewrite (6.2.2) into a simpler equivalent problem, making use of the following result.

Proposition 6.2.1. *Let $\boldsymbol{U} \in \mathbb{C}^{n\times p}, p < n$, be a given full rank matrix, and $\boldsymbol{Q} \in \mathcal{U}(n,p)$ a unitary matrix. Then there exists $\boldsymbol{A} \in \mathbb{C}^{p\times p}$ and $\boldsymbol{B} \in \mathbb{C}^{(n-p)\times p}$ such that $\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{A} + \boldsymbol{Q}^{\perp}\boldsymbol{B}$, where $\boldsymbol{Q}^{\perp} \in \mathcal{U}(n, n-p)$. Furthermore, $\boldsymbol{A} = \boldsymbol{Q}^{\dagger}\boldsymbol{U}$ and $\boldsymbol{B} = \boldsymbol{Q}^{\perp^{\dagger}}\boldsymbol{U}$.*

Proof. Refer to Appendix 6.7.1 □

As a result, Proposition 6.2.1 implies any $\boldsymbol{U}_{l_j}^{n+1} \in \mathbb{C}^{N\times d}$, can be written as $\boldsymbol{U}_{l_j}^{n+1} = \boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^n + \boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^n$, and consequently, the second constraint in (6.2.2) can be removed without changing the domain of the optimization problem. Then, by applying the one-to-one mapping $G : \boldsymbol{U}_{l_j}^{n+1} \to \boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^n + \boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^n$, we rewrite (6.2.2) as,

$$\begin{aligned} \min_{\boldsymbol{A}_{l_j}^n, \boldsymbol{B}_{l_j}^n} \; & f_{l_j}(\boldsymbol{A}_{l_j}^n, \boldsymbol{B}_{l_j}^n) = \mathrm{tr}[(\boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^n + \boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^n)^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}(\boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^n + \boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^n)] \\ \text{s.t. } & \|\boldsymbol{A}_{l_j}^n\|_F^2 + \|\boldsymbol{B}_{l_j}^n\|_F^2 = P_r \; . \end{aligned} \tag{6.2.3}$$

6.2.3 Turbo Optimization

Due to the fact that f_{l_j} is not jointly convex in $\boldsymbol{A}_{l_j}^n$ and $\boldsymbol{B}_{l_j}^n$, alternately optimizing each of the variables stands out as a possible solution. Furthermore, even when one of the variables is fixed, the resulting optimization problem is still a non-convex one, due to the non-affine equality constraint. Still, it is possible to find the globally optimum solution for each of the variables, as shown in Lemma 6.2.1. By repeating this process multiple times, we wish to produce a *non-increasing sequence* $\{f_{l_j}(\boldsymbol{A}_{l_j}^{n,m}, \boldsymbol{B}_{l_j}^{n,m})\}_m$ (m being the turbo iteration index) that converges to a non-negative limit. Thus, in addition to the main outer F-B iteration, n, we now have an inner loop (or turbo iteration), where $\boldsymbol{A}_{l_j}^{n,m}$ and $\boldsymbol{B}_{l_j}^{n,m}$ are sequentially optimized. With this in mind, for a given $\boldsymbol{B}_{l_j}^{n,m}$, the *sequential updates* $\boldsymbol{A}_{l_j}^{n,m+1}, \boldsymbol{B}_{l_j}^{n,m+1}$ are defined as follows,

$$\underbrace{\boldsymbol{B}_{l_j}^{n,m+1} \triangleq \underset{\boldsymbol{B}}{\operatorname{argmin}} \; f_{l_j}\left(\underbrace{\boldsymbol{A}_{l_j}^{n,m+1} \triangleq \underset{\boldsymbol{A}}{\operatorname{argmin}} \; f_{l_j}(\boldsymbol{A}, \boldsymbol{B}_{l_j}^{n,m})}_{J1}, \; \boldsymbol{B}\right)}_{J2},$$

where the inner optimization problems are elaborated below,

$$(J1):\ \boldsymbol{A}_{l_j}^{n,m+1} = \underset{\boldsymbol{A}}{\operatorname{argmin}}\ f_{l_j}(\boldsymbol{A}, \boldsymbol{B}_{l_j}^{n,m}) \\ \text{s. t. } h_1(\boldsymbol{A}) = \ \|\boldsymbol{A}\|_F^2 + \|\boldsymbol{B}_{l_j}^{n,m}\|_F^2 - P_r = 0\ ,$$

$$(J2):\ \boldsymbol{B}_{l_j}^{n,m+1} = \underset{\boldsymbol{B}}{\operatorname{argmin}}\ f_{l_j}(\boldsymbol{A}_{l_j}^{n,m+1}, \boldsymbol{B}) \\ \text{s. t. } h_2(\boldsymbol{B}) = \|\boldsymbol{B}\|_F^2 + \|\boldsymbol{A}_{l_j}^{n,m+1}\|_F^2 - P_r = 0\ .$$

Remark 6.1. Both $(J1)$ and $(J2)$ are non-convex due to the quadratic equality constraint. Note that applying convex relaxation by replacing the equality by an inequality (thus forming a convex superset) will not help: indeed one can show in that that the sequences of optimal updates within the turbo iteration, are such that $\{\boldsymbol{A}_{l_j}^{n,m}\}_m \to \boldsymbol{0}$ and $\{\boldsymbol{B}_{l_j}^{n,m}\}_m \to \boldsymbol{0}$ (consequently, $\boldsymbol{U}_{l_j}^{n+1} = \boldsymbol{0}$, implying that the algorithm converges to a point that does not necessarily correspond to stationary points of the leakage function).

The following lemma provides the solution to the different subproblems of our proposed algorithms.

Lemma 6.2.1. *Consider the following non-convex quadratic program,*

$$\min_{\boldsymbol{X}}\ f(\boldsymbol{X}) = \operatorname{tr}[(\gamma_1\boldsymbol{\Theta} + \gamma_2\boldsymbol{T}\boldsymbol{X})^\dagger \boldsymbol{Q}(\gamma_1\boldsymbol{\Theta} + \gamma_2\boldsymbol{T}\boldsymbol{X})] \\ \text{s. t. } h(\boldsymbol{X}) = \|\boldsymbol{X}\|_F^2 - \zeta = 0\ ,\ \zeta > 0\ , \tag{6.2.4}$$

where $\boldsymbol{Q} \succeq \boldsymbol{0}$, $\boldsymbol{\Theta} \neq \boldsymbol{0}$, $0 \leq \gamma_1$, $\gamma_2 \leq 1$. *Then, the (globally optimum) solution* $\boldsymbol{X}^\star$ *is given by*

$$\boldsymbol{X}^\star(\mu^\star) = -\gamma_1\gamma_2 \left(\gamma_2^2\ \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T} + \mu^\star\boldsymbol{I}\right)^{-1} \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{\Theta} \tag{6.2.5}$$

where $\mu^\star$ *is the unique solution to*

$$\|\boldsymbol{X}^\star(\mu)\|_F^2 = \zeta$$

in the interval $-\gamma_2^2\ \lambda_1[\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}] < \mu < \gamma_1\gamma_2\|\boldsymbol{\Theta}^\dagger\boldsymbol{Q}\boldsymbol{T}\|_F/\sqrt{\zeta}$. *Moreover,* $\|\boldsymbol{X}^\star(\mu)\|_F^2$ *is monotonically decreasing in* μ, *for* $\mu > -\gamma_2^2\ \lambda_1[\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}]$.

Proof. Refer to Appendix 6.7.2. □

Though it might seem that (6.1.5) can be solved using Lemma 6.2.1, i.e., by setting $\boldsymbol{\Theta} = \boldsymbol{0}$, this does make the necessary and sufficient conditions inconsistent (refer to Appendix 6.7.2). On the other hand, it becomes clear at this point that $(J1)$ is a special case of (6.2.4), by letting $\boldsymbol{X} = \boldsymbol{A}$, $\boldsymbol{\Theta} = \boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^{n,m}$, $\boldsymbol{T} = \boldsymbol{\Delta}_{l_j}$, $\gamma_1 = \gamma_2 = 1$,

$\zeta = P_r - \|\boldsymbol{B}_{l_j}^{n,m}\|_F^2$ (keeping in mind that $\|\boldsymbol{A}_{l_j}^{n,m}\|_F^2 + \|\boldsymbol{B}_{l_j}^{n,m}\|_F^2 = P_r$, $\forall m$, it is evident that $\zeta > 0$). Applying the result of Lemma 6.2.1, we now write the solution to $(J1)$ as,

$$\begin{aligned}\boldsymbol{A}_{l_j}^{n,m+1}(\mu) &= -(\boldsymbol{\Delta}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j} + \mu\boldsymbol{I})^{-1}\boldsymbol{\Delta}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}\boldsymbol{B}_{l_j}^{n,m},\\ \mu \in \{\ \mu \mid\ & g(\mu) = \|\boldsymbol{A}_{l_j}^{n,m+1}(\mu)\|_F^2 + \|\boldsymbol{B}_{l_j}^{n,m}\|_F^2 - P_r = 0\ ,\\ & \mu > -\lambda_1[\boldsymbol{\Delta}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}]\}. \end{aligned} \tag{6.2.6}$$

Since the function $g(\mu)$ is monotonically decreasing, the solution can be efficiently found using bisection.

The process of solving $(J2)$ follows exactly the same reasoning as above. By letting $\boldsymbol{X} = \boldsymbol{B}, \boldsymbol{\Theta} = \boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^{n,m+1}, \boldsymbol{T} = \boldsymbol{\Phi}_{l_j}$, $\gamma_1 = \gamma_2 = 1$, $\zeta = P_r - \|\boldsymbol{A}_{l_j}^{n,m+1}\|_F^2$, $\zeta > 0$. Then, the application of of Lemma 6.2.1 immediately yields the solution to $(J2)$,

$$\begin{aligned}\boldsymbol{B}_{l_j}^{n,m+1}(\mu) &= -(\boldsymbol{\Phi}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j} + \mu\boldsymbol{I})^{-1}\boldsymbol{\Phi}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}\boldsymbol{A}_{l_j}^{n,m+1},\\ \mu \in \{\ \mu \mid\ & g(\mu) = \|\boldsymbol{B}_{l_j}^{n,m+1}(\mu)\|_F^2 + \|\boldsymbol{A}_{l_j}^{n,m+1}\|_F^2 - P_r = 0\ ,\\ & \mu > -\lambda_1[\boldsymbol{\Phi}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}]\}. \end{aligned} \tag{6.2.7}$$

6.2.4 Reverse network optimization

Due to the inherent nature of the leakage function, the reverse network optimization follows the same reasoning as the one presented above. Thus, to avoid unnecessary repetition, we just limit ourselves to stating the results, skipping all the derivations. The update rule for the transmit filter as is set as follows (similarly to (6.2.1)),

$$\boldsymbol{V}_{l_j}^{n+1} = \boldsymbol{\Lambda}_{l_j}\boldsymbol{C}_{l_j}^{n} + \boldsymbol{\Gamma}_{l_j}\boldsymbol{D}_{l_j}^{n}, \tag{6.2.8}$$

where $\boldsymbol{\Lambda}_{l_j} \in \mathcal{U}(M,d)$ and $\boldsymbol{\Gamma}_{l_j} \in \mathcal{U}(M, M-d)$ are such that $\boldsymbol{\Lambda}_{l_j}^{\dagger}\boldsymbol{\Gamma}_{l_j} = \boldsymbol{0}$. Furthermore, $\boldsymbol{C}_{l_j}^{n} \in \mathbb{C}^{d\times d}$ and $\boldsymbol{D}_{l_j}^{n} \in \mathbb{C}^{(M-d)\times d}$ are the combining weights of $\boldsymbol{\Lambda}_{l_j}$ and $\boldsymbol{\Gamma}_{l_j}$, respectively. Then, the resulting sequential optimization problems are given as follows,

$$\begin{aligned}(J3):\ \boldsymbol{C}_{l_j}^{n,m+1} &= \underset{\boldsymbol{C}}{\operatorname{argmin}}\ \bar{f}_{l_j}(\boldsymbol{C}, \boldsymbol{D}_{l_j}^{n,m})\\ &\text{s. t. } h_3(\boldsymbol{C}) = \ \|\boldsymbol{C}\|_F^2 + \|\boldsymbol{D}_{l_j}^{n,m}\|_F^2 - P_t = 0,\end{aligned}$$

$$\begin{aligned}(J4):\ \boldsymbol{D}_{l_j}^{n,m+1} &= \underset{\boldsymbol{D}}{\operatorname{argmin}}\ \bar{f}_{l_j}(\boldsymbol{C}_{l_j}^{n,m+1}, \boldsymbol{D})\\ &\text{s. t. } h_4(\boldsymbol{D}) = \|\boldsymbol{D}\|_F^2 + \|\boldsymbol{C}_{l_j}^{n,m+1}\|_F^2 - P_t = 0,\end{aligned}$$

where P_t is the transmit filter power constraint, and $\bar{f}_{l_j}$, the leakage at transmitter l_j, is given by

$$\begin{aligned}\bar{f}_{l_j}(\boldsymbol{C}_{l_j}^n, \boldsymbol{D}_{l_j}^n) &= \mathrm{tr}(\boldsymbol{V}_{l_j}^{n+1^\dagger} \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{V}_{l_j}^{n+1}) \\ &= \mathrm{tr}[(\boldsymbol{\Lambda}_{l_j} \boldsymbol{C}_{l_j}^n + \boldsymbol{\Gamma}_{l_j} \boldsymbol{D}_{l_j}^n)^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} (\boldsymbol{\Lambda}_{l_j} \boldsymbol{C}_{l_j}^n + \boldsymbol{\Gamma}_{l_j} \boldsymbol{D}_{l_j}^n)].\end{aligned}$$

Again, the same block coordinate descent structure can be employed to optimize the weight matrices $\boldsymbol{C}_{l_j}^n$ and $\boldsymbol{D}_{l_j}^n$. Using the same reasoning as earlier, one can obtain the optimal updates, using the result of Lemma 6.2.1, to yield,

$$\begin{aligned}\boldsymbol{C}_{l_j}^{n,m+1}(\mu) &= -(\boldsymbol{\Lambda}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Lambda}_{l_j} + \mu \boldsymbol{I})^{-1} \boldsymbol{\Lambda}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Gamma}_{l_j} \boldsymbol{D}_{l_j}^{n,m}, \\ \mu \in \{\ \mu \mid g(\mu) &= \|\boldsymbol{C}_{l_j}^{n,m+1}(\mu)\|_F^2 + \|\boldsymbol{D}_{l_j}^{n,m}\|_F^2 - P_t = 0\ , \\ &\qquad \mu > -\lambda_1[\boldsymbol{\Lambda}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Lambda}_{l_j}]\}, \end{aligned} \tag{6.2.9}$$

$$\begin{aligned}\boldsymbol{D}_{l_j}^{n,m+1}(\mu) &= -(\boldsymbol{\Gamma}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Gamma}_{l_j} + \mu \boldsymbol{I})^{-1} \boldsymbol{\Gamma}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Lambda}_{l_j} \boldsymbol{C}_{l_j}^{n,m+1}, \\ \mu \in \{\ \mu \mid g(\mu) &= \|\boldsymbol{D}_{l_j}^{n,m+1}(\mu)\|_F^2 + \|\boldsymbol{C}_{l_j}^{n,m+1}\|_F^2 - P_t = 0\ , \\ &\qquad \mu > -\lambda_1[\boldsymbol{\Gamma}_{l_j}^\dagger \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Gamma}_{l_j}]\}. \end{aligned} \tag{6.2.10}$$

The resulting algorithm, Iteratively Weighted Updates with Rank Reducing updates (IWU-RR) is shown in Algorithm 5.

Algorithm 5 Iterative Weight Update with Rank-Reduction (IWU-RR)

T : number of F-B iterations, I : number of turbo iterations
for $n = 0, 1, ..., T-1$ **do**
  // forward network optimization
  Update receiver interference covariance matrix
  for $m = 0, 1, ..., I-1$ **do**
    Compute $\{\boldsymbol{A}_{l_j}^{n,m+1}\}_{l_j}$ in (6.2.6), $\{\boldsymbol{B}_{l_j}^{n,m+1}\}_{l_j}$ in (6.2.7)
  end for
  Check rank and perform rank-reduction (Remark 6.2)
  Update receive filter in (6.2.1)
  // reverse network optimization
  Update transmitter interference covariance matrix
  for $m = 0, 1, ..., I-1$ **do**
    Compute $\{\boldsymbol{C}_{l_j}^{n,m+1}\}_{l_j}$ in (6.2.9), $\{\boldsymbol{D}_{l_j}^{n,m+1}\}_{l_j}$ in (6.2.10)
  end for
  Check rank and perform rank-reduction (Remark 6.2)
  Update transmit precoder in (6.2.8)
end for

Remark 6.2. If the weight combining matrices at the output of the turbo iteration (for, say, the receive filter update) are rank deficient, then resulting receive filter is rank deficient as well. The rank-reduction process is done by eliminating linearly dependent columns of $\boldsymbol{A}_{l_j}^{n,m}$ and $\boldsymbol{B}_{l_j}^{n,m}$, and appropriately scaling each of them, to fulfill the power constraint.

6.2.5 Convergence Analysis

As shown earlier, although the problem solved within the turbo iteration is a non-convex one, i.e., (6.2.3), we can still show that the application of the updates for $\boldsymbol{A}_{l_j}^{n,m+1}$ and $\boldsymbol{B}_{l_j}^{n,m+1}$ (given in (6.2.6) and (6.2.7), respectively), cannot increase the leakage at each receiver.

Theorem 6.2.1. *For fixed $\{\boldsymbol{V}_{l_j}^{n+1}\}_{l_j}$, the leakage within the receiver turbo iteration is non-increasing, i.e. the sequence $\{f_{l_j}(\boldsymbol{A}_{l_j}^{n,m}, \boldsymbol{B}_{l_j}^{n,m})\}_m$ is non-increasing, and converges to a non-negative limit $f_{l_j}^{n,st} \geq 0$, where $\boldsymbol{A}_{l_j}^{n,m+1}$ and $\boldsymbol{B}_{l_j}^{n,m+1}$ are given in (6.2.6) and (6.2.7).*

Proof. The proof immediately follows from showing that for a fixed F-B iteration number l, the following holds,

$$f_{l_j}(\boldsymbol{A}_{l_j}^{n,m+1}, \boldsymbol{B}_{l_j}^{n,m+1}) \overset{(b)}{\leq} f_{l_j}(\boldsymbol{A}_{l_j}^{n,m+1}, \boldsymbol{B}_{l_j}^{n,m}) \overset{(a)}{\leq} f_{l_j}(\boldsymbol{A}_{l_j}^{n,m}, \boldsymbol{B}_{l_j}^{n,m}), \forall m. \tag{6.2.11}$$

Note that (a) follows immediately from the definition and solution of $(J1)$. Consequently, the application of the update $\boldsymbol{A}_{l_j}^{n,m} \leftarrow \boldsymbol{A}_{l_j}^{n,m+1}$, given by (6.2.6), cannot increase the cost function. Similarly, points that satisfy (6.2.7) mimimize $(J2)$ (as shown by Lemma 6.2.1). Thus, the update $\boldsymbol{B}_{l_j}^{n,m} \leftarrow \boldsymbol{B}_{l_j}^{n,m+1}$ given in (6.2.7) cannot increase the cost function, and (b) follows. Therefore, the sequence $\{f_{l_j}(\boldsymbol{A}_{l_j}^{n,m}, \boldsymbol{B}_{l_j}^{n,m})\}_m$ is non-increasing, and since the leakage function is non-negative, we conclude that $\{f_{l_j}(\boldsymbol{A}_{l_j}^{n,m}, \boldsymbol{B}_{l_j}^{n,m})\}_m$ converges to some non-negative limit $f_{l_j}^{n,st}$. □

With this in mind, not only does Theorem 6.2.1 establish the convergence of the turbo iteration to some limit, but also that the leakage is non-increasing with each of the updates (as immediately seen from (6.2.11)). Although Theorem 6.2.1 shows the convergence of the turbo iteration, to some limit, one cannot claim that this limit corresponds to a stationary point of the function, because the variables in (6.2.3) are coupled [RHL12]. Moreover, recall that we do not wish our algorithm to converge to stationary points of the leakage function since the latter correspond to rank-one solutions (following the discussion in Sect. II-B). Consequently, showing the convergence of the block coordinate descent method to stationary points becomes much less critical in our case, as long as we can establish the non-increasing nature of the leakage. In addition, it is not hard to see that exactly the same reasoning can be used to extrapolate the result of Theorem 6.2.1 to show that the updates for the transmit filter weights (given in (6.2.9) and (6.2.10)), can only decrease the

leakage at the given transmitter, and thus establishing the convergence of the turbo iteration for the transmit filter weights.

6.2.6 Convergence to lower-rank solutions

For convenience, we define T as the maximum number of F-B iterations, and I as the maximum number of turbo iterations, for our algorithm. Strong (empirical) evidence suggests that *the proposed algorithm will gradually reduce the transmit / receive filter rank, and converge to rank-one solutions, as* $T, I \to \infty$. As a result, operating the algorithm with large values of T, I will result in a multiplexing gain of 1 degree-of-freedom per user (highly suboptimal especially if multistream transmission is desired). Conversely, by allowing the algorithm to gradually reduce the rank of a given transmit / receive filter, we exploit the "transient phase" of this algorithm stopping before convergence to rank-one solutions (i.e. for small values of T, I). In addition, recall that reducing the transmit / receive filter rank also reduces the dimension of the interference that is caused to other receivers (this is beneficial in the interference-limited regime): this makes the alignment of interference "easier" and greatly speeds up the convergence. Note as well that although having small values of T, I is extremely desirable (the associated communication and computational overhead will be relatively low), having them too small will evidently result in poor performance, e.g., $T = 0, I = 0$. This does suggest the existence of a trade-off on T and I, between the performance and overhead. Unfortunately, a mathematical characterization of the latter reveals to be impossible, and we will rely on empirical evidence to select them.

6.3 Rank-Preserving Updates

6.3.1 Proposed Update Rule and Problem Formulation

An inherent consequence of the coupled nature of the weight updates for $\boldsymbol{A}_{l_j}^n$ and $\boldsymbol{B}_{l_j}^n$, i.e., (6.2.6) and (6.2.7) (as well as the turbo-like structure of the algorithm), is the fact that if any of the latter are rank-deficient, then the other one will be rank-deficient as well. Moreover, imposing an explicit rank constraint would make the problem extremely hard to solve (since most rank-constrained problems are NP-hard). Alternately, one way to have the algorithm yield full-rank solutions, is to use another update rule (shown below) where this effect is absent, i.e.,

$$\boldsymbol{U}_{l_j}^{n+1} = \sqrt{1 - \beta_{l_j}^{n^2}}\boldsymbol{U}_{l_j}^n + \beta_{l_j}^n \boldsymbol{\Delta}_{l_j}^n \boldsymbol{Z}_{l_j}^n,\ 0 \leq \beta_{l_j}^n \leq 1, \tag{6.3.1}$$

where $\boldsymbol{\Delta}_{l_j}^n \in \mathcal{U}(N, N-d)$ is such that $\boldsymbol{\Delta}_{l_j}^n \subseteq (\boldsymbol{U}_{l_j}^n)^\perp$, $\boldsymbol{Z}_{l_j}^n \in \mathbb{C}^{(N-d)\times d}$ is the combining weight matrix for the receiver update, and $\beta_{l_j}^n$ is the step size for the receive filter update. Note that due to the dependence of the update on the current receive filter, $\boldsymbol{U}_{l_j}^n$, it is easy to verify that $\boldsymbol{U}_{l_j}^{n+1}$ is full rank, if $\boldsymbol{U}_{l_j}^n$ is. In addition, if

both $\boldsymbol{U}_{l_j}^n$ and $\boldsymbol{Z}_{l_j}^n$ satisfy the power constraint, i.e., $\|\boldsymbol{U}_{l_j}^n\|_F^2 = P_r$ and $\|\boldsymbol{Z}_{l_j}^n\|_F^2 = P_r$, then $\|\boldsymbol{U}_{l_j}^{n+1}\|_F^2 = P_r$.

Similarly to (6.2.2), by incorporating the above update structure, the resulting optimization problem at each receiver is stated as follows,

$$\begin{aligned} \min_{\boldsymbol{U}_{l_j}^{n+1}} \; & f_{l_j}(\boldsymbol{U}_{l_j}^{n+1}) = \mathrm{tr}(\boldsymbol{U}_{l_j}^{n+1^\dagger} \boldsymbol{Q}_{l_j}^{n+1} \boldsymbol{U}_{l_j}^{n+1}) \\ \text{s.t. } & \|\boldsymbol{U}_{l_j}^{n+1}\|_F^2 = P_r \\ & \boldsymbol{U}_{l_j}^{n+1} = \sqrt{1-\beta_{l_j}^{n^2}}\boldsymbol{U}_{l_j}^n + \beta_{l_j}^n \boldsymbol{\Delta}_{l_j}^n \boldsymbol{Z}_{l_j}^n. \end{aligned} \tag{6.3.2}$$

A few comments are in order at this point regarding the similarities and fundamental differences between the rank-reducing update proposed earlier, and the rank-preserving update above. Given that both result in non-convex optimization problems, they both rely on a coordinate descent approach to optimize each of their respective variables. In addition, it is clear that the rank-reducing update in (6.2.1) is more generic than the rank-preserving update in (6.3.1). As a result, the relaxation argument that was put forth to motivate the use of the update in (6.2.1) (Sect. 6.2.1), no longer holds here. Furthermore, both algorithms have exactly the same structure: in that sense, after updating its interference covariance matrix, receiver l_j wishes to optimize both its combining weight and step-size, i.e. $\beta_{l_j}^n$ and $\boldsymbol{Z}_{l_j}^n$, such as to minimize the resulting interference leakage at the next iteration. Plugging (6.3.1) into (6.3.2) yields the cost function at receiver l_j,

$$\begin{aligned} f_{l_j}(\beta_{l_j}^n, \boldsymbol{Z}_{l_j}^n) = (1-\beta_{l_j}^{n^2})\mathrm{tr}(\boldsymbol{U}_{l_j}^{n^\dagger} \boldsymbol{Q}_{l_j}^{n+1} \boldsymbol{U}_{l_j}^n) + \beta_{l_j}^{n^2} \mathrm{tr}(\boldsymbol{Z}_{l_j}^{n^\dagger} \boldsymbol{\Delta}_{l_j}^{n^\dagger} \boldsymbol{Q}_{l_j}^{n+1} \boldsymbol{\Delta}_{l_j}^n \boldsymbol{Z}_{l_j}^n) \\ + 2\beta_{l_j}^n \sqrt{1-\beta_{l_j}^{n^2}} Re[\mathrm{tr}(\boldsymbol{U}_{l_j}^{n^\dagger} \boldsymbol{Q}_{l_j}^{n+1} \boldsymbol{\Delta}_{l_j}^n \boldsymbol{Z}_{l_j}^n)]. \end{aligned} \tag{6.3.3}$$

6.3.2 Inner Optimization

Again, we will use block coordinate descent to mitigate the non-convexity of (6.3.3), implying that receiver l_j optimizes both its weight combining matrix and step size ($\boldsymbol{Z}_{l_j}^n$ and $\beta_{l_j}^n$), alternately and sequentially, within the turbo iteration, to produce a non-increasing sequence $\{f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z}_{l_j}^{n,m})\}_m$ that will converge to some non-negative limit. Thus, given $\beta_{l_j}^{n,m}$ at the m^{th} turbo iteration, the *sequential* updates $\boldsymbol{Z}_{l_j}^{n,m+1}$ and $\beta_{l_j}^{n,m+1}$ are chosen, as follows,

$$\underbrace{\beta_{l_j}^{n,m+1} \triangleq \underset{\beta}{\mathrm{argmin}}\; f_{l_j}\left(\beta, \underbrace{\boldsymbol{Z}_{l_j}^{n,m+1} \triangleq \underset{\boldsymbol{Z}}{\mathrm{argmin}}\; f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z})}_{K1}\right)}_{K2},$$

$$\text{where } (K1): \ \boldsymbol{Z}_{l_j}^{n,m+1} = \underset{\boldsymbol{Z}}{\operatorname{argmin}} \ f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z}) \\ \text{s.t. } h_1(\boldsymbol{Z}) = \|\boldsymbol{Z}\|_F^2 = P_r.$$

Note that $(K1)$ in non-convex due the quadratic equality constraint, but can be solved using Lemma 6.2.1 by letting $\boldsymbol{X} = \boldsymbol{Z}$, $\boldsymbol{\Theta} = \boldsymbol{U}_{l_j}^n$, $\boldsymbol{T} = \boldsymbol{\Delta}_{l_j}^n$, $\gamma_1 = \sqrt{1-\beta_{l_j}^{n,m}}$, $\gamma_2 = \beta_{l_j}^{n,m}$, $\zeta = P_r$. Applying the result of Lemma 6.2.1 the optimal update is given by,

$$\boldsymbol{Z}_{l_j}^{n,m+1}(\mu) = -\beta_{l_j}^{n,m}\sqrt{1-\beta_{l_j}^{n,m}}\left(\beta_{l_j}^{n,m}\boldsymbol{\Delta}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}^{n} + \mu I\right)^{-1}\boldsymbol{\Delta}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{U}_{l_j}^{n}, \\ \mu \in \{\ \mu \mid g(\mu) = \|\boldsymbol{Z}_{l_j}^{n,m+1}(\mu)\|_F^2 - P_r = 0, \ \ \mu > -\beta_{l_j}^{n,m}\lambda_1[\boldsymbol{\Delta}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}^{n}]\}. \quad (6.3.4)$$

Given $\boldsymbol{Z}_{l_j}^{n,m+1}$, the optimization for $\beta_{l_j}^{n,m}$ is formulated as follows,

$$(K2): \beta_{l_j}^{n,m+1} = \underset{\beta}{\operatorname{argmin}} \ f_{l_j}(\beta, \boldsymbol{Z}_{l_j}^{n,m+1}) = (1-\beta^2)e_1 \\ + \beta\sqrt{1-\beta^2}e_2 + \beta^2 e_3 \\ \text{s.t. } \ 0 \le \beta \le 1, \quad (6.3.5)$$

where, for the sake of clarity, we let

$$e_1 = \operatorname{tr}(\boldsymbol{U}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{U}_{l_j}^{n}), \\ e_2 = 2Re[\operatorname{tr}(\boldsymbol{U}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}^{n}\boldsymbol{Z}_{l_j}^{n,m+1})], \\ e_3 = \operatorname{tr}(\boldsymbol{Z}_{l_j}^{n,m+1^\dagger}\boldsymbol{\Delta}_{l_j}^{n^\dagger}\boldsymbol{Q}_{l_j}^{n+1}\boldsymbol{\Delta}_{l_j}^{n}\boldsymbol{Z}_{l_j}^{n,m+1}).$$

The main issue that one has to carefully consider while optimizing $\beta_{l_j}^{n,m}$ is that the sign and magnitude of e_2 in (6.3.5) may vary depending on the particular instance and channel realization. Furthermore, we also need to rule out the fact that f_{l_j} might in fact be concave in $\beta_{l_j}^{n,m}$ (since by finding the stationary points, we would be maximizing our cost function), or having many extrema. The result of Lemma 6.3.1 addresses all those issues (whose proof is given in Appendix 6.7.3).

Lemma 6.3.1. *The function $p(x) = (1-x^2)e_1 + x\sqrt{1-x^2}e_2 + x^2e_3$ is convex on the interval $[0,\ 1]$, and thus has a single unique global minimum given by $x^* = \left(\frac{1}{2} + \frac{e_1-e_3}{2\sqrt{(e_1-e_3)^2+e_2^2}}\right)^{1/2}$.*

Proof. Refer to Appendix 6.7.3. □

Lemma 6.3.1 establishes the uniqueness of the solution to $(K2)$, by showing that $f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z}_{l_j}^{n,m+1})$ is indeed convex in $\beta_{l_j}^{n,m}$. Thus, the update for $\beta_{l_j}^{n,m}$ can be

simply expressed as,

$$\beta_{l_j}^{n,m+1} = x^* = \left(\frac{1}{2} + \frac{e_1 - e_3}{2\sqrt{(e_1 - e_3)^2 + e_2^2}}\right)^{1/2}. \tag{6.3.6}$$

6.3.3 Reverse Network Optimization

We again exploit the duality that is inherent to the structure of the leakage function, to apply the same reasoning to obtain the optimal updates for the reverse network optimization phase. Thus, skipping all the details, we limit ourselves to just presenting the results. Similarly to the receiver update, each transmitter updates its filter according to the following rule,

$$\boldsymbol{V}_{l_j}^{n+1} = \sqrt{1 - \alpha_{l_j}^{n^2}}\boldsymbol{V}_{l_j}^{n} + \alpha_{l_j}^{n}\boldsymbol{\Phi}_{l_j}^{n}\boldsymbol{W}_{l_j}^{n},\ 0 \leq \alpha_{l_j}^{n} \leq 1, \tag{6.3.7}$$

where $\boldsymbol{\Phi}_{l_j}^{n} \in \mathcal{U}(M, M-d)$ is such that $\boldsymbol{\Phi}_{l_j}^{n} \in (\boldsymbol{V}_{l_j}^{n})^{\perp}$, and $\boldsymbol{W}_{l_j}^{n} \in \mathbb{C}^{(M-d)\times d}$ is the matrix of combining weights. Thus, the resulting optimization problems solved within the turbo iteration are as follows,

$$\begin{aligned}
(K3): \boldsymbol{W}_{l_j}^{n,m+1} &= \underset{\boldsymbol{W}}{\operatorname{argmin}}\ \bar{f}_{l_j}(\alpha_{l_j}^{n}, \boldsymbol{W}) \\
&\quad \text{s.t.}\ \ h_2(\boldsymbol{W}) = \|\boldsymbol{W}\|_F^2 = P_t, \\
(K4): \alpha_{l_j}^{n,m+1} &= \underset{\alpha}{\operatorname{argmin}}\ \bar{f}_{l_j}(\alpha, \boldsymbol{W}_{l_j}^{n,m+1}) \\
&\quad \text{s.t.}\ \ 0 \leq \alpha \leq 1
\end{aligned}$$

where the interference leakage at transmitter l_j is given by,

$$\begin{aligned}
\bar{f}_{l_j}(\alpha_{l_j}^{n}, \boldsymbol{W}_{l_j}^{n,m}) = (1 - \alpha_{l_j}^{n^2})\mathrm{tr}(\boldsymbol{V}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{V}_{l_j}^{n}) + \alpha_{l_j}^{n^2}\mathrm{tr}(\boldsymbol{W}_{l_j}^{n,m^\dagger}\boldsymbol{\Phi}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}^{n}\boldsymbol{W}_{l_j}^{n,m}) \\
+ 2\alpha_{l_j}^{n}\sqrt{1 - \alpha_{l_j}^{n^2}}Re[\mathrm{tr}(\boldsymbol{V}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}^{n}\boldsymbol{W}_{l_j}^{n,m})].
\end{aligned} \tag{6.3.8}$$

Finally, the optimal updates within the turbo iteration are as follows,

$$\begin{aligned}
&\boldsymbol{W}_{l_j}^{n,m+1}(\mu) = -\alpha_{l_j}^{n}\sqrt{1 - \alpha_{l_j}^{n^2}}\left(\alpha_{l_j}^{n^2}\boldsymbol{\Phi}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}^{n} + \mu I\right)^{-1}\boldsymbol{\Phi}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{V}_{l_j}^{n}, \\
&\mu \in \{\ \mu \mid \|\boldsymbol{W}_{l_j}^{n,m+1}(\mu)\|_F^2 - P_t = 0,\ \ \mu > -\alpha_{l_j}^{n^2}\lambda_1[\boldsymbol{\Phi}_{l_j}^{n^\dagger}\bar{\boldsymbol{Q}}_{l_j}^{n+1}\boldsymbol{\Phi}_{l_j}^{n}]\}.
\end{aligned} \tag{6.3.9}$$

Using Lemma 6.3.1, the optimal update for $\alpha_{l_j}^{n}$ is,

$$\alpha_{l_j}^{n} = \left(\frac{1}{2} + \frac{b_1 - b_3}{2\sqrt{(b_1 - b_3)^2 + b_2^2}}\right)^{1/2}, \tag{6.3.10}$$

where

$$b_1 = \mathrm{tr}(\boldsymbol{V}_{l_j}^{n^\dagger} \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{V}_{l_j}^{n}),$$
$$b_2 = 2Re[\mathrm{tr}(\boldsymbol{V}_{l_j}^{n^\dagger} \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Phi}_{l_j}^{n} \boldsymbol{W}_{l_j}^{n,m+1})],$$
$$b_3 = \mathrm{tr}(\boldsymbol{W}_{l_j}^{n,m+1^\dagger} \boldsymbol{\Phi}_{l_j}^{n^\dagger} \bar{\boldsymbol{Q}}_{l_j}^{n+1} \boldsymbol{\Phi}_{l_j}^{n} \boldsymbol{W}_{l_j}^{n,m+1}).$$

Algorithm 6 Iterative Weight Update with Rank-Preservation (IWU-RP)

T : number of F-B iterations, I : number of turbo iterations
for $n = 0, 1, ..., T-1$ **do**
  // forward network optimization
  Update receiver interference covariance matrix
  for $m = 0, 1, ..., I-1$ **do**
    Compute $\{\boldsymbol{Z}_{l_j}^{n,m+1}\}_{l_j}$ in (6.3.4), $\{\beta_{l_j}^{n,m+1}\}_{l_j}$ in (6.3.6)
  end for
  Update receive filter in (6.3.1)
  // reverse network optimization
  Update transmitter interference covariance matrix
  for $m = 0, 1, ..., I-1$ **do**
    Compute $\{\boldsymbol{W}_{l_j}^{n,m+1}\}_{l_j}$ in (6.3.9), $\{\alpha_{l_j}^{n,m+1}\}_{l_j}$ in (6.3.10)
  end for
  Update transmit precoder in (6.3.7)
end for

6.3.4 Convergence of turbo iteration

The convergence of the turbo iteration (for both the receive and transmit filter updates) can be established using a similar reasoning as the one used in Sect 6.2.5. In other words, we show that the application of each update cannot increase the cost function, i.e.,

$$f_{l_j}(\beta_{l_j}^{n,m+1}, \boldsymbol{Z}_{l_j}^{n,m+1}) \overset{(b)}{\leq} f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z}_{l_j}^{n,m+1}) \overset{(a)}{\leq} f_{l_j}(\beta_{l_j}^{n,m}, \boldsymbol{Z}_{l_j}^{n,m}).$$

The proof follows exactly the same argument in as the one in Theorem 6.2.1, i.e., by showing that the updates $\boldsymbol{Z}_{l_j}^{n,m} \leftarrow \boldsymbol{Z}_{l_j}^{n,m+1}$ in (6.3.4), and $\beta_{l_j}^{n,m} \leftarrow \beta_{l_j}^{n,m+1}$ in (6.3.6) cannot increase the cost function.

6.4 Implementation Aspects and Complexity

The major drawback for previously proposed distributed schemes that rely on F-B iterations, is that they assume a large number of F-B iterations to deliver their

intended performance, ranging from hundreds to thousands (as we shall see in the next section) - a prohibitively high cost since they correspond to actual channel uses between the transmitter and receiver. The chief advantage of the proposed approach is the fact that it greatly reduces the latter communication overhead to a few iterations, while still retaining a very high performance (as simulations will show).

We will use the flop count as a surrogate measure of complexity, although it is well known that the latter is a rather coarse one. Assume for simplicity that $d = N/2 = M/2 \stackrel{\triangle}{=} n$ (this is consistent with the simulation parameters), and denote by $\mathcal{C}$ the complexity per F-B iteration. Note that the latter quantity will be largely dominated by the computationally demanding operations such as matrix product, matrix inversion, and eigenvalue decomposition (EVD). With this in mind, for $\boldsymbol{X} \in \mathbb{C}^{m\times p}, \boldsymbol{Y} \in \mathbb{C}^{p\times n}$, then $\boldsymbol{XY}$ needs $8mnp$ flops. Furthermore, inverting an $n\times n$ matrix requires $2n^3-n$ flops, while computing the EVD of an $n\times n$ Hermitian matrix using the SVD requires $126n^3$ flops[3], resulting in each update in IWU-RR requiring $(67/4)n^3 - n$. Thus, keeping in mind that each iteration involves $2K$ such updates repeated I times, and that EVD is applied to an $n/2 \times n/2$ matrix, the complexity of IWU-RR is

$$\mathcal{C}_{IWU-RR} = (2K)(I)(16n^3 + 67n^3 - 4n) = 2KI(83n^3 - 4n).$$

The same logic applies in the case of IWU-RP, except that each update requires $(53/4n^3 - n)$, to yield

$$\mathcal{C}_{IWU-RP} = 2KI(95/2n^3 - n).$$

Given that the complexity of the bisection method is negligible in comparison with the above, and that the latter depends on the channel realization, and many of the problem parameters (making it extremely difficult to characterize), we have ignored the cost of the bisection method in both cases. Finally, for distributed IA, the cost is largely dominated by the EVD of an $n \times n$ matrix, to yield

$$\mathcal{C}_{DIA} = 2K(126n^3).$$

Since our schemes employ relatively small values of I, the complexity (per F-B iteration) is similar for all the above schemes (albeit slightly lower for distributed IA). However, our simulations generally indicated that our proposed algorithms require a much smaller number of F-B iterations to reach a predetermined tolerance value. Consequently, the overall complexity of our schemes will be much lower.

[3]Generally speaking, the complexity of operations such EVD or SVD, are data dependent: though it is well-known that they are $\mathcal{O}(n^3)$, the exact values depend on the matrix itself. For simplicity, we approximate the complexity of an $n \times n$ SVD as $126n^3$ [GVL96].

6.5 Simulation Results

As stated earlier, the performance of the proposed schemes is largely dependent on the number of F-B iterations T, as well as the number of turbo iterations I. Since any explicit optimization of the latter quantities is a rather tedious task - if not infeasible, we will rely on simulations to evaluate their effect, as well as both algorithms' performance. For that matter, we fix the maximum number of F-B iterations to a small value, e.g., $T = 2$ (since we wish to keep the communication overhead at a low level), and evaluate the algorithms' performance for several values of I. In addition, initializing distributed IA with random rank-d unitary transmit filters, the stopping criterion in all subsequent simulations is a maximum number of F-B iterations T, thus keeping the overhead the same for all schemes. Although in this case, the proposed schemes will have higher computational overhead with respect to distributed IA, this will easily be offset by the gains in performance (as this section will clearly show).

We choose the matrices $\{\mathbf{\Delta}_{l_j}\}, \{\mathbf{\Phi}_{l_j}\}$ (for the receive filter optimization), and $\{\mathbf{\Lambda}_{l_j}\}, \{\mathbf{\Gamma}_{l_j}\}$ (for the transmit filter optimization) as random unitary matrices obtained by applying the QR decomposition to random matrices with Gaussian i.i.d entries. Because the latter matrices are fixed throughout the entire algorithm, we can see that their choice is irrelevant, firstly since it is not based on some a priori channel information (for instance, the performance will improve by choosing $\{\mathbf{\Delta}_{l_j}\}$ to span the range of $\boldsymbol{H}_{l,l_j}$). Moreover, we generate the channel matrices as i.i.d. circular Gaussian random variables, which are stochastically invariant to unitary transformations. All the sum-rate curves are averaged over 1000 channel realizations. We reiterate the important fact that our schemes only optimize the interference subspace, without any regard to the signal or noise. Thus, a comparison with schemes such as max-SINR [CJ08] and (weighted) MMSE [SSB$^+$09], [SRLH11] is somewhat not relevant for this work, since they also optimize the desired signal subspace.

6.5.1 Evolution of Interference Leakage versus T and I

Using insights from the feasibility of IA [RLL12,YGJK10] , we test the robustness of the proposed schemes against the following scenario, known a priori to be infeasible. Though this might seem to put distributed IA at a disadvantage (given that the latter is designed to handle feasible scenarios), scenarios that are known to be feasible are few, and might not always be of practical interest. Thus, robustness to infeasible IA configurations is desirable. Fig. 6.2 shows the (average) evolution of the leakage with the number of F-B iterations, for both our schemes (plotted for several values of I), and distributed IA. Although both schemes outperform distributed IA for any value of I, *the gap between IWU-RR and the benchmark is indeed impressive* (~ 3 to 5 orders of magnitude, depending on the value of I). As expected, this gain stems from the ability of IWU-RR to perform rank-reduction, thereby decreasing the dimension of the interference at the corresponding receiver. In addition we

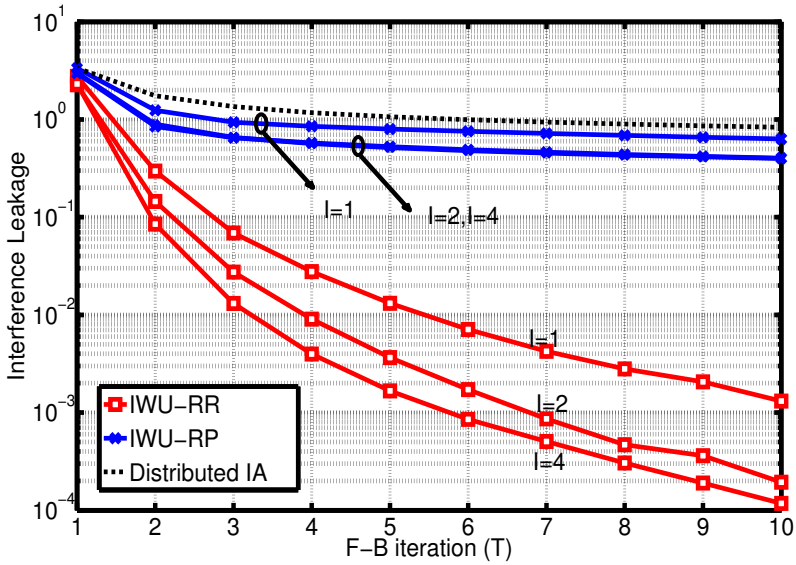

Figure 6.2: Interference Leakage as a function of T, I (4×4 MIMO, 4 users, $d = 2$)

observe that the gain from each additional turbo iteration is decreasing: this is clearly visible in the case of IWU-RP, where the curves corresponding to $I = 2$ and $I = 4$ are almost identical, implying that *only a few turbo iteration are needed to give the desired performance boost.*

6.5.2 Sum-Rate Performance

Next, we simulate the ergodic sum-rate of both our schemes for a 10×10 MIMO, 4-user MIMO IC with $d = 4$, known to be proper [RLL12], and fix the number of F-B iterations to 2, for all algorithms. We use the distributed IA algorithm in [CJ08] as a benchmark, but most importantly, we also include the rank-one solution to (6.1.5), given by SDR. Fig. 6.3 reflects the effect of the turbo iteration on the sum-rate performance of both algorithms: *by running just a few turbo iterations, we see that both schemes significantly outperform distributed IA, especially in the high SNR region, when the gain becomes very large!* In addition we observe that indeed the rank-one solution of SDR offers extremely poor performance in terms of sum-rate (as discussed in Sect. II.B). Moreover, we observe that the high-SNR slope for IWU-RR ($I = 10$) is higher than that of SDR, implying that on average, IWU-RR yields transmit / receive filters whose rank is larger than 1.

Moreover, we observe from Fig. 6.3 that the performance gap is very pronounced, e.g., the high-SNR spectral efficiency of IWU-RR with 10 turbo iterations is almost

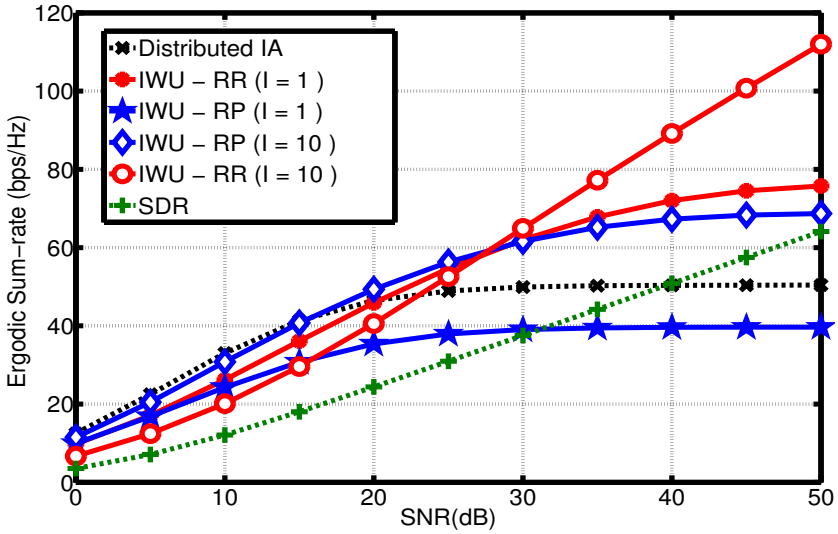

Figure 6.3: Sum-rate of proposed schemes for 10×10 MIMO IC, 4 users ($d = 4$, $T = 2$), for different number of turbo iterations

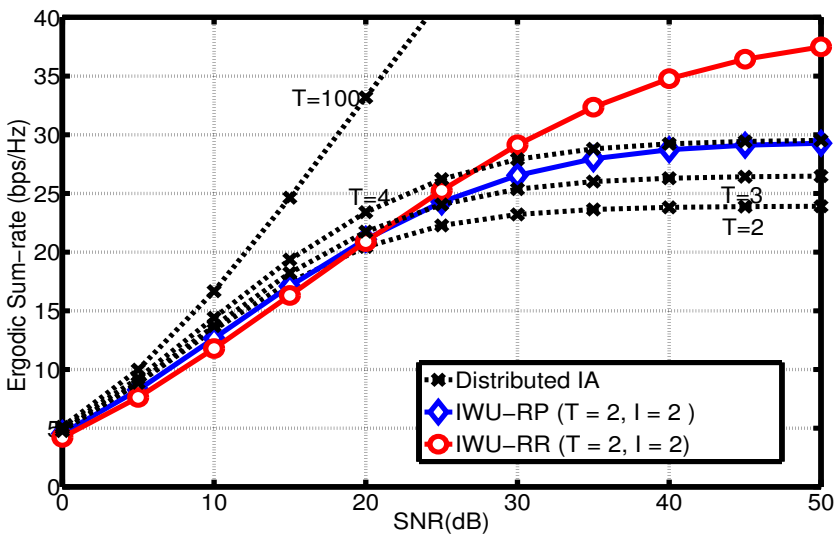

Figure 6.4: Sum-rate of proposed schemes for a 4×4 MIMO IC, 3 users ($d = 2$, $T = 2$), v/s distributed IA for different number of F-B iterations

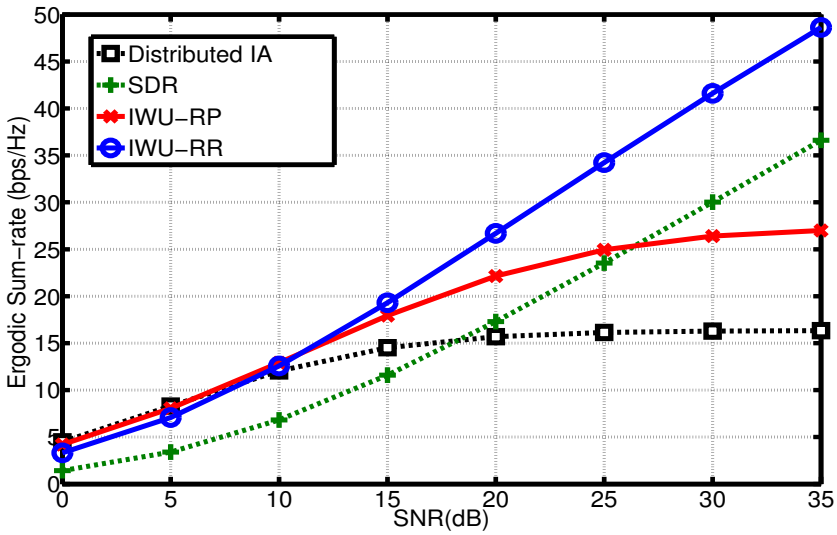

Figure 6.5: Sum-rate performance in more realistic setting (8×8 MIMO IC, 4-users, $d=4$, $T=2$, $I=4$)

double that of distributed IA. Interestingly, note that for $I = 10$ IWU-RR can achieve, albeit not optimal, some degrees-of-freedom gain (shown by the linear scaling of the sum-rate at high SNR), with just 2 F-B iterations. The latter does strongly suggest that *the gains of the current approach become more accentuated, as the dimensions of the system grow.*

Remark 6.3. One might be led to think at this point that the impressive gain in sum-rate for the proposed schemes comes from the fact that the rank reduction transforms the initial IA problem into one of smaller dimensions (while distributed IA is solving the original problem), and thus that the latter simulations do not provide a basis for a fair comparison. However, this argument can be directly refuted by comparing the sum-rate performance of distributed IA, with the rank-preserving scheme (IWU-RP): as seen in Fig. 6.3, although both schemes yield full-rank precoders, IWU-RP still significantly outperforms distributed IA (the gap also increases with the number of turbo iterations, and as the dimensions of the problem grow). This seems to suggest that those gains follow from introducing the turbo iteration (for both schemes), and additionally from solving a relaxed problem (in the case of IWU-RR).

Next, we fix both the number of F-B and turbo iterations in our schemes to 2 and simulate the performance of distributed IA for a varying number of F-B iterations T (for a feasible 4×4 MIMO IC, with $d=2$). Fig. 6.4 clearly shows that for $T=2$ and $T=3$, distributed IA has a similar performance as both our schemes

in the medium-to-low SNR region (and a worse one in the high-SNR region). It is only for $T = 4$ that it starts to outperform them in the medium-to-low SNR region only. This implies that *the overhead requirement of distributed IA is at least* 50% *more than our schemes*, for this particular case (further simulations suggest that this trend increases with the system dimensions). Moreover, we see that distributed IA delivers its "optimal" performance after a large enough number of F-B are run (corresponding to extremely high communication overhead): this suggests that the poor performance of dist IA in all simulations is due to the fact that there is significant interference leakage for small values of T.

6.5.3 Performance in more realistic setup

In view of having a more realistic assessment - albeit still far from accurate - of the algorithms' performance, in somewhat more practical environments, we simulate 8×8 MIMO transmission with 4 cells, 1 user per cell, 4 streams per user (fixing $T = 2$ for all algorithms, and $I = 4$ for our algorithms). We modify the gain of all interfering channels (both intra and inter), such that the resulting SIR is -5dB, to (coarsely) emulate cell edge users. We can see from Fig. 6.5 that though both schemes have a similar performance as distributed IA in the very low-SNR region, they outperform it for SNR values greater than 7dB (the gap being increasing with the SNR): *IWU-RR outperforms distributed IA by* $\sim$ 30% *at* $15dB$ *of SNR, and* $\sim$ 80% *at* $20dB$. This indeed shows that our schemes are good candidates for operating in such practical scenarios. On another note, we also see that IWU-RR and SDR have a similar high-SNR slope (thus implying that IWU-RR finds a rank-one solution in almost all cases). However, the massive gap between IWU-RR and SDR, indicates that the solution provided by the IWU-RR yields significantly higher effective channel gain than the solution found by the SDR.

6.5.4 Discussions

It is interesting to notice in Fig. 6.3-6.5 that the gains for both schemes seem to happen in the medium-to-high SNR region: this is expected, since in that regime, reducing interference is vital to increasing the sum-rate. The observed performance boosts for both IWU-RR and IWU-RP are attributed to the introduction of the turbo-iteration. Furthermore, in the case of IWU-RR, the massive performance gain additionally comes from the fact it is solving a relaxed problem. On another note, Fig. 6.3 shows that indeed the optimal rank-one solution to (6.1.5) provided by SDR is massively suboptimal in terms of sum-rate performance. This also provides a clear motivation for our work, where the proposed algorithms were mainly aimed at avoiding this rank-one solution.

Though negligible, one can indeed see a degradation in performance of both schemes, with respect to distributed IA, in the low-SNR region (as seen from Fig. 6.3-6.6). Despite the that fact that full-rank filters are known to be optimal in the high-SNR regime (thanks to the insights from interference alignment), in the

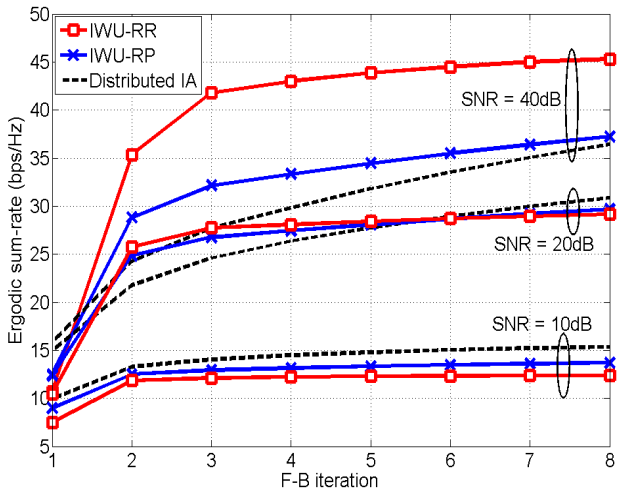

Figure 6.6: Ergodic sum-rate of proposed schemes vs distributed IA, as a function of operating SNR (4×4, 3-user MIMO IC $d = 2$, $T = 2$)

very low-SNR (interference-free) regime however, matched filtering is the optimal strategy, and consequently rank-one filters are optimal as well. We note that although our rank-reducing algorithm does find a rank-one solution, it might be the "wrong one", i.e., different from the matched filtering direction: this is due to the fact that both our algorithm and distributed IA look for solutions that reduce interference, that most likely are not aligned with the matched filtering direction. On the other hand, the full-rank solution given by distributed IA is likely to transmit a reasonable amount of energy along that direction. This might explain the reason that distributed IA exhibits better performance than IWU-RR, in the low-SNR region. Moreover, recall that schemes such as the proposed ones and distributed IA do not take into account the desired signal and noise subspace. As a result, one can at best speculate about their low SNR behavior (since the SNR is not part of their mathematical formulation). However, referring to Fig. 6.6, we can see that this degradation is minimal (around 5% for IWU-RP and 8% for IWU-RR, over the benchmark scheme). A possible alternative to mitigate this issue is to select the scheme based on the operating SNR, i.e., select IWU-RP in the low-SNR region, since it has a similar performance as distributed IA (as seen from Fig. 6.3-6.6): this can be easily implemented since both algorithms have the exact same structure, and only the updates have to be changed.

In conclusion, though both schemes are extremely similar in their algorithmic structure (i.e. both update the filter weights within a turbo iteration), both are distributed, optimize the same metric, and require the same (local) CSI quantities

at each node, they indeed have some fundamental differences. The fact that the filter update equations are different has several implications: the update IWU-RR does not necessarily lead to full rank filters, and though it was shown that IWU-RR attempts to solve the relaxed problem in (4), such claim cannot be made for IWU-RP mainly due to the different constraints on the update structure. Finally, we compared their performance in several scenarios via simulations, and suggested reasons for the behavior we observed.

6.6 Conclusion

Within the context of the leakage minimization problem, we proposed two distinct schemes based on rank-reducing (IWU-RR) and rank-preserving (IWU-RP) filter updates, where the transmit and receive filter weights are iteratively refined in a turbo-like structure. We then showed that they are well suited for delivering high spectral efficiency (compared to the well-known distributed IA algorithm), while generating very small overhead (typically, only a few F-B iterations). Though the introduction of the so-called turbo iteration significantly boosted the performance of both schemes, it is clear that its impact was much more significant when combined with the rank-reducing updates in IWU-RR, thus allowing it to achieve a performance that otherwise required a much larger number of F-B iterations. In that sense, the proposed schemes enabled us to tradeoff the communication overhead associated with the F-B iterations - a rather expensive resource, with computational complexity (an immensely cheaper resource).

6.7 Appendix

6.7.1 Proof of Proposition 6.2.1

Given $\boldsymbol{U}$ and $\boldsymbol{Q}$, and using the fact that $\boldsymbol{Q}$ and $\boldsymbol{Q}^{\perp}$ are unitary and orthogonal, the proof is simple after noting that any subspace $\boldsymbol{U}$ can be expressed as a sum of its components over orthogonal directions (a result that trivially follows from the orthogonal decomposition theorem), i.e. $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{U} + \boldsymbol{P}^{\perp}\boldsymbol{U}$, where $\boldsymbol{P}$ and $\boldsymbol{P}^{\perp}$ are any two orthogonal projection matrices. In particular, let $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^{\dagger}$ and $\boldsymbol{P}^{\perp} = \boldsymbol{Q}^{\perp}\boldsymbol{Q}^{\perp^{\dagger}}$, then $\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{Q}^{\dagger}\boldsymbol{U} + \boldsymbol{Q}^{\perp}\boldsymbol{Q}^{\perp^{\dagger}}\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{A} + \boldsymbol{Q}^{\perp}\boldsymbol{B}$, where $\boldsymbol{A} = \boldsymbol{Q}^{\dagger}\boldsymbol{U}$ and $\boldsymbol{B} = \boldsymbol{Q}^{\perp^{\dagger}}\boldsymbol{U}$

6.7.2 Proof of Lemma 6.2.1

The result is a special case of [BE06], which shows that strong duality holds for all complex valued quadratic problems with up to two quadratic inequality constraints. It is straightforward to show that (6.2.4) and its dual are strictly feasible. Furthermore, since the equality constraint $\|\boldsymbol{X}\|_F^2 = \zeta$ is equivalent to the two inequality constraints $\|\boldsymbol{X}\|_F^2 \leq \zeta$ and $\|\boldsymbol{X}\|_F^2 \geq \zeta$, the results of [BE06] show that the globally optimum solution of (6.2.4) can be obtained from its dual. For the specific

formulation (6.2.4), the solution takes a particularly simple form. Adding the Lagrange multipliers of the two inequality constraints into a single dual variable μ, the necessary and sufficient conditions of [BE06, Theorem 2.4] can be written as

$$\left(\gamma_2^2\, \boldsymbol{T}^\dagger \boldsymbol{Q}\boldsymbol{T} + \mu^\star \boldsymbol{I}\right) \boldsymbol{X}^\star = -\gamma_1\gamma_2\, \boldsymbol{T}^\dagger \boldsymbol{Q}\boldsymbol{\Theta} \tag{6.7.1}$$

$$\|\boldsymbol{X}\|_F^2 = \zeta \tag{6.7.2}$$

$$\left(\gamma_2^2\, \boldsymbol{T}^\dagger \boldsymbol{Q}\boldsymbol{T} + \mu^\star \boldsymbol{I}\right) \succeq 0\ . \tag{6.7.3}$$

The last inequality is fulfilled when $\mu > -\gamma_2^2\ \lambda_1[\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}]$ ($\mu = -\gamma_2^2\ \lambda_1[\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}]$ can be excluded since it results in $\|\boldsymbol{X}\|_F^2 = \infty$). Next, we study $g(\mu) \stackrel{\triangle}{=} \|\boldsymbol{X}^\star(\mu)\|_F^2 - \zeta$. Let $\sigma_1, ..., \sigma_d$ be the eigenvalues of $\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}$ (sorted in increasing order), and $\boldsymbol{v}_1, ..., \boldsymbol{v}_d$ their corresponding eigenvectors. We first rewrite $g(\mu)$ as

$$\begin{aligned} g(\mu) &= \gamma_1^2\gamma_2^2\ \mathrm{tr}[\ \boldsymbol{\Theta}^\dagger\boldsymbol{Q}\boldsymbol{T}(\gamma_2^2\ \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T} + \mu\boldsymbol{I})^{-2}\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{\Theta}\ ] - \zeta \\ &= \gamma_1^2\gamma_2^2\ \mathrm{tr}[\boldsymbol{X}_o^\dagger(\ \gamma_2^2\ \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T} + \mu\boldsymbol{I}\ )^{-2}\boldsymbol{X}_o] - \zeta\ , \end{aligned}$$

where $\boldsymbol{X}_o = \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{\Theta}$. Note that we can express the matrix $(\gamma_2^2\ \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T} + \mu\boldsymbol{I})^{-2}$ as a function of $\sigma_i, \boldsymbol{v}_i, \mu$, as $(\gamma_2^2\ \boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T} + \mu\boldsymbol{I})^{-2} = \sum_{i=1}^d (\gamma_2^2\sigma_i + \mu)^{-2}\boldsymbol{v}_i\boldsymbol{v}_i^\dagger$. Thus, we rewrite $g(\mu)$ as follows,

$$\begin{aligned} g(\mu) &= \gamma_1^2\gamma_2^2\ \mathrm{tr}[\boldsymbol{X}_o^\dagger(\sum_{i=1}^d(\gamma_2^2\sigma_i + \mu)^{-2}\boldsymbol{v}_i\boldsymbol{v}_i^\dagger)\boldsymbol{X}_o] - \zeta \\ &= \sum_{i=1}^d \frac{\gamma_1^2\gamma_2^2\ \mathrm{tr}(\boldsymbol{X}_o^\dagger\boldsymbol{v}_i\boldsymbol{v}_i^\dagger\boldsymbol{X}_o)}{(\gamma_2^2\sigma_i + \mu)^2} - \zeta = \sum_{i=1}^d \frac{(\gamma_1\gamma_2\ c_i)^2}{(\gamma_2^2\sigma_i + \mu)^2} - \zeta, \end{aligned} \tag{6.7.4}$$

where $c_i = \|\boldsymbol{X}_o^\dagger\boldsymbol{v}_i\|_2$. A quick look at this last expression reveals that indeed $g(\mu)$ is strictly monotonically decreasing in μ, for $\mu > -\gamma_2^2\sigma_1 = -\gamma_2^2\lambda_1[\boldsymbol{T}^\dagger\boldsymbol{Q}\boldsymbol{T}]$. Consequently, $g(\mu) = 0$ has a unique solution. To find the upper bound on μ to use in a bisection search, note that if $\mu > 0$ then

$$g(\mu) < \gamma_1^2\gamma_2^2 \sum_{i=1}^d \frac{\|\boldsymbol{X}_o^\dagger\boldsymbol{v}_i\|^2}{\mu^2} - \zeta = \frac{\gamma_1^2\gamma_2^2}{\mu^2}\|\boldsymbol{X}_o^\dagger\boldsymbol{V}_o\|_F^2 - \zeta = \left(\frac{\gamma_1\gamma_2\|\boldsymbol{X}_o^\dagger\|_F}{\mu}\right)^2 - \zeta\ .$$

where $\boldsymbol{V}_o = \begin{bmatrix}\boldsymbol{v}_1 & \dots & \boldsymbol{v}_d\end{bmatrix}$. Consequently if $\mu \geq \gamma_1\gamma_2\|\boldsymbol{X}_o^\dagger\|_F/\sqrt{\zeta}$, we get $g(\mu) < 0$. This concludes the proof.

6.7.3 Proof of Lemma 6.3.1

Let $\mathcal{S}_k$ be the set of local and global minima of $(K2)$, which can written as,

$$\mathcal{S}_k = \{\ x \mid p'(x) = 0,\ p''(x) \geq 0,\ \ 0 \leq x < 1\},$$

where $p(x) = (1-x^2)e_1 + x\sqrt{1-x^2}e_2 + x^2e_3$. We will show that the above set has a single element, thereby establishing that $(K2)$ is a convex problem, and derive the solution.
Defining $a = e_1 - e_3$, we start by finding the zero-differential points of $p(x)$, i.e., $p'(x) = 0 \Rightarrow e_2\frac{1-2x^2}{\sqrt{1-x^2}} = 2ax \Rightarrow 4(a^2+e_2^2)x^4 - 4(a^2+e_2^2)x^2 + e_2^2 = 0$ $(e.1)$ where the last equation stems from squaring both sides. Note that some of the roots of $(e.1)$ will not correspond to zero-differential points (we will remedy this fact later). Letting $X = x^2$, we can write the solution of $(e.1)$ as,

$$X_1 = 1/2 + a/2\sqrt{a^2+e_2^2}, \quad X_2 = 1/2 - a/2\sqrt{a^2+e_2^2}.$$

Moreover, since we are interested in solutions to $(K2)$ that lie in the interval $[0,\ 1]$, we verify that indeed X_1, X_2 lie in this interval. This can be easily done by considering two cases, $a \geq 0$ and $a \leq 0$. Using exactly the same manner, we can show that if $a \leq 0$, then $0 \leq X_1 \leq 1/2$ and $1/2 \leq X_2 \leq 1$, thus concluding that both lie in the interval $[0,\ 1]$. This said, by discarding negative solutions, the solution to $(e.1)$ is $x_1 = \sqrt{X_1},\ x_2 = \sqrt{X_2}$, i.e.,

$$x_1 = \sqrt{X_1} = \sqrt{1/2 + a/2\sqrt{a^2+e_2^2}},$$
$$x_2 = \sqrt{X_2} = \sqrt{1/2 - a/2\sqrt{a^2+e_2^2}}.$$

Note that both x_1 and x_2, lie in the interval $[0, 1]$. Recall that not all the solutions of $(e.1)$ correspond to zero-differential points of $p(x)$ - in fact it is easy to show that $p'(x_1) = 0$ and $p'(x_2) \neq 0$, implying that $p(x)$ has a single unique zero-differential point at x_1. Thus, it remains to show that $p''(x_1) = 0$. Using the fact that $x_1^2 = X_1$, $x_2^2 = X_2$, and noting that $X_1 + X_2 = 1$, we rewrite this condition as,

$$p''(x_1) \geq 0 \Leftrightarrow -2a - e_2\left[\left(\frac{X_1}{X_2}\right)^{3/2} + 3\left(\frac{X_1}{X_2}\right)^{1/2}\right] \geq 0$$

The last equation can be easily shown, by plugging in the values for X_1 and X_2 (we will omit the derivations since they are rather straightforward and easily reproduced).

Thus, we conclude that the set of global minima of $p(x)$, $\mathcal{S}_k$, has a single element, thereby establishing that $p(x)$ is convex and has single global minimum given by $\mathcal{S}_k = \{\sqrt{1/2 + a/2\sqrt{a^2+e_2^2}}\}$.

Part III

Cloud Radio Access Networks

Chapter 7

Antenna Domain Formation

7.1 Densification

In the context of cellular systems (and this thesis in particular), *densification*, refers to having more BSs per unit area, and more antennas per BS. Moreover, *ultra-dense deployments* have been identified as one of the key scenarios for 5G systems [MET14]. From a historical perspective, densification has given the most significant gains in data rates. This is due to the fact that, in general, more BSs / antennas, lead to more degrees-of-freedom. This is of course contingent upon having effective ways of dealing with the resulting interference (since densification results in more interference). Indeed, the fundamental insights provided by techniques such as Interference Alignment [CJ08, MAMK08], and Coordinated Multi-point [GHH$^+$10], clearly state that the effective managing of interference (via coordination among BSs) is necessary to achieve the optimal degrees-of-freedom, in several communication scenarios.

While we have motivated and investigated mechanisms for distributed coordination in Part II (focusing on ones with low-overhead), we attempt to shed light on the opposing paradigm of *centralized coordination*. Intuitively speaking, tighter coordination can be achieved, when BSs stations are connected via high-capacity links to a so-called aggregation node, that can be used to share CSI and/or data among the different BSs. This is the basic setup of the so-called *Cloud Radio Access Network (Could-RAN)*.

Typically, a Cloud-RAN consists of *Remote Radio Heads (RRHs)* (assumed to have limited baseband/processing capabilities) which are connected to the so-called *Aggregation Nodes (ANs)* (assumed to have perfect and global CSI), via wired/wireless links. In that sense, aggregation nodes act as centralized compute nodes, that gather all the required CSI from a cluster of connected radio-heads, perform the required optimization (e.g., precoding), and send the resulting parameters to the relevant radio-heads. An *Antenna Domain (AD)* is the collection of radio-heads connected to a particular aggregation node. It is envisioned that each antenna domain consists of a few (up to tens of) radio-heads, and serves a few dozen (up to a hundred) users. The investigation of such setups, i.e., where base stations

are connected via backhauls, was originally done in [ZQL13].

So far, all such approaches assume the presence of one antenna domain / aggregation node: The problem reduces to managing *intra-AD interference* only, assuming no *inter-AD interference* is present. The authors in [LHMJL13] investigated dynamic clustering of base stations, where users within each cluster are served in a Joint Transmission (JT)-like manner. The same model was adopted in [ZTCY15] and [TCZY15], where the authors consider the problem of forming BS clusters in the presence of caching and multi-cast transmission. A similar model for coordination was employed in [DY16], focusing on energy efficient transmission instead. In [RGIG15], (looser) coordination among the radio-heads within the antenna domain was investigated (where Coordinated Beamforming (CB)-type precoding was employed). Obviously, the model adopted by all such approaches - where *all* the network is coordinated by one aggregation node, is *not scalable*.

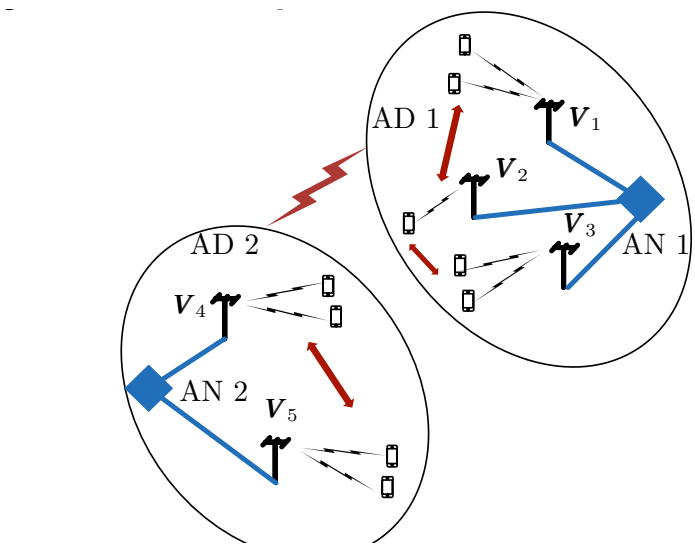

Figure 7.1: Toy Example: intra-AD vs inter-AD interference ($\boldsymbol{V}_1, ..., \boldsymbol{V}_5$ are precoders at the RRHs)

Going to the multi-AD/multi-AN setting, although the management of both inter-AD and intra-AD interference (Fig. 7.1) becomes a critical problem, it remains unaddressed yet. This is what we refer to as the *Antenna Domain Formation (ADF)* problem:

(A) Given a set of radio-heads, each serving a set of users (Fig. 7.2), what is the optimal assignment of radio-heads to aggregation nodes, that minimizes the inter-AD interference?

(B) Given an initial state (i.e., assignment of users to radio-heads, and radio-heads

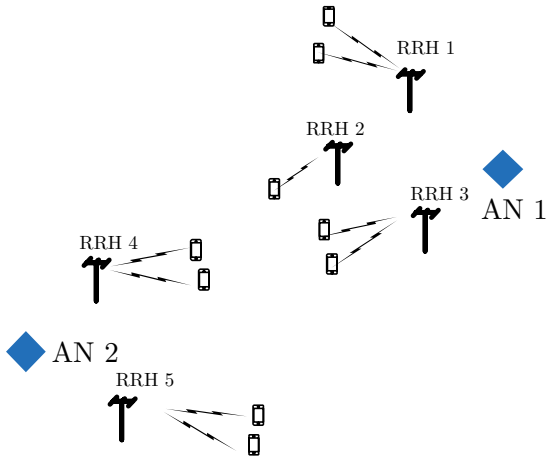

Figure 7.2: Antenna Domain Formation (A)

to aggregation nodes in Fig 7.3), what is the optimal assignment of users to antenna domains, using the total interference leakage as performance metric (Fig 7.4)?

A) was addressed in our most recent work [GRIG16], where both intra-AD and inter-AD interference were to be balanced. Intra-AD interference was inherently present due to the CB-type of precoding (Weighted-MMSE [SRLH11]) that was used within each antenna domain. Though closely related, this work will not be included in the thesis.

In contrast, investigating B) under the assumption that intra-AD interference is nulled by the proposed precoding, will be the main objective of this part in the thesis. We focus on theoretical aspects of the so-called *ADF problem*: Given an initial state (i.e., assignment of users to radio-heads, and radio-heads to aggregation nodes), we study the optimal assignment of users to antenna domains, using the total interference leakage as performance metric. In contrast to our earlier work [GRIG16], we assume tighter coordination within each antenna domain. In this chapter we describe the basic setup for the Cloud-RAN system under consideration, and outline the main assumptions. Moreover, we formulate the ADF problem as an integer optimization problem, and provide a small illustrative example to motivate the problem.

In addition to the notation defined in Chap. 1, we introduce the following. For any two vectors $\boldsymbol{x}, \boldsymbol{y}$ (resp. matrices $\boldsymbol{X}, \boldsymbol{Y}$), inequalities such as $\boldsymbol{x} \leq \boldsymbol{y}$ (resp. $\boldsymbol{X} \leq \boldsymbol{Y}$) hold element-wise. While $\mathbf{1}_n$ denotes the $n \times 1$ vector of ones, $\mathbf{0}_n$ denotes the $n \times 1$ vector of zeros, $\boldsymbol{e}_n$ is the nth elementary vector of appropriate dimension. Given a set $\mathcal{X}$, $|\mathcal{X}|$ denotes its cardinality, and $\text{conv}(\mathcal{X})$ its convex hull.

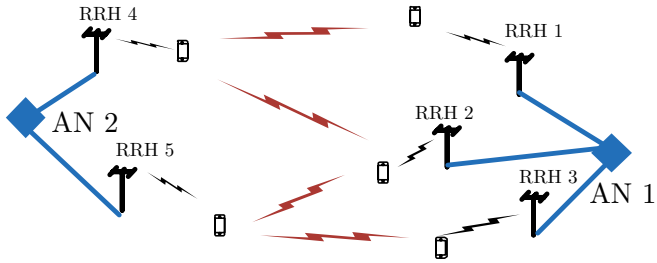

Figure 7.3: Antenna Domain Formation (B): Initial state

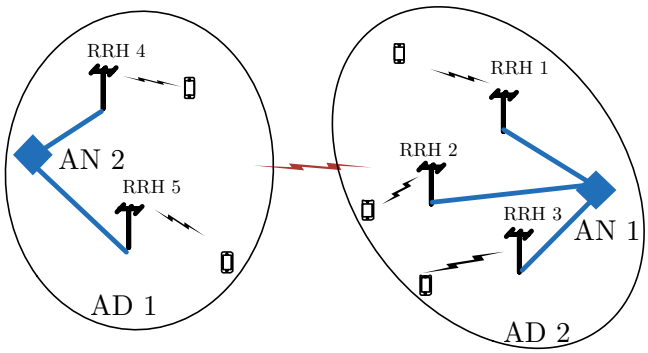

Figure 7.4: Antenna Domain Formation (B): Resulting antenna domain structure

7.2 Model and Assumptions

Consider a large area, comprising of A *aggregation nodes*, N_T *remote radio-heads*, and K_T *users*. The k*th* aggregation node, is connected to N radio-heads, via wireless/wired links, where each radio-head is serving a set of users. We refer to the collection of radio-heads connected to each aggregation node, as an *antenna domain*. Thus, from a system-level perspective, each antenna domain is serving a set of users (thereby abstracting the operation of the radio-heads in the system). A small toy example of the considered system model is illustrated in Fig. 7.5. With that in mind, each antenna domain comprises of N radio-heads and K users. Note that for simplicity of notation, we assume that N and K are the same with in each antenna domain (keeping in mind that the results of this chapter are still applicable for cases where N and K vary across antenna domains). We denote by $\mathcal{A}$ the set of aggregation nodes, and $\mathcal{I}$ the set of all users, i.e., $\mathcal{I} = \{j_n \mid 1 \leq j \leq A, 1 \leq n \leq K\}$. Since all the quantities defined above are time-varying, the proposed model is for

each scheduling time-slot (thus, any time-related indexes are omitted).

Radio-heads are assumed to have limited baseband/processing capabilities, restricted to precoding only. Moreover, aggregation nodes act as "large" centralized processors, that gather channel state information (CSI) from all the users, perform the required processing/optimization, and communicate the optimal precoders to the radio-heads. To avoid complicating the notation, we assume that each radio head is equipped with M antennas, while each user has a single antenna. This can be extended to multiple antennas at the receiver, and different antenna configuration in a straightforward manner

We outline the main assumptions used throughout this part of the thesis.

Assumption 7.2.1 (Synchronization)**.** The different radio-heads within each AD are tightly synchronized, e.g., phase-level synchronization, essentially acting as a large virtual antenna array. Each antenna domain consists of a small number of RRHs (typically $2 \sim 4$).

Assumption 7.2.2 (CSI)**.** Global and perfect CSI is assumed to be a priori available at each of the aggregation nodes.

Assumption 7.2.3 (Zero intra-AD interference)**.** The precoding within each antenna domain, is designed in such a way that no intra-antenna domain interference is present, i.e., no interference among users within the same antenna domain.

We underline at this point that such assumptions are quite common for Cloud-RAN related performance studies (e.g., [ZTCY15, TCZY15, LHMJL13]). Though the resulting systems tend to have large overhead and complexity, we reiterate that the main aim of this work is the study of performance bounds for antenna domain systems, rather than practical design paradigms. More light is shed on the overhead in Sect. 8.4.3.

Let j_n denote the index of the nth user, in the jth antenna domain, $j_n \in \mathcal{I}$. Then, its received signal is given by (assuming a *downlink transmission* scenario),

$$y_{j_n} = \sum_{q=1}^{K} \boldsymbol{h}_{j,j_n} \boldsymbol{v}_{j_q} s_{j_q} \sqrt{p}_j + \sum_{i \neq j}^{A} \sum_{m=1}^{K} \boldsymbol{h}_{i,j_n} \boldsymbol{v}_{i_m} s_{i_m} \sqrt{p}_j + n_{j_n} \tag{7.2.1}$$

where $\boldsymbol{h}_{i,j_n} \in \mathbb{C}^{1 \times MN}$ is the (MISO) channel from antenna domain i to user j_n, $\boldsymbol{v}_{i_m} \in \mathbb{C}^{MN \times 1}$ the beamforming vector to user $i_m \in \mathcal{I}$, s_{i_m} the data symbol for user $i_m \in \mathcal{I}$ such that $\mathbb{E}[s_{i_m} s_{i_m}^{\dagger}] = 1$, n_{j_n} the AWGN noise for user $j_n \in \mathcal{I}$ such that $\mathbb{E}[n_{j_n} n_{j_n}^{\dagger}] = \sigma_{j_n}^2$, and p_j the transmit power for antenna domain $j \in \mathcal{A}$. Moreover, the proposed precoding design, i.e., zero intra-AD interference (Sect. 7.2), translates to the following,

$$\boldsymbol{h}_{j,j_n} \boldsymbol{v}_{j_q} = \begin{cases} \beta_j, & \forall n = q \\ 0 \ , & \forall n \neq q \end{cases} \quad , \ \forall j \in \mathcal{A} \tag{7.2.2}$$

where $\beta_j > 0$ is a free parameter that is chosen to satisfy the maximum transmit power constraint on the transmit precoder (per antenna domain), i.e.,

$$\sum_{q=1}^{K} \|\boldsymbol{v}_{j_q}\|_2^2 \leq K, \ \forall j \in \mathcal{A} \tag{7.2.3}$$

The resulting received signal and SINR are,

$$\begin{aligned} y_{j_n} &= \beta_j s_{j_n} \sqrt{p}_j + \sum_{i \neq j}^{A} \sum_{m=1}^{K} \boldsymbol{h}_{i,j_n} \boldsymbol{v}_{i_m} s_{i_m} \sqrt{p}_i + n_{j_n} \\ \gamma_{j_n} &= \frac{\beta_j^2 p_j}{\sum_{i \neq j}^{A} \sum_{m=1}^{K} p_i |\boldsymbol{h}_{i,j_n} \boldsymbol{v}_{i_m}|^2 + \sigma_{j_n}^2} \end{aligned} \tag{7.2.4}$$

where $\text{SNR}_{j_n} = p_j \beta_j^2 / \sigma_{j_n}^2$ is the SNR of user j_n. Assuming optimal encoding/decoding, and treating interference as noise, the achievable sum-rate of the network is given by,

$$R_\Sigma = \sum_{j=1}^{A} \sum_{n=1}^{K} \log_2(1 + \gamma_{j_n}) \tag{7.2.5}$$

7.2.1 Motivation

We denote by ψ_{i_m,j_n} the so-called *interference coupling coefficient* between users i_m and j_n,

$$\psi_{i_m,j_n} = \begin{cases} p_i |\boldsymbol{h}_{i,j_n} \boldsymbol{v}_{i_m}|^2, \ \forall \ (i_m, j_n) \in \mathcal{I}^2, & i_m \neq j_n, \\ 0 \quad , \quad \forall i_m = j_n \end{cases}$$

ψ_{i_m,j_n} denotes the interference that user $i_m \in \mathcal{I}$ causes to user $j_n \in \mathcal{I}$. Moreover, we recall that $\psi_{i_m,j_n} \neq \psi_{j_n,i_m}$. Let $\boldsymbol{\Psi} \in \mathbb{R}_+^{K_T \times K_T}$ be the matrix formed by gathering all the coupling coefficients, i.e.,

$$[\boldsymbol{\Psi}]_{j_n,i_m} = \begin{cases} \psi_{j_n,i_m}, \ \forall i \neq j \\ 0, \forall i = j \end{cases} \quad , \forall \ (j_n, i_m) \in \mathcal{I}^2, \tag{7.2.6}$$

and

$$x_{k,j_n} \in \{0,1\}, \ \forall \ j_n \in \mathcal{I}, \quad k \in \mathcal{A} \tag{7.2.7}$$

be the assignment variable for user j_n to antenna domain k. With that in mind, g_{j_n}, the total interference leakage seen by user $j_n \in \mathcal{I}$, is given by,

$$g_{j_n}(\{x_{k,j_n}\}) \triangleq \sum_{k \in \mathcal{A}} \sum_{\substack{l \in \mathcal{A} \\ l \neq k}} \left(\sum_{\substack{i_m \in \mathcal{I} \\ i_m \neq j_n}} x_{k,i_m} \psi_{i_m,j_n} x_{l,j_n} \right) \tag{7.2.8}$$

where $\{x_{k,j_n}\}$ denotes the set of all assignment variables. The total interference leakage, f, is then

$$f(\{x_{k,j_n}\}) \triangleq \sum_{j_n \in \mathcal{I}} g_{j_n}(\{x_{k,i_m}\}) \tag{7.2.9}$$

and can be rewritten as,

$$f(\{x_{k,j_n}\}) \triangleq \sum_{k \in \mathcal{A}} \sum_{\substack{l \in \mathcal{A} \\ l \neq k}} \left(\sum_{j_n \in \mathcal{I}} \sum_{\substack{i_m \in \mathcal{I} \\ i_m \neq j_n}} x_{k,i_m} \psi_{i_m,j_n} x_{l,j_n} \right) \tag{7.2.10}$$

Recall that due to the proposed precoding (i.e., zero intra-AD interference), the inter-AD interference leakage coincides with the total interference leakage, f, in the system.

Example 7.2.1 (Motivating Example)**.** Consider the following toy example with $A = 2$, $K = 2$, $N = 1$ (Fig. 7.5). Then the cost in (7.2.10) reduces to,

$$f(\{x_{k,j_n}\}) = \sum_{i_m \in \mathcal{I}} \sum_{j_n \neq i_m} x_{1,i_m} \psi_{i_m,j_n} x_{2,j_n}, \ \mathcal{I} = \{1_1, \cdots, 2_2\}$$

Now the intuition behind the above cost becomes clear: the cost of having users i_m and j_n in different antenna domains is $\psi_{i_m,j_n} + \psi_{j_n,i_m}$, and zero otherwise. That same criterion is the reason that intra-AD interference is not accounted for, in f. The last equation shows that the total interference leakage in this network (Fig. 7.5) corresponds to setting all the assignment variables to one, i.e., $f(\{x_{k,j_n} = 1\})$ - a "naive" assignment. Thus, better performance can be reaped-off with a "smarter" assignment. This is the main motivation for using a cost function such as (7.2.10).

7.3 Problem Formulation

In the last part, we motivated the effect of assigning users to antenna domains. That is the so-called ADF problem, that is formalized below.

Definition 7.3.1 (Antenna Domain Formation (ADF))**.** Given an initial state (i.e., assignment of users to radio-heads, and radio-heads to aggregation nodes), the ADF problem is given by the optimal assignment of users to antenna domains, w.r.t. minimizing the total interference leakage in the system. The corresponding optimization problem is the integer programming problem shown below,

$$(P) \begin{cases} \min\limits_{\{x_{k,j_n}\}} f = \sum_{k=1}^{A} \sum_{l \neq k}^{A} (\sum_{j_n} \sum_{i_m \neq j_n} x_{k,i_m} \psi_{i_m,j_n} x_{l,j_n}) \\ \text{s. t. } \sum_{i_m} x_{k,i_m} = \rho_k, \ \forall k \in \mathcal{A} \\ \quad \sum_{k=1}^{A} x_{k,i_m} \leq 1, \ \forall i_m \in \mathcal{I} \\ \quad x_{k,i_m} \in \{0,1\}, \ \forall (k,i_m) \in \mathcal{A} \times \mathcal{I} \end{cases} \tag{7.3.1}$$

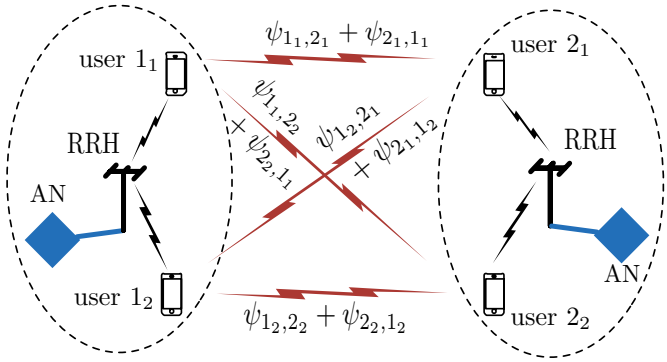

Figure 7.5: Toy Example (interference marked in red): total interference equal to inter-AD interference

The first constraint specifies that $\rho_k \in \mathbb{Z}_+$ users are to be assigned to each antenna domain, i.e., the *loading constraint*. Such a constraint is needed for the sake of load balancing on the backhaul (i.e., to prevent highly asymmetric cases where all users get assigned to one antenna domain, while the rest are idle). Moreover, the second constraint, i.e., the *assignment constraint*, ensures that each user is assigned to *at most* one antenna domain. As a result, when $\sum_{k\in\mathcal{A}} < K_T$, some users are not assigned to any antenna domain. Another way to interpret (P) is from a user assignment/selection perspective: given an initial state with K users (where K large), the goal is to select the optimal subset (of size $\rho_k < K$), that minimizes the interference leakage.

We first start by rewriting (P) in vector and matrix form - both of which will be used later in the text (keeping in mind that all are equivalent). Let $\boldsymbol{x}_k$ be the *aggregate assignment vector* for antenna domain k to all other users, and $\boldsymbol{X}$ the *aggregate assignment matrix* for the system,

$$\boldsymbol{x}_k = [x_{k,1_1}, \cdots, x_{k,A_K}]^T,\ \boldsymbol{x}_k \in \mathbb{B}^{K_T}, \forall k \in \mathcal{A}$$
$$\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_A] \in \mathbb{B}^{K_T \times A}$$

Proposition 7.3.1. *(P) can be rewritten in equivalent vector form,*

$$(P) \begin{cases} \min f(\{\boldsymbol{x}_k\}) = \sum_{k=1}^{A} \sum_{l\neq k}^{A} \boldsymbol{x}_k^T \boldsymbol{\Psi} \boldsymbol{x}_l \\ \text{s. t.} \quad \sum_{k=1}^{A} \boldsymbol{x}_k \leq \mathbf{1}_{K_T}, \\ \qquad \mathbf{1}_{K_T}^T \boldsymbol{x}_k = \rho_k,\ \boldsymbol{x}_k \in \mathbb{B}^{K_T}, \quad \forall k \in \mathcal{A} \end{cases} \tag{7.3.2}$$

and matrix form

$$(P) \begin{cases} \min f(\boldsymbol{X}) = \text{tr}(\boldsymbol{X}^T \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Omega}) \\ \text{s. t. } \boldsymbol{X}\mathbf{1}_A \leq \mathbf{1}_{K_T}, \quad \boldsymbol{X} \in \mathcal{S}_\rho \end{cases} \tag{7.3.3}$$

where we denote by $\mathcal{S}_\rho$ *the set of all* $K_T \times A$ *binary matrices, that satisfy the loading constraint, i.e.,*

$$\mathcal{S}_\rho \triangleq \{\boldsymbol{Q} \in \mathbb{B}^{K_T \times A} \mid \boldsymbol{Q}^T \mathbf{1}_{K_T} = \boldsymbol{\rho}\} \tag{7.3.4}$$

$\boldsymbol{\rho} \triangleq [\rho_1, ..., \rho_A]^T$, *and* $\boldsymbol{\Omega} \triangleq \mathbf{1}_A \mathbf{1}_A^T - \boldsymbol{I}_A$.

Proof. The derivations are shown in Appendix 8.7.1. □

It can be seen from (7.3.2) that f is not jointly convex in all the variables, due to the coupling among them. However, we underline the inherent *multi-linear* nature of f (taken separately in each variable, f is linear), that we exploit for the optimization.

Chapter 8

Proposed Approach

In the previous chapter, we motivated and formulated the ADF problem, (P) in (7.3.1), as an integer optimization problem. In this chapter, we then employ *Block-Coordinate Descent (BCD)* - that we have earlier developed in [GRIG16], to iteratively solve the problem. The lack of theoretical guarantees on the obtained solution, as well as the complicated nature of the problem, motivates us to find useful and meaningful *lower bounds* on the ADF problem (since it represents the total interference leakage). For that purpose, we derive the corresponding *Dantzig-Wolfe (DW)* decomposition (a Linear Program (LP) with exponentially many variables), and adapt the *Column Generate Method (CGM)* to compute the DW lower bound. We also derive the dual problem (a natural lower bound), characterize the duality gap, and show that the DW lower bound is tighter than of the dual problem (and consequently all related methods such as dual subgradient ascent, and Lagrange relaxation). Finally, we provide some numerical results that highlight the performance of our proposed algorithm.

8.1 Algorithm

Our proposed approach consists of two parts, where the first one concerns the optimal assignment of users to antenna domains, i.e., obtaining a solution to the so called ADF problem in (7.3.1). Parts of the approach ware presented in our earlier work [GRIG16] - albeit for a different system model. They are still summarized here for completeness. In the second part, we develop the precoding mechanism. We first present the following definition.

Definition 8.1.1 (Integrality Property for Linear Programming)**.** Consider the following binary linear program (LP),

$$(P)\ \boldsymbol{x}^\star = \min_{\boldsymbol{x}}\ \boldsymbol{c}^T\boldsymbol{x},\ \text{s. t. } \boldsymbol{x} \in \mathcal{C},\ \ \boldsymbol{x} \in \mathbb{B}^N,$$

and its *continuous relaxation* (CR) (also known as LP relaxation),

$$(CR)\ \hat{\boldsymbol{x}} = \min_{\boldsymbol{x}}\ \boldsymbol{c}^T\boldsymbol{x},\ \text{s. t. } \boldsymbol{x} \in \mathcal{C},\ \ \boldsymbol{0}_N \leq \boldsymbol{x} \leq \boldsymbol{1}_N,$$

where $\mathcal{C}$ is a convex set. The set $\mathcal{C}$ is said to satisfy the *integrality property* if all its vertexes correspond to integers: it is well-known for such cases, that the so-called *continuous relaxation* (CR) is optimal [Fra05], and consequently, $\hat{\boldsymbol{x}}$ is integer as well, and $\hat{\boldsymbol{x}} = \boldsymbol{x}$.

8.1.1 Algorithm Description

Due to the coupled nature of the objective function in (7.3.2), we leverage the well known Block-Coordinate Descent (BCD) method, that has been applied to several areas of signal processing, namely, transmitter and receiver optimization in cellular networks [SRLH11, SSB+09, GKBS15, GCJ11]. In what follows, n denotes the iteration number, i.e., $\boldsymbol{x}_k^{(n)}$ denotes the value of $\boldsymbol{x}_k$ at the nth iteration. We denote by $\boldsymbol{z}_k^{(n)} = \{\boldsymbol{x}_1^{(n+1)}, ..., \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_{k+1}^{(n)}, ..., \boldsymbol{x}_A^{(n)}\}$ the block of fixed variables, for the kth update at the nth iteration.

Our exposition here will be summarized, since the full details of the algorithm are shown in [GRIG16]. All the derivations/formulations of this part are based on (P), as shown in (7.3.2). We let $f(\boldsymbol{x}_k, \boldsymbol{z}_k^{(n)})$ denote the function $f(\boldsymbol{x}_k)$, when the variables in block $\boldsymbol{z}_k^{(n)}$ are fixed, which can be written as,

$$\begin{aligned} f(\boldsymbol{x}_k, \boldsymbol{z}_k^{(n)}) &= \boldsymbol{x}_k^T \boldsymbol{\Psi} \left(\sum_{l=1}^{k-1} \boldsymbol{x}_l^{(n+1)} + \sum_{l=k+1}^{A} \boldsymbol{x}_l^{(n)} \right), \\ &\triangleq \boldsymbol{x}_k^T \boldsymbol{r}_k^{(n)}, \end{aligned} \tag{8.1.1}$$

where $\boldsymbol{r}_k^{(n)}$ is referred to as the *residual* of antenna domain k, at the nth iteration. Looking at the above equation, $f(\boldsymbol{x}_k, \boldsymbol{z}_k^{(n)})$ is linear in $\boldsymbol{x}_k$, implying that f is linear in each block of variables. The application of BCD yields the following update for $\boldsymbol{x}_k$, at the nth iteration.

$$\boldsymbol{x}_k^{(n+1)} = \begin{cases} \underset{\boldsymbol{x}_k}{\text{argmin}}\ f(\boldsymbol{x}_k, \boldsymbol{z}_k^{(n)}) \\ \text{s. t. } \mathbf{1}_{K_T}^T \boldsymbol{x}_k = \rho_k,\ \boldsymbol{x}_k \leq \boldsymbol{\omega}_k,\ \boldsymbol{x}_k \in \mathbb{B}^{K_T}, \end{cases} \tag{8.1.2}$$

where $\boldsymbol{\omega}_k \triangleq \mathbf{1}_{K_T} - \sum_{l \neq k} \boldsymbol{x}_l$ is the set of feasible assignments for $\boldsymbol{x}_k$. The above problem belongs to the class of Mixed-Integer Linear Programs (MILPs). Moreover, it is a special case of the *generalized assignment problem (GAP)*. Though the generic formulation of GAP is known to be NP-hard, we exploit the particular structure of (8.1.2), to show that it is equivalent to a LP. Let $\mathcal{C}$ be the set formed by the first two constraints in (8.1.2), i.e., $\mathcal{C} = \{\mathbf{1}_{K_T}^T \boldsymbol{x}_k = \rho_k,\ \boldsymbol{x}_k \leq \boldsymbol{\omega}_k\}$. Recalling that ρ_k and $\boldsymbol{\omega}_k$ are integers (by construction), one can see that the vertexes of $\mathcal{C}$ are integers, and thus satisfies the integrality property (as presented in Definition 8.1.1). Thus, following the result of Definition 8.1.1, its continuous relaxation will yield

the optimal solution. Thus, the last problem is equivalent to,

$$\boldsymbol{x}_k^{(n+1)} = \begin{cases} \underset{\boldsymbol{x}_k}{\text{argmin}}\ f(\boldsymbol{x}_k, \boldsymbol{z}_k^{(n)}) \\ \text{s. t. } \mathbf{1}_{K_T}^T \boldsymbol{x}_k = \rho_k,\ \boldsymbol{x}_k \leq \boldsymbol{\omega}_k,\ \mathbf{0}_{K_T} \leq \boldsymbol{x}_k. \end{cases} \tag{8.1.3}$$

As seen from (8.1.1), when $\{\boldsymbol{x}_l\}_{l \neq k}$ are fixed, the cost function decouples in $\boldsymbol{x}_k$'s and can thus be solved *locally* at antenna domain k, in a *fully distributed manner*: Each aggregation node solves its own subproblem - a linear program, without any loss in optimality. The process is formalized in Algorithm 7. In a nutshell, the optimal update for $\boldsymbol{x}_k$ at antenna domain k, is a function of the assignments at all the other antenna domains (that thus have to be shared): Given assignments from other antenna domains, $(\boldsymbol{x}_1^{(n+1)}, ..., \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_{k+1}^{(n)}, ..., \boldsymbol{x}_A^{(n)})$, antenna domain k forms the residual $\boldsymbol{r}_k^{(n)}$, and can proceed to solve its optimization problem locally, and update $\boldsymbol{x}_k^{(n+1)}$.

Algorithm 7 ADF via BCD

Input: $\boldsymbol{\Psi},\ K_T, \boldsymbol{\rho},\ A$
for $n = 0, 1, \cdots, L-1$ **do**
  // *procedure at each aggregation node*
  obtain $(\boldsymbol{x}_1^{(n+1)}, ..., \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_{k+1}^{(n)}, ..., \boldsymbol{x}_A^{(n)})$ at antenna domain k
  compute residual $\boldsymbol{r}_k^{(n)}$ using (8.1.1)
  compute feasible assignment $\boldsymbol{\omega}_k$ using (8.1.2)
  compute $\boldsymbol{x}_k^{(n+1)}$ as solution to (8.1.3)
end for
Output: $\boldsymbol{X}^{(L)} = [\boldsymbol{x}_1^{(L)}, ..., \boldsymbol{x}_A^{(L)}]$

8.1.2 Convergence

Let $\{\boldsymbol{x}_k^{(n)}\}$ be the sequence iterates produced by the BCD in (8.1.3), and $\{\boldsymbol{x}_k^o\} \triangleq \lim_{n\to\infty}\{\boldsymbol{x}_k^{(n)}\}$. The monotonic nature of the BCD iterates was established in our earlier work [GRIG16], and is presented below for completeness.

Lemma 8.1.1 (Monotonicity)**.** *With each update* $\boldsymbol{x}_k^{(n)} \to \boldsymbol{x}_{k+1}^{(n)}$, *$f$ is non-increasing. Moreover, the sequence of function iterates* $\{f(\boldsymbol{x}_1^{(n)}, ..., \boldsymbol{x}_A^{(n)})\}_n$ *converges to a limit point* $f(\{\boldsymbol{x}_k^o\})$.

Proof. Refer to Appendix 8.7.4 □

Although the above result establishes the convergence of the proposed BCD method, it only establishes convergence to a limit. However, showing that this limit is a stationary point of f is not possible under the BCD framework, due to

the coupled nature of the assignment constraint (7.3.2). Even the strongest BCD convergence results such as [Tse01] cannot establish convergence to a stationary point.

8.1.3 Precoding

This far, we have only focused on the specifics of the ADF problem, while ignoring the precoding. The main idea behind the precoder design is to null all intra-AD interference, as shown in (7.2.2) and (7.2.3). More intuition could be gained by rewriting the signal model (7.2.1) in vector form,

$$\boldsymbol{y}_j = \boldsymbol{H}_{j,j}\boldsymbol{V}_j\boldsymbol{s}_j\sqrt{p_j} + \sum_{i\neq j}^{A}\boldsymbol{H}_{i,j}\boldsymbol{V}_i\boldsymbol{s}_i\sqrt{p}_i + \boldsymbol{n}_j, \tag{8.1.4}$$

where $\boldsymbol{y}_j$ is the vector of received signals for users served by antenna domain j. In the above,

$$\boldsymbol{H}_{i,j} \triangleq \begin{bmatrix} \boldsymbol{h}_{i,j_1} \\ \vdots \\ \boldsymbol{h}_{i,j_K} \end{bmatrix}, \boldsymbol{V}_i \triangleq [\boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_{i_K}] \text{and } \boldsymbol{s}_i \triangleq \begin{bmatrix} s_{i_1} \\ \vdots \\ s_{i_K} \end{bmatrix} \tag{8.1.5}$$

denote the channel between the antennas of antenna domain i and the users of antenna domain j, the matrix of precoding vectors for antenna domain i, and the vector of transmit symbols for users of antenna domain i, respectively. Then, zero intra-AD interference condition, i.e., (7.2.2) and the maximum transmit power constraint, i.e., (7.2.3) are equivalently written,

$$\boldsymbol{H}_{i,i}\boldsymbol{V}_i = \beta_i\boldsymbol{I}_K, \text{ and } \|\boldsymbol{V}_i\|_F^2 \leq K \ , \ \forall i \in \mathcal{A} \tag{8.1.6}$$

Note that the total interference leakage f, can be equivalently written as,

$$\begin{aligned} f &\triangleq \sum_{i=1}^{A}\sum_{j\neq i}^{A}\|\boldsymbol{H}_{i,j}\boldsymbol{V}_i\|_F^2 = \sum_{i=1}^{A}\sum_{j\neq i}^{A}\mathrm{tr}(\boldsymbol{H}_{i,j}\boldsymbol{V}_i\boldsymbol{V}_i^\dagger\boldsymbol{H}_{i,j}^\dagger) \\ &= \sum_{i=1}^{A}\mathrm{tr}[\boldsymbol{V}_i^\dagger(\sum_{j\neq i}^{A}\boldsymbol{H}_{i,j}^\dagger\boldsymbol{H}_{i,j})\boldsymbol{V}_i] \triangleq \sum_{i=1}^{A}\mathrm{tr}(\boldsymbol{V}_i^\dagger\boldsymbol{R}_i\boldsymbol{V}_i) \triangleq h(\{\boldsymbol{V}_i\}). \end{aligned} \tag{8.1.7}$$

Note that while f in (7.2.10) denotes the interference leakage expressed as a function of the assignment variables, h denotes interference leakage expressed as function of the precoders. Though the two are equal, we make that distinction for clarity. The precoder optimization problem for antenna domain i, is then formulated as follows.

$$\boldsymbol{V}_i^\star = \begin{cases} \underset{\boldsymbol{V}_i}{\text{argmin}} \ \ h(\boldsymbol{V}_i) = \mathrm{tr}(\boldsymbol{V}_i^\dagger\boldsymbol{R}_i\boldsymbol{V}_i) \\ \text{s. t. } \boldsymbol{H}_{i,i}\boldsymbol{V}_i = \beta_i\boldsymbol{I}_K \ . \end{cases} \tag{8.1.8}$$

Note that the transmit power constraint can be explicitly enforced, since it can be satisfied by changing the free parameter β_i. The solution for this problem is a special case of the next result (by setting $d_i = K$), where the solution is parametrized as a function of some d_i, and solved for the general case.

Proposition 8.1.1. *Consider the following convex problem.*

$$\boldsymbol{V}_i^\star = \begin{cases} \underset{\boldsymbol{V}_i \in \mathbb{C}^{MN \times d_i}}{\operatorname{argmin}} & \operatorname{tr}(\boldsymbol{V}_i^\dagger \boldsymbol{R}_i \boldsymbol{V}_i) \\ \text{s. t. } \boldsymbol{H}_{i,i}\boldsymbol{V}_i = \beta_i \boldsymbol{I}_{d_i}, \ d_i \in \mathbb{Z}_{++} \end{cases} \tag{8.1.9}$$

where $\boldsymbol{R}_i = \sum_{j \neq i}^{A} \boldsymbol{H}_{i,j}^\dagger \boldsymbol{H}_{i,j}$, $\boldsymbol{H}_{i,j} \in \mathbb{C}^{d_i \times MN}$, β_i *is a free parameter chosen to satisfy the transmit power constraint,* $\|\boldsymbol{V}_i\|_F^2 = d_i$. *The globally optimal solution is given by*

$$\boldsymbol{V}_i^\star = \frac{\sqrt{d_i}\ \boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \left(\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\right)^{-1}}{\|\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \left(\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\right)^{-1}\|_F}. \tag{8.1.10}$$

Moreover, for $d_i \leq MN$, *the problem is feasible almost surely.*

Proof. Refer to Appendix 8.7.2 □

8.2 Relaxations and Performance Bounds

It should be clear at this stage that problems such as (P) are quite challenging. This is further highlighted by the findings of the previous section: despite the widespread effectiveness of methods such as BCD, one is not able to show any stationarity of the obtained solution (i.e. no local optimality can be established). Moreover, it is hard to theoretically ascertain how 'close' is an obtained solution to optimality. To compensate for those shortcomings, finding meaningful lower bounds on (P) is of interest: that is particularly relevant for our case, since the cost function, f, represents an actual physical quantity. Moreover, as discussed earlier in Sec. 7.2.1, finding lower bounds on the interference leakage f, result in finding upper bounds on the sum-rate. For the problem at hand we derive the corresponding Dantzig-Wolfe (DW) decomposition, and establish that although the resulting problem is a LP, it has exponentially many variables. We thus adapt the Column Generation Method (CGM), for our particular problem. We also derive the dual problem for (P), and show that it yields a looser lower bound on (P). We thus conclude that methods that are based on the dual problem (e.g., Dual Subgradient Ascent, Lagrange Relaxation), offer worse lower bounds, than ones based on the DW decomposition (e.g., CGM).

8.2.1 Preliminaries

We here summarize some relevant concepts and definitions that will be applied extensively, later in the work.

Definition 8.2.1 (Inner Representation of Bounded Polyhedron)**.** Let $\mathcal{P}$ be a *bounded polyhedron* (the intersection of finitely many half-spaces), i.e. $\mathcal{P} = \{\boldsymbol{x} \in \mathbb{R}^n | \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}\}$. Then, every point $\boldsymbol{x} \in \mathcal{P}$ is expressed as a convex combination of its extreme points,

$$\boldsymbol{x} = \sum_{j=1}^{J} \boldsymbol{\psi}_j w_j, \ \sum_j w_j = 1, \ w_j \geq 0, \ \forall j \in \mathcal{V}, \tag{8.2.1}$$

where $\mathcal{V} = \{\boldsymbol{\psi}_j\}_{j=1}^{J}$ is the set of extreme points of $\mathcal{P}$.

Definition 8.2.2 (Special LPs)**.** Consider the following LP,

$$(LP)\ \boldsymbol{x}^\star = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}}\ \boldsymbol{c}^T\boldsymbol{x}, \ \text{s. t. } \mathbf{1}_n^T\boldsymbol{x} = 1, \ \boldsymbol{x} \geq \mathbf{0}_n.$$

Let $\mathcal{V}$ be the set of vertexes (extreme points) for (LP). Note that, $\mathcal{V}$ can be written as $\mathcal{V} = \{\boldsymbol{e}_i\}_{i=1}^{n}$, where $\boldsymbol{e}_i$ is the ith elementary vector in $\mathbb{R}^n$. Moreover, for LPs, the optimal solution lies within $\mathcal{V}$ - a fundamental result for LPs.

$$\begin{aligned}(LP): \ \boldsymbol{x}^\star &= \operatorname{argmin} \boldsymbol{c}^T\boldsymbol{x}, \ \text{s. t. } \boldsymbol{x} \in \mathcal{V} \\ \Leftrightarrow \ i^\star &= \ \operatorname{argmin}_{1\leq i\leq n} \boldsymbol{c}^T\boldsymbol{e}_i \ ,\end{aligned}$$

and consequently, $\boldsymbol{x}^\star = \boldsymbol{e}_{i^\star}$. For such problems, the solution reduces to searching over the cost $\boldsymbol{c}$. A simple toy example is shown in Fig. 8.1.

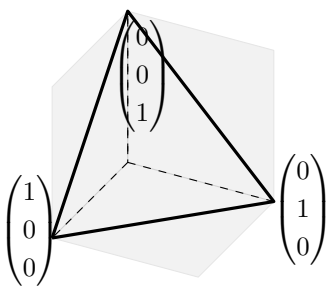

Figure 8.1: Feasible region of a special LP in $\mathbb{R}^3$ (solid lines). All vertexes are elementary vectors, i.e., binary.

In what follows, we define the following notation,

$$\begin{aligned}
&\mathcal{S}_\rho \triangleq \{\boldsymbol{Q}_j\}_{j=1}^{S},\ S = |\mathcal{S}_\rho| \\
&\alpha_j \triangleq \mathrm{tr}(\boldsymbol{Q}_j^T \boldsymbol{\Psi} \boldsymbol{Q}_j \boldsymbol{\Omega}),\ \forall j = 1, \cdots, S \\
&\boldsymbol{q}_j \triangleq \boldsymbol{Q}_j \mathbf{1}_A - \mathbf{1}_{K_T},\ \forall \boldsymbol{Q}_j \in \mathcal{S}_\rho,\ \ \boldsymbol{q}_j \in \mathbb{Z}^{K_T} \\
&\boldsymbol{\Gamma} \triangleq [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_S],\ \boldsymbol{\Gamma} \in \mathbb{Z}_+^{K_T \times S}.
\end{aligned} \tag{8.2.2}$$

Moreover, note that $\mathcal{S}_\rho$ has a *decomposable structure*, i.e. $\mathcal{S}_\rho \triangleq \mathcal{W}(\rho_1) \times \cdots \times \mathcal{W}(\rho_A)$ where $\mathcal{W}(\rho_i) = \{\boldsymbol{x} \in \mathbb{B}^{K_T} \mid \mathbf{1}_{K_T}^T \boldsymbol{x} = \rho_i\}$. Thus,

$$S \triangleq |\mathcal{S}_\rho| = \prod_{i=1}^{A} |\mathcal{W}(\rho_i)| \text{ where } |\mathcal{W}(\rho_i)| = \binom{K_T}{\rho_i}. \tag{8.2.3}$$

Remark 8.1. As it will become clear in this section, some of the decompositions/relaxations in question are computationally demanding. However, we underline the fact that such methods are intended for benchmarking purposes: they are intended to run during an offline training phase, where "enough" computational resources are available. Thus, the computation of quantities such as $\boldsymbol{\Psi}, \mathcal{S}_\rho, \alpha_j, \boldsymbol{q}_j$ and $\boldsymbol{\Gamma}$ is not a limiting factor.

8.2.2 Dantzig-Wolfe Decomposition

Initially proposed in their seminal paper [GBD60b], the Dantzig-Wolfe decomposition has been widely adopted by the operations research community, for finding bounds on integer programming problems. Based on our above definitions in (8.2.2), we can rewrite $\mathcal{S}_\rho$ and (P) as,

$$\mathcal{S}_\rho = \{\boldsymbol{X} = \sum_{j=1}^{S} w_j \boldsymbol{Q}_j \mid \sum_{j=1}^{S} w_j = 1,\ w_j \in \mathbb{B}, \forall j\}, \tag{8.2.4}$$

$$(P) \begin{cases} \min\ f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^T \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Omega}) \\ \text{s. t.}\ \ \boldsymbol{X} \in \mathcal{S}_\rho,\ \ \boldsymbol{X}\mathbf{1}_A \leq \mathbf{1}_{K_T}. \end{cases} \tag{8.2.5}$$

The above problem is still difficult to tackle, due to the combinatorial nature of $\boldsymbol{X} \in \mathcal{S}_\rho$. The DW decomposition then proceeds by relaxing $\boldsymbol{X} \in \mathcal{S}_\rho$, into a convex one, by taking its *convex hull*, i.e.,

$$\mathrm{conv}(\mathcal{S}_\rho) = \{\boldsymbol{X} = \sum_{j=1}^{S} w_j \boldsymbol{Q}_j \mid \mathbf{1}_S^T \boldsymbol{w} = 1,\ \mathbf{0}_S \leq \boldsymbol{w}\}. \tag{8.2.6}$$

As a result, every point in $\text{conv}(\mathcal{S}_\rho)$ is represented as a *convex combination* of the *extreme points* of $\text{conv}(\mathcal{S}_\rho)$ (detailed in Definition 8.2.1). Since $\mathcal{S}_\rho \subseteq \text{conv}(\mathcal{S}_\rho)$, the resulting problem (P_{DW}), is a lower bound on (P),

$$(P_{DW}) \begin{cases} \min \ f(\boldsymbol{X}) = \text{tr}(\boldsymbol{X}^T \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Omega}) \\ \text{s. t.} \ \ \boldsymbol{X} \in \text{conv}(\mathcal{S}_\rho), \ \ \boldsymbol{X}\mathbf{1}_A \leq \mathbf{1}_{K_T}. \end{cases} \tag{8.2.7}$$

Note that the assignment constraint can be written in an equivalent form,

$$\begin{aligned} \boldsymbol{X}\mathbf{1}_A \leq \mathbf{1}_{K_T} &\Leftrightarrow (\sum_j w_j \boldsymbol{Q}_j)\mathbf{1}_A \leq \mathbf{1}_{K_T} \Leftrightarrow (\sum_j w_j \boldsymbol{Q}_j \mathbf{1}_A) \leq \mathbf{1}_{K_T} \\ &\Leftrightarrow \sum_j w_j \boldsymbol{q}_j + (\sum_j w_j)\mathbf{1}_{K_T} \leq \mathbf{1}_{K_T} \Leftrightarrow \boldsymbol{\Gamma}\boldsymbol{w} \leq \mathbf{0}_{K_T}, \end{aligned}$$

where the last one follows from the fact that $\sum_j w_j = 1$ (as defined by the DW decomposition). Moreover, recalling that $\alpha_j \triangleq \text{tr}(\boldsymbol{Q}_j^T \boldsymbol{\Psi} \boldsymbol{Q}_j \boldsymbol{\Omega}), \ \forall j$, and letting $\boldsymbol{w} = (w_1, \cdots, w_S)^T$, (8.2.7) is equivalent to,

$$(P_{DW}) \begin{cases} \min\limits_{\boldsymbol{w}} \boldsymbol{\alpha}^T \boldsymbol{w} \\ \text{s. t.} \ \boldsymbol{\Gamma}\boldsymbol{w} \leq \mathbf{0}_{K_T}, \ \ \mathbf{1}_S^T \boldsymbol{w} = 1, \ \ \boldsymbol{w} \geq \mathbf{0}_S. \end{cases} \tag{8.2.8}$$

A few remarks are in order at this stage. Note that despite the combinatorial and non-convex nature of (P), the DW always results in a linear program (provided that $\mathcal{S}_\rho$ is a bounded polyhedron). However, there is the additional caveat that though (8.2.8) is a LP, it has an exponential number of variables, S: it is unfit for conventional LP solvers. We thus adapt the Column Generate Method (CGM), for our particular problem.

Remark 8.2. We note that (8.2.6) clearly shows that the DW decomposition is a mapping from $\boldsymbol{X}$ in (7.3.3), to $\boldsymbol{w}$ in (8.2.8). However, this mapping is evidently not one-to-one, since $\boldsymbol{X}$ uniquely reconstructs from $\boldsymbol{w}$, but not vice versa [LD05].

Solution via Column Generation Method

The Column Generation Method (CGM), attempts to iteratively solve (8.2.8), thereby mitigating the need for directly solving it: starting from $\boldsymbol{\Gamma}_0$ - a matrix consisting of a subset of m_o columns of $\boldsymbol{\Gamma}$, one first solves the resulting *restricted master problem* (RMP), i.e. a reduced version of (8.2.8). Then, at the lth iteration, one selects an additional column that is added to $\boldsymbol{\Gamma}_0$ (or multiple ones), and solves the resulting RMP. Given a subset $\mathcal{X}$ of $\mathcal{S}_\rho$, we define, $\boldsymbol{\Gamma}(\mathcal{X}) \in \mathbb{Z}_+^{K_T \times |\mathcal{X}|}$ as the matrix generated by the $\mathcal{X}$ columns of $\boldsymbol{\Gamma}$, and $\boldsymbol{\alpha}(\mathcal{X}) \in \mathbb{R}^{|\mathcal{X}|}$ the corresponding sub-vector of $\boldsymbol{\alpha}$.

The procedure is formalized below. Let $\mathcal{T}_o \subset \mathcal{S}_\rho$ be the initial subset of columns for $\boldsymbol{\Gamma}$, such that $|\mathcal{T}_o| = m_o$. At iteration $l \geq 1$, given the previous selected columns

$\mathcal{T}_{l-1}$, and the corresponding optimal solutions for the RMP, $\pi^\star_{l-1}$ and $\boldsymbol{\mu}^\star_{l-1}$, the vectors of *reduced costs* is defined as,

$$\boldsymbol{d}_l \triangleq \boldsymbol{\alpha}(\mathcal{Z}_{l-1}) - \hat{\boldsymbol{\Gamma}}(\mathcal{Z}_{l-1})^T \boldsymbol{\mu}^\star_{l-1} - \pi^\star_{l-1}\mathbf{1}_{|\mathcal{Z}_{l-1}|}, \tag{8.2.9}$$

where $\mathcal{Z}_{l-1} \triangleq \mathcal{S}_\rho/\mathcal{T}_{l-1}$ and $\hat{\boldsymbol{\Gamma}}(\mathcal{T}_{l-1})^T = [-\boldsymbol{\Gamma}(\mathcal{T}_{l-1})^T, \boldsymbol{I}_{K_T}]$. Then, the index of the column to be updated is defined as,

$$i^\star_l \triangleq \underset{i\in\mathcal{Z}_{l-1}}{\operatorname{argmin}} \, [\boldsymbol{d}_l]_i, \tag{8.2.10}$$

and the set of *active columns* is updated as follows,

$$\mathcal{T}_l = \mathcal{T}_{l-1} \cup \{i^\star_l\}.$$

Essentially, $i^\star_l$ is the index of the column in $\boldsymbol{\Gamma}$, that is added to the RMP. Then, the updated RMP at iteration l, is denoted by (R_l),

$$(R_l): \quad \boldsymbol{w}^\star(\mathcal{T}_l) \begin{cases} \underset{\boldsymbol{w}(\mathcal{T}_l)}{\operatorname{argmin}} \, \boldsymbol{\alpha}(\mathcal{T}_l)^T \boldsymbol{w}(\mathcal{T}_l) \\ \text{s. t. } \boldsymbol{\Gamma}(\mathcal{T}_l)\boldsymbol{w}(\mathcal{T}_l) \leq \mathbf{0}_{K_T} \\ \quad \mathbf{1}^T_{m_l}\boldsymbol{w}(\mathcal{T}_l) = 1, \;\; \boldsymbol{w}(\mathcal{T}_l) \geq \mathbf{0}_{m_l} \end{cases} \tag{8.2.11}$$

The above problem is a simple LP, and assuming that it is feasible, strong duality holds. Then, it can be verified that its equivalent dual is written as,

$$(\boldsymbol{\mu}^\star_l, \pi^\star_l) \quad \begin{cases} \underset{\boldsymbol{\mu}_l \geq \mathbf{0} \, , \, \pi_l}{\operatorname{argmax}} \; \pi_l \\ \text{s. t. } \hat{\boldsymbol{\Gamma}}(\mathcal{T}_l)^T \boldsymbol{\mu}_l + \pi_l \mathbf{1}_{m_l} \leq \boldsymbol{\alpha}(\mathcal{T}_l) \; , \end{cases} \tag{8.2.12}$$

where $m_l \triangleq |\mathcal{T}_l| = m_o + l$, and $\hat{\boldsymbol{\Gamma}}(\mathcal{T}_l)^T = [-\boldsymbol{\Gamma}(\mathcal{T}_l)^T, \boldsymbol{I}_{K_T}]$. The steps are detailed in Table 8.1. Note that, in the worst case, CGM ends up adding all columns in $\boldsymbol{\Gamma}$, i.e., solving the original problem (8.2.8). However, most often, the algorithm will terminate much earlier than that.

When all reduced costs are non-negative, the optimal solution has been found, i.e., the solution of the current RMP is the same as the original problem. Let L be that iteration number, and $\boldsymbol{w}^\star(\mathcal{T}_L)$, $(\boldsymbol{\mu}^\star_L, \pi^\star_L)$ be the corresponding optimal primal-dual pair corresponding to (R_L). Then, the optimal solution $\boldsymbol{w}^\star$ of the original problem, (8.2.8) is given by,

$$\boldsymbol{w}^\star_i = \begin{cases} \boldsymbol{w}^\star_i(\mathcal{T}_L) \text{ if } i \in \mathcal{T}_L \\ 0, \text{ otherwise} \end{cases} \quad , 1 \leq i \leq S. \tag{8.2.13}$$

Looking at (8.2.13), the solution that CGM yields consists only of the component in $\boldsymbol{w}$ that have a contribution to the solution (8.2.8), while setting the rest to zero. Interestingly, in most cases, despite the exponential size of $\boldsymbol{w}$, it will have only a few non-zero entries. It is a well-known fact that despite its iterative nature, CGM is an exact method, i.e., $\boldsymbol{w}^\star$ in (8.2.13) is the globally optimal solution to (P_{DW}).

Initialization: $\mathcal{T}_0, m_0$
for $l = 1, 2, \cdots, S - m_o$ **do**
  $\mathcal{Z}_l \leftarrow \mathcal{S}_\rho / \mathcal{T}_l$
  // *update* $\boldsymbol{\mu}_l^\star, \pi_l^\star$
  Generate $\boldsymbol{\Gamma}(\mathcal{T}_l), \hat{\boldsymbol{\Gamma}}(\mathcal{T}_l), \boldsymbol{\alpha}(\mathcal{T}_l)$
  Compute $\boldsymbol{\mu}_l^\star, \pi_l^\star$ by solving (R_l)
  // *update reduced costs and active columns*
  $\boldsymbol{d}_l \leftarrow \boldsymbol{\alpha}(\mathcal{Z}_l) - \hat{\boldsymbol{\Gamma}}(\mathcal{Z}_l)^T \boldsymbol{\mu}_l^\star - \pi_l^\star \mathbf{1}_{|\mathcal{Z}_l|}$
  $i^\star \leftarrow \underset{i \in \mathcal{Z}_l}{\operatorname{argmin}} \, [\boldsymbol{d}_l]_i$
  if $[\boldsymbol{d}_l]_{i^\star} \leq 0$
    $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{i^\star\}$
    Compute $\boldsymbol{\Gamma}(\mathcal{Z})$, $\boldsymbol{\alpha}(\mathcal{Z})$ and solve (R_l) again
  else $d_j \geq 0$
    Compute optimal solution in (8.2.13)
end for
Output: $\boldsymbol{w}^\star$

Table 8.1: DW solution via CGM

Remark 8.3. Note that the algorithm can be extended to taking $\Delta \geq 1$ columns at each iteration, that correspond to columns with negative reduced cost, thereby speeding up the algorithm. However, for simplicity of exposition, we stick with the above formulation, where one column is added at each iteration.

Bound on DW decomposition

In a last step, we shed light on the tightness of the proposed DW decomposition. Using the already established framework, we derive two simple (yet potentially loose) bounds.

Lemma 8.2.1 (Bounds on DW decomposition gap)**.** *Let $\boldsymbol{X}^\star$ and $\boldsymbol{w}^\star$ be optimal solution for the primal problem (P_3) in* (7.3.3)*, and DW problem (P_{DW}) in* (8.2.8)*, respectively. Then the following holds,*

$$0 \leq f(\boldsymbol{X}^\star) - f_{DW}(\boldsymbol{w}^\star) \leq \eta \sigma_{\max}[\boldsymbol{\Psi}] - \min_{1 \leq j \leq S} \alpha_j \leq \eta(\sigma_{\max}[\boldsymbol{\Psi}] - \sigma_{\min}[\boldsymbol{\Psi}]) \quad (8.2.14)$$

where $\eta \triangleq \sum_k \sum_{l \neq k} \rho_k \rho_l$.

Proof. Refer to Appendix 8.7.5. □

Interestingly, while the first bound is tighter, the second one is more informative: The DW bound is tighter as the largest and smallest singular values of $\boldsymbol{\Psi}$ get closer. In the limit case, the DW bound is exact, when *all* the singular values of $\boldsymbol{\Psi}$ are the same.

8.2.3 Dual Problem

In addition to being a natural lower bound on (P), the dual problem, (D), is the basis of several techniques for obtaining lower bounds. For instance, it is the "optimal bound" that the *Lagrange Relaxation* - one of most widely adopted methods for finding lower bounds, can yield. Moreover, methods such as *Dual Subgradient Ascent* - the analog of gradient ascent for non-differentiable problems, converge to the optimal solution of (D). With that in mind we derive the dual problem (D), associated with (P), characterize the resulting duality gap, and show that the DW decomposition offers a tighter bound than the dual problem (and hence all the associated methods described above).

Suboptimality of Dual Problem Bound

Proposition 8.2.1. *The dual problem, (D), is defined as,*

$$(D) \quad \max_{\boldsymbol{\lambda} \geq \mathbf{0}_{K_T}} d(\boldsymbol{\lambda}) = \{ \min_{\boldsymbol{X} \in \mathcal{S}_\rho} \operatorname{tr}(\boldsymbol{X}^T \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Omega}) + \boldsymbol{\lambda}^T (\boldsymbol{X} \mathbf{1}_A - \mathbf{1}_{K_T}) \} \ , \tag{8.2.15}$$

can be written as follows,

$$(D) \begin{cases} \max\limits_{\boldsymbol{\mu}} \ \boldsymbol{c}^T \boldsymbol{\mu} \\ \text{s. t. } \bar{\boldsymbol{\Gamma}}^T \boldsymbol{\mu} \leq \boldsymbol{\alpha}, \ \boldsymbol{\mu} \geq \mathbf{0}_{K_T}, \end{cases} \tag{8.2.16}$$

where $\boldsymbol{c} = [\mathbf{0}_N, \ 1]^T$, and $\bar{\boldsymbol{\Gamma}}^T = [-\boldsymbol{\Gamma}^T, \ \mathbf{1}_S]$.

Proof. Refer to Appendix 8.7.6 □

(D) in (8.2.16) is a LP, and since strong duality holds, we work with its (equivalent) dual form. Moreover, pluging in the values of $\bar{\boldsymbol{\Gamma}}$ and $\boldsymbol{c}$, (D) in (8.2.16) is equivalent to,

$$(D) \begin{cases} \min\limits_{\boldsymbol{w}} \ \boldsymbol{\alpha}^T \boldsymbol{w} \\ \text{s. t. } \boldsymbol{\Gamma} \boldsymbol{w} \leq \mathbf{0}_S, \ \mathbf{1}_S^T \boldsymbol{w} \geq 1, \ \boldsymbol{w} \geq \mathbf{0}_S. \end{cases} \tag{8.2.17}$$

Comparing (D) in (8.2.17) to (P_{DW}) in (8.2.8) quickly reveals that (D) is a relaxation of (P_{DW}). Consequently, the bound provided by the DW decomposition is tighter than that of the dual. Thus, methods such as Lagrange Relaxation and Dual Subgradient Ascent (that yield a solution to (D)) will result in looser bounds on (P), when compared to methods based on the (P_{DW}).

Characterization of Duality Gap

As the dual problem is the object of several investigations in this work, it is natural to inquire about the wideness of the *duality gap*: the difference between the optimal solution of (P), and that of (D). Indeed, such a gap could be large (or potentially

unbounded). We note at this point that an exact characterization of the duality gap is clearly infeasible (since one needs optimal solutions for both (P), and (D)). We thus provide a bound on the gap, in the result below.

Lemma 8.2.2 (Bound on Duality Gap)**.** *Let $\boldsymbol{X}^\star$ and $\boldsymbol{\lambda}^\star$ be optimal solution for the primal problem (P) in* (7.3.3) *and the dual (D) in* (8.2.15)*, respectively. Then the duality gap satisfies,*

$$0 \leq f(\boldsymbol{X}^\star) - d(\boldsymbol{\lambda}^\star) \leq \eta(\sigma_{\max}[\boldsymbol{\Psi}] - \sigma_{\min}[\boldsymbol{\Psi}]) + \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star - \sum_k \rho_k \min_i [\boldsymbol{\lambda}^\star]_i \,, \tag{8.2.18}$$

where $\eta \triangleq \sum_k \sum_{l \neq k} \rho_k \rho_l$.

Proof. Refer to Appendix 8.7.7 □

Discussions

In this section we investigated potential bounds on the ADF problem. Motivated by the lack of optimality claims on the BCD solution, we derived problems that correspond to lower bounds on the ADF problem: the DW decomposition, and the dual problem. After deriving the latter, we provided an upper bound on the duality gap to ensure it is bounded, and concluded that the dual problem is a relaxation of the DW problem. Consequently, the DW problem offers as good a bound as possible (or better) with respect to the dual problem. This in turn implies that methods based on the DW decomposition (e.g., CGM) yield tighter approximations than methods based on the dual problem (e.g., dual subgradient ascent, Lagrange relaxation). We derived informative bounds on the gap between the DW and the ADF problem. Focusing on the DW problem, we argued that it has exponentially many variables. We thus adapted the CGM to iteratively solve the DW problem, as it is ill-suited for conventional solvers. We shed light on the tightness of the DW decomposition in the numerical results section.

8.3 The two antenna domain case

We focus in this section on the case of two antenna domains, since the problem takes a rather simple form. Moreover, we propose an equivalent reformulation of the ADF problem, that enables a straightforward and systematic solution. Firstly, the cost function is given by $f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^T(\boldsymbol{\Psi} + \boldsymbol{\Psi}^T)\boldsymbol{x}_2$. Moreover, note that in this case, the assignment constraint is always satisfied and thus no longer needed. We assume full-load conditions with equal loading for the antenna domains (i.e., $\rho_1 = \rho_2 = K_T/2$). For this special case, $\boldsymbol{x}_2 = \mathbf{1}_{K_T} - \boldsymbol{x}_1$. Thus, the optimization problem can be expressed in terms of $\boldsymbol{x}_1$ only (and one can drop all subscripts). With

that in mind, the loading constraint is expressed as, $\mathbf{1}_{K_T}^T \boldsymbol{x} = \rho$. Letting $\bar{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \boldsymbol{\Psi}^T$, when $A = 2$, (P) takes the following simple form,

$$(P): \ f(\boldsymbol{x}^\star) = \min_{\boldsymbol{x} \in \mathcal{S}_\rho} f(x) = (\boldsymbol{x}^T \bar{\boldsymbol{\Psi}} \mathbf{1}_{K_T} - \boldsymbol{x}^T \bar{\boldsymbol{\Psi}} \boldsymbol{x}), \tag{8.3.1}$$

where $\mathcal{S}_\rho = \{\boldsymbol{x} \in \mathbb{B}^{K_T} \mid \mathbf{1}_{K_T}^T \boldsymbol{x} = \rho\}$.

8.3.1 Equivalent formulation

We use a "DW-like" transformation to reformulate problems such as (P), into an equivalent form. The result below is given for the generic case.

Lemma 8.3.1. *Let $p(\boldsymbol{Z})$ be any arbitrary (possibly non-convex) function, and consider the following integer programming problem*

$$(Q) \quad \boldsymbol{Z}^\star = \text{argmin } p(\boldsymbol{Z}) \text{ s. t. } \boldsymbol{Z} \in \mathcal{S}, \tag{8.3.2}$$

where $\mathcal{S} = \{\boldsymbol{W}_j \mid j = 1, ..., n\}$ is a finite discrete set. Then, the problem is equivalent to,

$$(Q) \quad \boldsymbol{t}^\star = \begin{cases} \text{argmin } p_d(\boldsymbol{t}) = \boldsymbol{t}^T \boldsymbol{\theta} \\ \text{s. t. } \ \boldsymbol{t}^T \mathbf{1}_n = 1, \ \boldsymbol{t} \geq \mathbf{0}_n, \end{cases} \tag{8.3.3}$$

where $[\boldsymbol{\theta}]_j \triangleq p(\boldsymbol{W}_j), \ j = 1, ..., n$.

Proof. Refer to Appendix 8.7.3. □

Lemma 8.3.1 can be directly applied to rewrite (8.3.1) in an equivalent form,

$$(P) \ \boldsymbol{w}^\star = \begin{cases} \text{argmin } \boldsymbol{w}^T \boldsymbol{\alpha} \\ \text{s. t. } \ \boldsymbol{w}^T \mathbf{1}_{K_T} = 1, \boldsymbol{w} \geq \mathbf{0}_{K_T}, \end{cases}$$

where $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_S]^T$, $\alpha_j = \boldsymbol{u}_j^T \bar{\boldsymbol{\Psi}} \mathbf{1}_{K_T} - \boldsymbol{u}_j^T \bar{\boldsymbol{\Psi}} \boldsymbol{u}_j, \ \forall j = 1, ..., S$, and $\mathcal{S}_\rho = \{\boldsymbol{u}_j\}_{j=1}^S$ Note that this last problem falls under the category of special LPs, and following the discussion in Definition 8.2.2, its solution is an elementary vector. Thus, the optimal solution to (P) is given by,

$$\boldsymbol{x}^\star = \boldsymbol{u}_{j^\star}, \text{ where } j^\star = \text{argmin}_{1 \leq j \leq S} \ \alpha_j. \tag{8.3.4}$$

Consequently, for the two antenna domain case, solving for $\boldsymbol{x}^\star$ reduces to just finding the minimum of the S-dimensional vector, $\boldsymbol{\alpha}$. Although this is similar in complexity to exhaustively searching for (P), it does provide a systematic means of doing that. Moreover, as argued in Remark 8.1, computing $\boldsymbol{\alpha}$ is not a limiting factor.

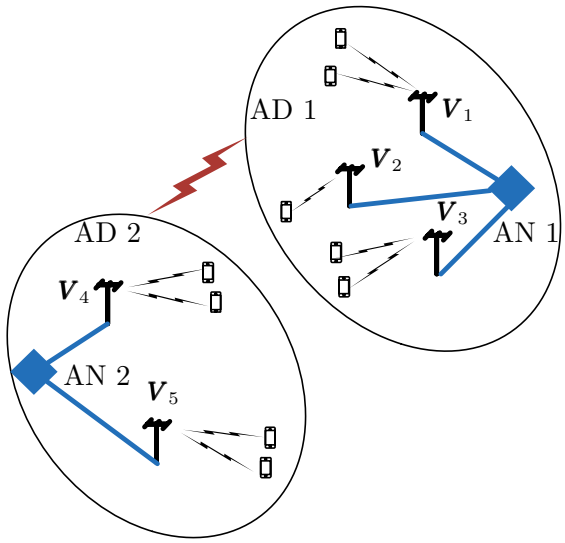

Figure 8.2: System-level Operation

8.4 Practical Aspects

8.4.1 System-Level Operation

We next detail the overall operation of the algorithm. Starting from a given deployment of aggregation nodes, radio-heads and users, each radio-head is first assigned to an aggregation node (based on some rule, e.g., minimal distance), and then synchronized within each antenna domain. Moreover, users are initially assigned to antenna domains, based on strongest channels. After the CSI acquisition phase (where each aggregation node acquires global CSI), the precoders are computed at each aggregation node, and the matrix of coupling coefficients (consisting of channels and precoders) is computed at each aggregation node. Algorithm 7 is then run across all the aggregation nodes to compute an ADF solution, that is in turn used to (re-)assign users to antenna domains. Finally, the precoders are recomputed based on the latter assignment. The overall system-wide operation of the proposed method is summarized in Algorithm 8.

8.4.2 Choice of loading factors

We highlight the existence of an interesting result, regarding the choice of loading factors: when $\sum_i \rho_i \leq MN$, then one can show that the leakage can be completely nulled.

Algorithm 8 Precoding and Antenna Domain Formation

// Start with a given users-to-antenna domain assignment
1. Compute precoders using (8.1.8)
2. Compute $\boldsymbol{\Psi}$ (based on CSI and precoders)
3. Compute ADF solution ($\boldsymbol{X}^{(L)}$ in Algorithm 7)
4. Assign users to antenna domain based on $\boldsymbol{X}^{(L)}$
5. Recompute precoders using (8.1.8)

Corollary 8.4.1. *Consider a special case of Proposition 8.1.1 by setting $d_i = \rho_i$, where $\sum_i \rho_i \leq MN$,*

$$\boldsymbol{V}_i^{\star} = \begin{cases} \underset{\boldsymbol{V}_i \in \mathbb{C}^{MN \times \rho_i}}{\operatorname{argmin}} \quad h(\boldsymbol{V}_i) = \operatorname{tr}(\boldsymbol{V}_i^{\dagger} \boldsymbol{R}_i \boldsymbol{V}_i) \\ \text{s. t. } \boldsymbol{H}_{i,i} \boldsymbol{V}_i = \beta_i \boldsymbol{I}_{\rho_i}. \end{cases} \tag{8.4.1}$$

Then, $h(\boldsymbol{V}_i^{\star}) = 0$, almost surely.

Proof. Refer to Appendix 8.7.8 □

Note that the same result of nulling all interference can be achieved by the so-called *global zero-forcing (ZF)*, wherein ZF is performed across all antenna domains thereby suppressing all interference: this turns the whole system into a noise-limited one. While global ZF would require synchronizing all radio-heads in the system, this requirement is absent in our case, and yet it still achieves the same performance. More light will be shed on this matter, in the numerical results section.

8.4.3 Communication Overhead and Complexity

In this section - included for completeness, we (roughly) estimate the cost associated with deploying the proposed scheme (Algorithm 8), in terms of *total communication overhead*. This overhead chiefly consists of ADF overhead (Algorithm 7), the CSI acquisition overhead, the data sharing overhead, and the radio-head synchronization overhead. We use the coarse measure of counting the total number of required *training symbols*, for each of the previous parts. We assume that the aggregation nodes form a fully connected network. We underline the fact that we are not advocating any specific algorithms for, say, channel estimation or radio-head synchronization. We are rather estimating the number of training symbols that one needs, using well-known methods.

At each iteration, aggregation node k updates its assignment vector, and broadcasts the updated vector to all $A-1$ other nodes. To estimate the total overhead, we assume that a given assignment vector (of size K_T) can be encoded 8-bits at a time (into a symbol), and then broadcast, thereby requiring $K_T/8$ symbols. Then

the total overhead is given by,

$$\mathcal{H}_{ADF} = AL(K_T/8) \quad \text{symbols} , \tag{8.4.2}$$

where L is the number of iterations of Algorithm 7. We assume a TDD uplink pilot-based channel estimation done in an orthogonal fashion: each of the K users sends out orthogonal pilot sequences that enables each of the antenna domains to estimate the MN channel gains. Moreover, each antenna domain has to broadcast its CSI to the other $A-1$, for a total of

$$\mathcal{H}_{CSI} = K_T N_T M \quad \text{symbols} . \tag{8.4.3}$$

The precoding implicitly assumes that radio-heads within an antenna domain act as virtual array (Sect. 7.2). Thus, the K data symbols for each antenna domain have to be broadcast to all other ones, for a total of

$$\mathcal{H}_{DS} = AK = K_T \quad \text{symbols} . \tag{8.4.4}$$

Finally, the overhead required to perform phase-level synchronization of the radio-heads within each antenna domain, was studied in our earlier work [RGIG15]. Using the latter results, we see that K training symbols are required to synchronize radio-heads within each antenna domain (if carried out in the uplink phase), thereby resulting in a total of,

$$\mathcal{H}_{SYNC} = AK = K_T \quad \text{symbols} . \tag{8.4.5}$$

Note that each of the aforementioned quantities can occur at the backhaul between aggregation nodes, the backhaul between the radio-heads, and/or over-the-air. Then, the total associated overhead is given as,

$$\Omega_{CENT} = K_T(2 + MN + AL/8) \quad \text{symbols} , \tag{8.4.6}$$

At each aggregation node, the *computational complexity* of the proposed approach (Algorithm 8) is dominated by the matrix inversion step (of size $MN \times MN$) to compute the precoder (Proposition 8.1.1), as well as solving a K_T dimensional linear program (Algorithm 7). The resulting complexity is approximated as $\mathcal{C} = \mathcal{O}(M^3N^3) + \mathcal{O}(K_T^3)$.

8.5 Numerical Results

8.5.1 Simulation Setup

Recall that A is the total number of antenna domains, N and K the number of radio-heads and users per antenna domain, respectively, and M the number of antennas at each radio-head. Aggregation nodes/radio-heads/users are dropped uniformly within the area of interest, of size $A\Delta^2$, $\Delta = 100$m. The position for aggregation

nodes/radio-heads/users are kept fixed throughout the simulation, and no mobility is considered. Then, for each simulation run, channels are generated randomly, and averaging is done over 100 different channel realizations. To emulate a realistic setting, channels between radio-heads and users are assumed to be spatially correlated Rician (Kronecker model), with pathloss and shadow fading. The parametrization is discussed at length in our earlier work [RGIG15][Sect. VII-A]. The system bandwidth is 200 MHz, and noise level is set to $\sigma_{j_n} = -91$ dBm (for all users). Moreover, we assume that the loading factors are identical, $\rho_i \triangleq \rho$ (i.e., the user load is split equally among the antenna domains). The performance metric under consideration is the sum-rate in (7.2.5), as well as the total interference leakage in the system, f.

For the assignment of radio heads to aggregation nodes, we benchmark our proposed ADF algorithm (Algorithm 8) against a simple *distance-based assignment* heuristic:

- each aggregation node picks the N-nearest radio-heads (to form an antenna domain)
- users are associated to radio-heads (and consequently antenna domains) based on strongest channels (ρ_i users are associated to antenna domain i)
- each antenna domain performs ZF to its users

Moreover, we use the following upper bound:

- *Global ZF:* whereby an equivalent system is used, with all interference set to zero, i.e., global ZF across all antenna domains (requires synchronization of all radio-heads in the system)

8.5.2 Sum-rate results

We first aim to investigate the sum-rate performance of a relatively small deployment with $A = 2, M = 4, N = 2$ radio-heads per antenna domain, and $K = 8$ users per antenna domain, while varying the loading factors ρ. Fig. 8.3 shows the resulting sum-rate, and one can clearly see an increase in the performance of both schemes, as ρ is decreased: this result is expected since interference decreases as less users are served. More importantly, we see a very significant performance gap between our proposed methods, and the benchmark, for all values of ρ. Note that sum-rate values are plotted in log scale, for clarity. Moreover the aforementioned gap is increasing with ρ, becoming massive for $\rho = 4$.

Similar trends are observed by moving on to a larger setup where $A = 4, M = 2, N = 6$ radio-heads per antenna domain, and $K = 6$ users per antenna domain, as evidenced in Fig. 8.4. However, we clearly see that in that case (Fig. 8.4), the performance gap is indeed more pronounced than the previous case (Fig. 8.3): while the performance of the benchmark increases with smaller ρ, this increase is significantly more pronounced for our algorithm. In particular, for the case where $\rho = 3$, the gap is over 20 times. As detailed earlier, this is due to the fact that by an

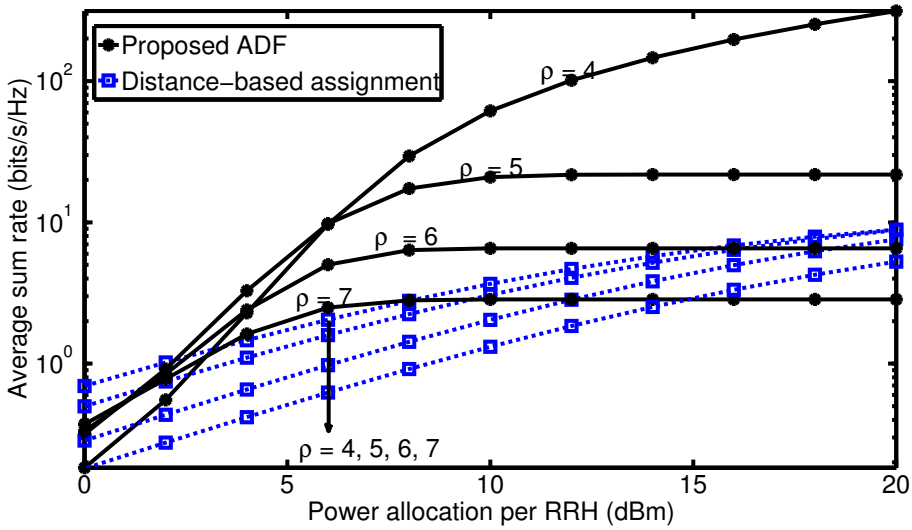

Figure 8.3: Average sum-rate performance for $A = 2, M = 4, N = 2, K = 8$

appropriate choice of ρ, the proposed scheme can totally suppress *all* interference in the system.

To shed further light on the latter effect, we investigated further deployments with $A = 2$, $N = 2$, $K = MN$, and where the loading factor is appropriately chosen as $\rho = K/2$. Fig. 8.5 shows the sum-rate for such a system, for various values of M. Most importantly, in this regime, our proposed algorithm coincides exactly with that of the global ZF upper bound. This is due to the fact that all the latter schemes are able to totally suppress all interference in the network.

8.5.3 Performance bounds

We compare in this section, the performance of the proposed BCD algorithm (Algorithm 7), against the globally optimal solution (found via exhaustive search), as well as the DW lower bound. We first look at the tightness of the DW decomposition, with respect to the globally optimal solution of (P). We consider a small scenario $(A = 2)$, assuming no fading, and looking at the (average) total interference leakage f, as metric. As seen in Table 8.2, the error form approximating the globally optimal solution of (P), by the DW lower bound (solved using CGM in Table 8.1) is quite tolerable (for $\rho = 3, 4$). We note that the case where $\rho = 1$ is too small, and not practically relevant. We also compare in Table 8.2 the performance of the proposed BCD algorithm (Algorithm 7) against that of the globally optimal solution. With that in mind, we observe a similar trend here, where the proposed BCD algorithm

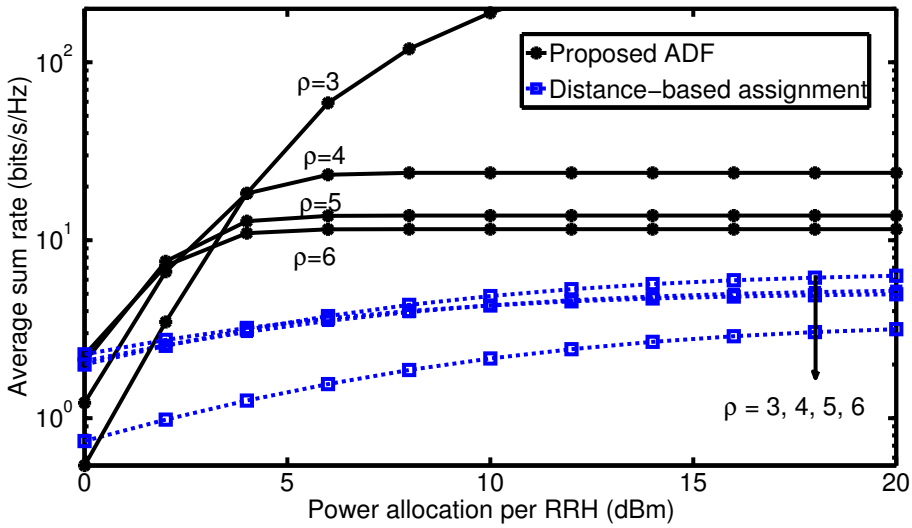

Figure 8.4: Average sum-rate performance for $A = 4, M = 2, N = 6, K = 6$

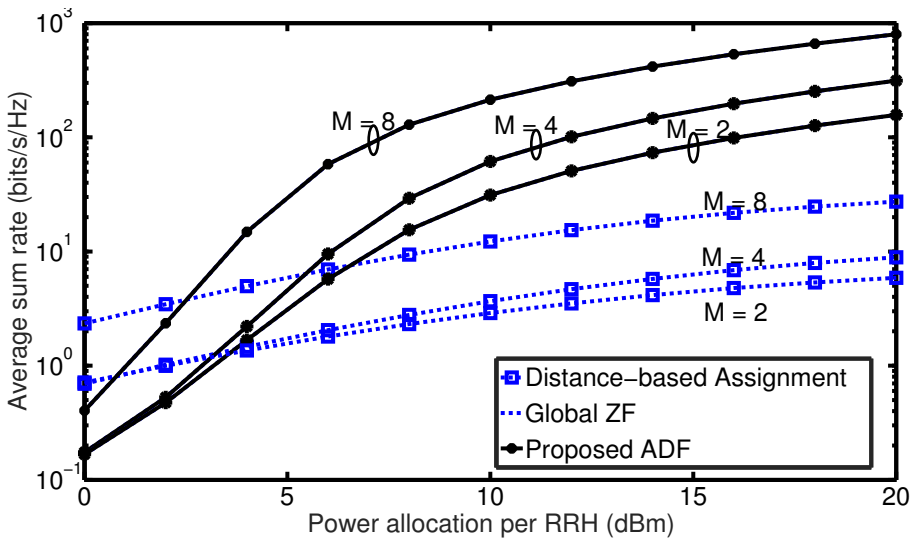

Figure 8.5: Average sum-rate performance for $A = 2, K = MN, \rho = K/2$

	$N=1$, $K=2$ $\rho=1$	$N=2$ $K=4$ $\rho=3$	$N=3$, $K=6$ $\rho=4$
Proposed	0.5329	7.5445	12.1334
Primal Opt	0.3443	6.7538	10.8226
DW decomp	0.2392	5.9249	9.3255
Error (DW) (%)	**30.53**	**13.99**	**16.05**
Error (Prop) (%)	**54.78**	**11.71**	**12.11**

Table 8.2: Average total inference leakage: proposed algorithm vs DW lower bound vs globally optimal, for $A = 2, M = 2, K = MN$

has a similar performance as the globally optimal solution, for relevant cases.

8.5.4 Discussions

A clear observation that follows from the above results (Fig. 8.5), is that massive performance gains can be achieved when the loading factor are appropriate chosen - an expected result. Though the performance of our proposed scheme is extremely close to that of global ZF (Sect. 8.4.2), it circumvents the corresponding need for synchronizing all radio-heads in the system. Not surprisingly, we observe that the performance depends on MN, the total number of transmit antennas in each antenna domain, rather than on M and N, individually. This fact could be exploited to greatly reduce the radio-head synchronization overhead, since it is independent of M and N (as shown in Sect. 8.4.3). Finally, our results also suggest that both the proposed BCD-based algorithm (Algorithm 7), and the the DW lower bound approximate well the globally optimal solution to the ADF problem, for practical cases.

8.5.5 Centralized vs Distributed Coordination

Note that this point that two approaches were presented: the centralized ADF algorithm (Algorithm 8), a family of distributed ones, namely, max-DLT (Algorithm 3). We compare their performance in terms of sum-rate and communication overhead, in the simulation setup described this section (Chap. 8.5). We underline that each antenna domain and its users, can be thought of as cell and its users. Thus, we run max-DLT in a distributed manner, across all the antenna domains. Referring to the communication overhead section for distributed algorithms (Chap 4.3.2), we set the number cells to A, the number of data streams and receive antennas to one,

and the resulting communication overhead reduces to,

$$\Omega_{DIST} = 2AKT + AK + AK = 2K_T T + 2K_T = K_T(2T+2) \quad \text{symbols} \ , \quad (8.5.1)$$

where T is the number of forward-backward iterations for max-DLT. In the above equation, while the first term is the CSI acquisition overhead, the second and third terms represent synchronization overhead (among different radio-heads) and data sharing overhead (sharing the data among all the different antenna domains). Moreover, recall that the total communication overhead for the ADF algorithm (Algorithm 8) was shown to be,

$$\Omega_{CENT} = K_T(2 + MN + AL/8) \quad \text{symbols} \ , \quad (8.5.2)$$

With that in mind, note that $\Omega_{DIST} \leq \Omega_{CENT}$ when,

$$T \leq MN/2 + AL/16 \ . \quad (8.5.3)$$

For the low-overhead algorithm that we advocate in this thesis, this is indeed the case: the overhead of a distributed solution is better than that of its centralized counterpart.

We next compare their sum-rate performance, considering a system with $A = 2, M = 4, N = 2, K = 8$. Note that for such as system, $\Omega_{DIST} = 160$ symbols, $\Omega_{CENT} = 224$ symbols. The resulting sum-rates are shown in Fig 8.6. The result depends on the SNR, and loading factor ρ. In most of the cases, its seems that max-DLT performs better than the centralized ADF. The only exception is for low-load condition $\rho = 3$, where the ADF solution slightly outperforms the distributed one, in the high-SNR region only. In addition, note that in the low-to-medium SNR region, distributed solutions (via max-DLT) are always better than the centralized one (via ADF). Thus, one can conclude at this stage that given the communication overhead models for max-DLT and the proposed ADF algorithm, distributed approaches such as max-DLT offer better performance, with a lower overhead (this conclusion is heavily dependent on the communication overhead models). Moving to a larger and denser setting (Fig. 8.7), we observe the same trends described just above, where the performance gap between max-DLT and the ADF algorithm is more pronounced. However, we recall that our overhead models (for the proposed ADF method and max-DLT), are not fully comprehensive and do not model all the required overhead. Thus, the above conclusion holds for the overhead models in question : it does not hold for any real system. Note in addition, that approaches such as max-DLT have the additional advantage of taking both the signal and interference into account, whereas the proposed ADF solely relies on the interference leakage.

8.6 Conclusion

We formulated the ADF problem as an integer optimization problem (using the interference leakage at metric), and showed that it can be tackled using BCD. Motivated by the complicated nature of the problem, we argued the need for "good"

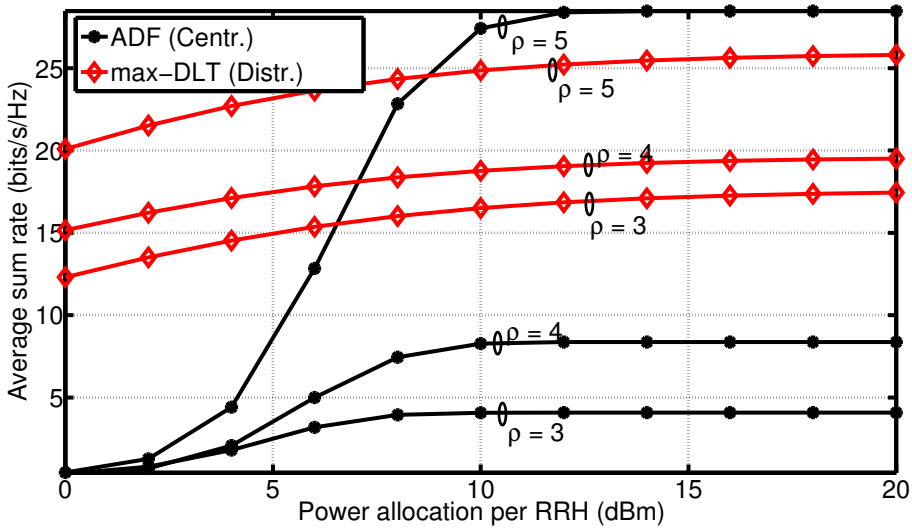

Figure 8.6: Average sum-rate performance for $A = 2, M = 4, N = 2, K = 8$ and $T = 4$

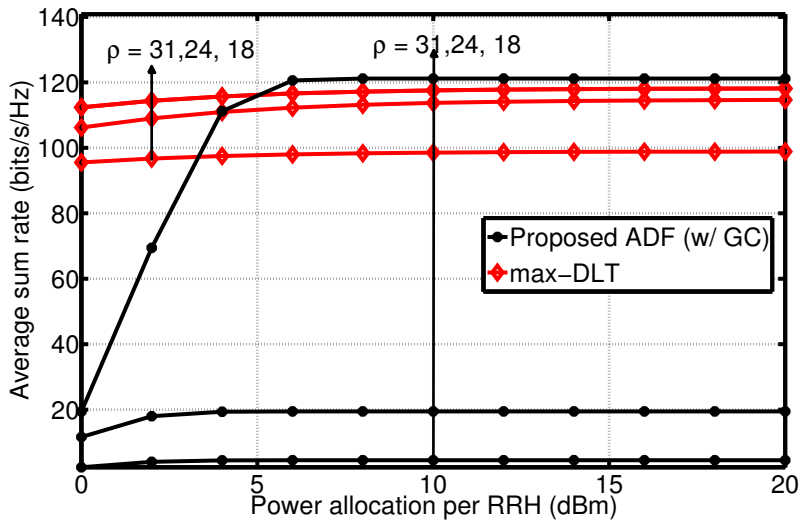

Figure 8.7: Average sum-rate performance for $A = 2, M = 4, N = 8, K = 24$ and $T = 4$

lower bounds on the problem (as well as the interference leakage). We investigated several "classical" lower bounds, such as the DW decomposition, the dual problem, and showed that the DW lower bound is tighter. Due to the exponential number of variables present in the DW lower bound, we adapted the Column Generation Method to (globally) solve it. Finally, sum-rate results clearly indicate a large performance gap between our proposed ADF algorithm, and the relevant benchmark. Moreover, in practical setups, the proposed ADF algorithm, and the advocated lower bound, seem approximate the optimal solution to the ADF problem, with acceptable error.

8.7 Appendix

8.7.1 Proof of Proposition 7.3.1

The fact that (P) can be rewritten in vector form, i.e., (7.3.2), is straightforward and can be skipped. As for rewriting (P) in matrix form (7.3.3), we first recall that for any $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$, $\mathbf{1}_m^T\boldsymbol{Q}\mathbf{1}_m = \sum_{i=1}^m\sum_{j=1}^m Q_{i,j}$, and rewrite the cost function in (P) as,

$$\begin{aligned} f &= \mathrm{tr}[\mathbf{1}_A^T(\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X})\mathbf{1}_A] - \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X}) \\ &= \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X}\mathbf{1}_A\mathbf{1}_A^T) - \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X}) = \mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{\Psi}\boldsymbol{X}\boldsymbol{\Omega}\right), \end{aligned} \tag{8.7.1}$$

where we used the fact that $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$, and let $\boldsymbol{\Omega} \triangleq \mathbf{1}_A\mathbf{1}_A^T - \boldsymbol{I}_A$. Moreover, the loading constraint can be rewritten as,

$$\sum_{i_m} x_{k,i_m} = \rho_k,\ \forall k \Leftrightarrow \mathbf{1}_{K_T}^T\boldsymbol{X} = [\rho_1, ..., \rho_A] \Leftrightarrow \boldsymbol{X}^T\mathbf{1}_N = \boldsymbol{\rho}.$$

Similarly, the assignment constraint can be reformulated as,

$$\sum_{k=1}^{A} x_{k,i_m} \leq 1, \forall i_m \in \mathcal{I} \Leftrightarrow \boldsymbol{X}\mathbf{1}_A \leq \mathbf{1}_{K_T}.$$

8.7.2 Proof of Proposition 8.1.1

Note that (8.1.8) is a convex problem (quadratic cost and linear constraint), it can be solved using standard Lagrangian techniques. The associated Lagrangian is,

$$\mathcal{L}(\boldsymbol{V}_i, \boldsymbol{M}_i) = \mathrm{tr}(\boldsymbol{V}_i^\dagger\boldsymbol{R}_i\boldsymbol{V}_i) + \mathrm{tr}\left[\boldsymbol{M}_i(\boldsymbol{H}_{i,i}\boldsymbol{V}_i - \beta_i\boldsymbol{I}_{d_i})\right],$$

where $\boldsymbol{M}_i \in \mathbb{C}^{d_i\times d_i}$ is the matrix of Lagrange multipliers. Differentiating the latter w.r.t. $\boldsymbol{V}_i$ and setting to zero yields,

$$\nabla_{\boldsymbol{V}_i}\mathcal{L} = \mathbf{0} \Leftrightarrow \boldsymbol{V}_i^\star = -\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\boldsymbol{M}_i^\star,$$

where $\boldsymbol{M}_i^\star$ is chosen to satisfy the linear constraint, i.e.,

$$\boldsymbol{H}_{i,i}(-\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \boldsymbol{M}_i^\star) = \beta_i \boldsymbol{I}_{d_i} \Leftrightarrow \boldsymbol{M}_i^\star = -\beta_i (\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger)^{-1}.$$

Combining the last two equations yields the optimal solution,

$$\boldsymbol{V}_i^\star = \beta_i \ \boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \left(\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\right)^{-1}.$$

The resulting transmit power is

$$\|\boldsymbol{V}_i^\star\|_F^2 = \beta_i^2 \|\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \left(\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\right)^{-1} \|_F^2.$$

Thus transmit power constraint is satisfied with equality for

$$\beta_i = \sqrt{d_i}/\|\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger \left(\boldsymbol{H}_{i,i}\boldsymbol{R}_i^{-1}\boldsymbol{H}_{i,i}^\dagger\right)^{-1} \|_F.$$

Note that when $d_i \leq MN$, then there always exists at least one $\boldsymbol{V}_i$, such that $\boldsymbol{H}_{i,i}\boldsymbol{V}_i = \beta_i \boldsymbol{I}_{d_i}$: the Moore-Penrose inverse of $\boldsymbol{H}_{i,i}/\beta_i$. Due to the generic nature of the channels, $\boldsymbol{H}_{i,i}/\beta_i$ is full-rank almost surely, its Moore-Penrose inverse exists almost surely, and the problem is feasible almost surely.

8.7.3 Proof of Lemma 8.3.1

The proof follows from considering the following "DW-like" mapping,

$$\begin{aligned}\mathcal{S} &= \{\boldsymbol{Z} = \sum_j t_j \boldsymbol{W}_j \mid \sum_j t_j = 1,\ t_j \in \mathbb{B}\ , \forall j = 1, ..., n\} \\ &= \{\boldsymbol{Z} = \sum_j t_j \boldsymbol{W}_j \mid \boldsymbol{t}^T \mathbf{1}_n = 1,\ \boldsymbol{t} \in \mathbb{B}^n\}. \ (g.1)\end{aligned}$$

Then, the cost in (Q) is written as $p(\boldsymbol{Z}) = \sum_j t_j p(\boldsymbol{W}_j)$. Letting $\boldsymbol{t} = [t_1, ..., t_n]^T$, and $\theta_j = p(\boldsymbol{W}_j)$, (Q) is equivalent to,

$$(Q) \begin{cases} \text{argmin } p_d(\boldsymbol{t}) = \boldsymbol{t}^T \boldsymbol{\theta} \\ \text{s. t. } \ \boldsymbol{t}^T \mathbf{1}_n = 1,\ \boldsymbol{t} \in \mathbb{B}^n. \end{cases} \tag{8.7.2}$$

It can be verified that the mapping in $(g.1)$ is *one-to-one* from Z to $\boldsymbol{t}$: every $\boldsymbol{t}$ yields a unique $\boldsymbol{Z}$, and every $\boldsymbol{Z}$ decomposes into a unique $\boldsymbol{t}$. The equivalence between the two problems follows from that.

8.7.4 Proof of Lemma 8.1.1

Note that the following is a direct consequence of (8.1.3)

$$\begin{aligned} f(\{\boldsymbol{x}_k^{(n)}\}) &\geq f(\boldsymbol{x}_1^{(n+1)}, \boldsymbol{z}_1^{(n)}) \geq f(\boldsymbol{x}_2^{(n+1)}, \boldsymbol{z}_2^{(n)}) ... \\ &\geq f(\boldsymbol{x}_A^{(n+1)}, \boldsymbol{z}_A^{(n)}) \triangleq f(\{\boldsymbol{x}_k^{(n+1)}\}), \end{aligned}$$

where the last equality follows from the fact that $f(\boldsymbol{x}_A^{(n+1)}, \boldsymbol{z}_A^{(n)})$ corresponds to the case where all variables $(\boldsymbol{x}_1, ..., \boldsymbol{x}_A)$, are updated. It follows that the sequence $\{f(\boldsymbol{x}_1^{(n)}, ..., \boldsymbol{x}_A^{(n)})\}_n$ converges to a limit point f_0

8.7.5 Proof of Lemma 8.2.1

Let $\eta = \sum_k \sum_{l \neq k} \rho_k \rho_l$. The left inequality follows immediately from the fact that the DW decomposition is always a lower bound on the problem - by construction (Sec 8.2.2). Moreover, the right one is obtained from upper bounding $f(\boldsymbol{X}^\star)$ and lower bounding $f_{DW}(\boldsymbol{w}^\star)$,

$$f(\boldsymbol{X}^\star) = \sum_k \sum_{l \neq k} {\boldsymbol{x}_k^\star}^T \boldsymbol{\Psi} \boldsymbol{x}_l^\star \leq \sum_k \sum_{l \neq k} \sigma_{\max}[\boldsymbol{\Psi}] \|\boldsymbol{x}_k^\star\|_2 \|\boldsymbol{x}_l^\star\|_2$$
$$\overset{(e.1)}{=} \sigma_{\max}[\boldsymbol{\Psi}] \sum_k \sum_{l \neq k} \rho_k \rho_l = \sigma_{\max}[\boldsymbol{\Psi}] \eta$$

where $(e.1)$ follows from the fact that $\boldsymbol{x}_k^\star$ must be feasible: thus, $\|\boldsymbol{x}_k^\star\|_2$ is the sum of all non-zero elements, and equal to ρ_k. Moreover, a simple/naive lower bound can be obtained on P_{DW} in (8.2.8), by relaxing the first constraint,

$$f_{DW}(\boldsymbol{w}^\star) \geq \min_{\substack{\mathbf{1}_S^T \boldsymbol{w} = 1, \\ \boldsymbol{w} \geq \mathbf{0}_S}} \boldsymbol{\alpha}^T \boldsymbol{w} \overset{(e.1)}{=} \min_{1 \leq j \leq S} \alpha_j = \min_j \ \mathrm{tr}(\boldsymbol{Q}_j^T \boldsymbol{\Psi} \boldsymbol{Q}_j \boldsymbol{\Omega}) \overset{(e.2)}{\geq} \eta \sigma_{\min}[\boldsymbol{\Psi}]$$

where $(e.1)$ follows from the fact that problem in a special LP, whose solution is obtained in Definition 8.2.2. Moreover, $(e.2)$ follows similar reasoning used for lower bounding $d(\boldsymbol{\lambda}^\star)$ in Appendix 8.7.7. The first and second bound follows from combining $(e.1)$ and $(e.2)$ respectively.

8.7.6 Proof of Proposition 8.2.1

We rewrite (8.2.15) in a series of equivalent problems,

$$(D) \max_{\boldsymbol{\lambda} \geq \mathbf{0}_{K_T}} d(\boldsymbol{\lambda}) = \{ \min_{\boldsymbol{Q}_j \in \mathcal{S}_\rho} \mathrm{tr}(\boldsymbol{Q}_j^T \boldsymbol{\Psi} \boldsymbol{Q}_j \boldsymbol{\Omega}) + \boldsymbol{\lambda}^T (\boldsymbol{Q}_j \mathbf{1}_A - \mathbf{1}_{K_T}) \}$$
$$(D) \max_{\boldsymbol{\lambda} \geq \mathbf{0}_{K_T}} d(\boldsymbol{\lambda}) = \{ \min_{1 \leq j \leq S} \alpha_j + \boldsymbol{\lambda}^T \boldsymbol{q}_j \}$$

$$(D) \begin{cases} \max\limits_{\boldsymbol{\lambda} \geq \mathbf{0}_{K_T},\ \zeta} \ \zeta \\ \text{s. t. } \ \alpha_j + \boldsymbol{\lambda}^T \boldsymbol{q}_j \geq \zeta, \ \forall j = 1, \cdots, S \end{cases}$$
$$(D) \begin{cases} \max\limits_{\boldsymbol{\lambda}, \zeta} \ \zeta \\ \text{s. t. } \ \boldsymbol{\alpha} + \boldsymbol{\Gamma}^T \boldsymbol{\lambda} \geq \zeta \mathbf{1}_S, \ \boldsymbol{\lambda} \geq \mathbf{0}_{K_T} \end{cases}$$

The result in (8.2.16) follows by letting $\boldsymbol{\mu} = [\boldsymbol{\lambda}, \ \zeta]^T$, $\boldsymbol{c} = [\mathbf{0}_N, \ 1]^T$, and $\bar{\boldsymbol{\Gamma}}^T = [-\boldsymbol{\Gamma}^T, \ \mathbf{1}_S]$.

8.7.7 Proof of Lemma 8.2.2

Let $\eta = \sum_k \sum_{l \neq k} \rho_k \rho_l$. The left inequality, stating that the dual solution is always a lower bound on the primal one, follows immediately from weak duality. Moreover, the right one is obtained from upper bounding $f(\boldsymbol{X}^\star)$ and lower bounding $d(\boldsymbol{\lambda}^\star)$,

$$\begin{aligned} f(\boldsymbol{X}^\star) &= \sum_k \sum_{l \neq k} \boldsymbol{x}_k^{\star^T} \boldsymbol{\Psi} \boldsymbol{x}_l^\star \leq \sum_k \sum_{l \neq k} \sigma_{\max}[\boldsymbol{\Psi}] \|\boldsymbol{x}_k^\star\|_2 \|\boldsymbol{x}_l^\star\|_2, \\ &\overset{(e.1)}{=} \sigma_{\max}[\boldsymbol{\Psi}] \sum_k \sum_{l \neq k} \rho_k \rho_l = \sigma_{\max}[\boldsymbol{\Psi}] \eta, \end{aligned}$$

where $(e.1)$ follows from the fact that $\boldsymbol{x}_k^\star$ must be feasible: thus, $\|\boldsymbol{x}_k^\star\|_2$ is the sum of all non-zero elements, and equal to ρ_k. Using (8.2.15), we formulate the optimal dual solution (and its lower bound) as,

$$\begin{aligned} & d(\boldsymbol{\lambda}^\star) \\ & \triangleq \min_{\boldsymbol{X} \in \mathcal{S}_\rho} \operatorname{tr}(\boldsymbol{X}^T \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Omega}) + \boldsymbol{\lambda}^{\star^T} (\boldsymbol{X} \mathbf{1}_A - \mathbf{1}_{K_T}) \\ & = \min_{\substack{\boldsymbol{x}_k \in \mathbb{B}^{K_T}, \forall k \\ \boldsymbol{x}_k^T \mathbf{1}_{K_T} = \rho_k, \forall k}} \sum_k \boldsymbol{x}_k^T \left(\sum_{l \neq k} \boldsymbol{\Psi} \boldsymbol{x}_l + \boldsymbol{\lambda}^\star \right) - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star \\ & \geq \min_{\substack{\boldsymbol{x}_k \in \mathbb{B}^{K_T}, \forall k \\ \boldsymbol{x}_k^T \mathbf{1}_{K_T} = \rho_k, \forall k}} \sum_k \left(\sum_{l \neq k} (\sigma_{\min}[\boldsymbol{\Psi}] \|\boldsymbol{x}_k\|_2 \|\boldsymbol{x}_l\|_2) + \boldsymbol{x}_k^T \boldsymbol{\lambda}^\star \right) - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star \\ & \overset{(e.2)}{=} \sigma_{\min}[\boldsymbol{\Psi}] \eta - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star + \sum_k \min_{\substack{\boldsymbol{x}_k \in \mathbb{B}^{K_T} \\ \boldsymbol{x}_k^T \mathbf{1}_{K_T} = \rho_k}} \boldsymbol{x}_k^T \boldsymbol{\lambda}^\star \\ & \overset{(e.3)}{=} \sigma_{\min}[\boldsymbol{\Psi}] \eta - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star + \sum_k \min_{\substack{\boldsymbol{x}_k \geq \mathbf{0} \\ \boldsymbol{x}_k^T \mathbf{1}_{K_T} = \rho_k}} \boldsymbol{x}_k^T \boldsymbol{\lambda}^\star \\ & \overset{(e.4)}{=} \sigma_{\min}[\boldsymbol{\Psi}] \eta - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star + \sum_k \rho_k \left(\min_{\substack{\boldsymbol{z}_k \geq \mathbf{0} \\ \boldsymbol{z}_k^T \mathbf{1}_{K_T} = 1}} \boldsymbol{z}_k^T \boldsymbol{\lambda}^\star \right) \\ & \overset{(e.5)}{=} \sigma_{\min}[\boldsymbol{\Psi}] \eta - \mathbf{1}_{K_T}^T \boldsymbol{\lambda}^\star + \sum_k \rho_k \min_i [\boldsymbol{\lambda}^\star]_i. \end{aligned}$$

Note that $(e.2)$ follows from the fact that $\|\boldsymbol{x}_k\|_2 = \rho_k$ for any feasible $\boldsymbol{x}_k$. $(e.3)$ is due to the fact that the problem is a MILP. Furthermore, we show that it satisfied the integrality property (as per Definition 8.1.1): then, relaxing the binary constraint into a continuous one, yields the optimal solution. Finally, $(e.4)$ is obtained by letting $\boldsymbol{z}_k = \boldsymbol{x}_k / \rho_k$, and $(e.5)$ from the fact that the problem is a Special LP whose solution is detailed in Definition 8.2.2. The final result follows by combining the above result with $(e.1)$.

8.7.8 Proof of Corollary 8.4.1

Let $\boldsymbol{Z}_i \triangleq \bigcup_{j \neq i} \text{span}(\boldsymbol{H}_{i,j})$, and $\boldsymbol{N}_i = \text{null}(\boldsymbol{Z}_i)$. Due to the generic random nature of the channels, then one can verify that $\dim(\boldsymbol{N}_i) = MN - \rho_i$, almost surely. In the case where $\sum_i \rho_i \leq MN$, then

$$\begin{aligned}
\dim(\boldsymbol{N}_i) \geq \rho_i &\Leftrightarrow \exists\, \boldsymbol{T}_i \in \mathbb{C}^{MN \times \rho_i}, \boldsymbol{T}_i \in \boldsymbol{N}_i \\
&\Leftrightarrow \exists\, \boldsymbol{T}_i \in \mathbb{C}^{MN \times \rho_i}, \boldsymbol{T}_i^\dagger \boldsymbol{H}_{i,j} = \boldsymbol{0}_{\rho_i \times \rho_i},\ \forall j \neq i, \\
&\Leftrightarrow \exists\, \boldsymbol{T}_i \in \mathbb{C}^{MN \times \rho_i}, \boldsymbol{T}_i^\dagger \boldsymbol{R}_i \boldsymbol{T}_i = \boldsymbol{0}_{\rho_i \times \rho_i}
\end{aligned}$$

Note that $h(\boldsymbol{V}_i) \geq 0$, since it is a quadratic form. Moreover, since the problem is convex and has a unique optimal solution, any solution that makes h zero, is globally optimal then. Then, consider solutions of the form, $\boldsymbol{V}_i = \boldsymbol{T}_i \boldsymbol{\Theta}_i$, where $\boldsymbol{\Theta}_i$ is an arbitrary unitary matrix. Then,

$$h(\boldsymbol{V}_i) = \text{tr}(\boldsymbol{\Theta}_i^\dagger \boldsymbol{T}_i^\dagger \boldsymbol{R}_i \boldsymbol{T}_i \boldsymbol{\Theta}_i) = \text{tr}(\boldsymbol{T}_i^\dagger \boldsymbol{R}_i \boldsymbol{T}_i) = 0$$

Chapter 9

Conclusions and Future Work

In this thesis, we have investigated optimization techniques for the most promising communication systems. Our contributions have addressed all three pillars for increasing data rates in cellular systems, namely, (A) exploiting the *massive spectrum of mmWave MIMO systems*, (B) *increasing the spectral efficiency* via BS coordination, and (C) *densification*.

Under (A), we motivated the *hybrid analog-digital MIMO* architecture as a key to scaling up the number of transmit/receive antennas for mmWave MIMO systems. After characterizing the optimal precoder/combiner structure, we proposed an algorithm (based on the Arnoldi Iteration) to blindly estimate the dominant subspaces of the mmWave MIMO channel. This is motivated by the fact the such channels are inherently *sparse* (in terms of eigenmodes): it is much more efficient to estimate the non-zero eigenmodes, rather than the entire channel. In addition, we also devised an iterative procedure to optimize the analog/digital precoder and combiner (based on estimates of the dominant subspace). Simulation results showed that the proposed approach significantly outperforms the only benchmark. We believe that such an approach - *subspace estimation exploiting channel reciprocity* (or later ones that build upon it), will be an essential component in the operation of mmWave MIMO systems. The above approach assumes narrow-band channels: this is hard to motivate, since mmWave MIMO systems have a large bandwidth, and are thus inherently frequency selective. In the future, we consider extending the proposed subspace estimation approach, to handle wide-band frequency selective channels. Moreover, we also envision to reduce the communication overhead resulting from the Arnoldi iteration.

Under (B), we investigated algorithms for distributed multi-user multi-cell coordination, under the framework of F-B iterations. Despite the plethora of different approaches developed over the years, we highlighted the fact that they all required a large enough number of iterations that would destroy the gains brought about by their use: we thus proposed two types of *low-overhead algorithms* that require just a few F-B iterations. In the first algorithm, max-DLT, we lower bounded the sum-rate using a so-called DLT bound, and derived the corresponding solution, dubbed as *non-homogeneous waterfilling*: we highlighted its ability to *turn off streams* with

low-SINR, thereby greatly speeding up the convergence of the algorithm. In the second type of algorithm, we used the interference leakage as metric. The increased convergence speed is due firstly due to relaxing the leakage minimization problem, and the introduction of a *turbo iteration* at the transmitters/receivers, within each F-B iteration, where the leakage is further decreased.

In (C), we focused on the opposing paradigm of *fully centralized coordination*, given by the Cloud-RAN architecture - the most prominent candidate for densification. In the case of multiple antenna domains, we highlighted the absence of prior work tackling both *intra-AD and inter-AD interference*, and formulated the so-called *antenna domain formation* problem, using the interference leakage as metric. We proposed an iterative algorithm, based on Block Coordinate Descent, to solve it. Motivated by the lack of theoretical guarantees on the optimality of such a solution, we derived lower bounds on the problem (and the interference leakage consequently), and compared them analytically. We also compared the performance of the proposed (centralized) ADF algorithm to (distributed) max-DLT (in the same Cloud-RAN simulation setup), and concluded that the low-overhead fast-converging max-DLT outperformed the centralized ADF algorithm. We also noticed that this performance gap increases as the deployment gets denser. We argued that this result is valid for the particular communication overhead model used here. We thus conclude at this point, that for the algorithms and communication scenarios considered in this thesis, distributed coordination algorithms (with focus of fast-convergence), are a clear winner in the case of densification. In the future, we plan to extend the numerical setup considered here, to include basic mobility of the users. We wish to investigate the robustness of the proposed ADF approach to changes in user positions, resulting from mobility: How is the sum-rate affected, if the ADF algorithms is run at higher intervals? In addition, metrics including fairness will also be considered.

Bibliography

[3GP11] Spatial channel model for multiple input multiple output (MIMO) simulations. *3GPP TR 25.996 V10.0*, Mar 2011.

[AEALH14] A. Alkhateeb, O. El Ayach, G. Leus, and R.W. Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, Oct 2014.

[AEB06] M. Aharon, M. Elad, and A Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov 2006.

[AHAS+12] O.E. Ayach, R.W. Heath, S. Abu-Surra, S. Rajagopal, and Zhouyue Pi. Low complexity precoding for large millimeter wave MIMO systems. In *Communications (ICC), 2012 IEEE International Conference on*, pages 3724–3729, June 2012.

[BB11] D.S. Baum and H. Bolcskei. Information-theoretic analysis of MIMO channel sounding. *Information Theory, IEEE Transactions on*, 57(11):7555–7577, Nov 2011.

[BB15a] R. Brandt and M. Bengtsson. Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems. *IEEE Transactions on Vehicular Technology*, PP(99):1–1, 2015.

[BB15b] R. Brandt and M. Bengtsson. Fast-convergent distributed coordinated precoding for TDD multicell MIMO systems. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, pages 457–460, Dec 2015.

[BE06] Amir Beck and Yonina Eldar. Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM Journal on Optimization*, 17(3):844–860, 2006.

[Ben02] M. Bengtsson. A pragmatic approach to multi-user spatial multiplexing. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*, pages 130–134, Aug 2002.

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[BJBO11] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten. Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell ofdma transmission. *IEEE Transactions on Signal Processing*, 59(12):6086–6101, Dec 2011.

[BJN+96] Cynthia Barnhart, Ellis L. Johnson, George L. Nemhauser, Martin W. P. Savelsbergh, and Pamela H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46:316–329, 1996.

[Bra83] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings Communications, Radar and Signal Processing*, 130(1):11–16, 1983.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[CJ08] V.R. Cadambe and S.A. Jafar. Interference alignment and degrees of freedom of the K -user interference channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, 2008.

[CJ09] V. R Cadambe and S. A Jafar. Reflections on interference alignment and the degrees of freedom of the k-user mimo interference channel. *IEEE Information Theory Society Newsletter*, 54(4):5–8, December 2009.

[CKGB16] W. M. Chan, T. Kim, H. Ghauch, and M. Bengtsson. Subspace estimation and hybrid precoding for wideband millimeter-wave mimo systems. *IEEE ASILOMAR*, Nov. 2016.

[CTRF02] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi. Joint transmitter receiver diversity for efficient space division multiaccess. *IEEE Transactions on Wireless Communications*, 1(1):16–27, Jan 2002.

[DCG04] T. Dahl, N. Christophersen, and D. Gesbert. Blind MIMO eigenmode transmission based on the algebraic power method. *IEEE Transactions on Signal Processing*, 52(9):2424–2431, Sept 2004.

[DHL+11] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela. Is the phy layer dead? *IEEE Communications Magazine*, 49(4):159–165, April 2011.

[DPCG07] T. Dahl, S.S. Pereira, N. Christophersen, and D. Gesbert. Intrinsic subspace convergence in TDD MIMO communication. *IEEE Transactions on Signal Processing*, 55(6):2676–2687, June 2007.

[DY16] Binbin Dai and Wei Yu. Energy efficiency of downlink transmission strategies for cloud radio access networks. *CoRR*, abs/1601.01070, 2016.

[EALH11] O. El Ayach, A. Lozano, and R.W. Heath. Optimizing training and feedback for MIMO interference alignment. In *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1717–1721, 2011.

[EALH12] O. El Ayach, A. Lozano, and R.W. Heath. On the overhead of interference alignment: Training, feedback, and cooperation. *IEEE Transactions on Wireless Communications*, 11(11):4192–4203, 2012.

[EARAS+14] O. El Ayach, S. Rajagopal, S. Abu-Surra, Zhouyue Pi, and R.W. Heath. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Transactions on Wireless Communications,*, 13(3):1499–1513, March 2014.

[Eri15] On the pulse of the networked society. *Ericsson Mobility Report*, November 2015.

[Fra05] Antonio Frangioni. About lagrangian methods in integer optimization. *Annals of Operations Research*, 139(1):163–193, 2005.

[GBD60a] Philip Wolfe George B. Dantzig. Decomposition principle for linear programs. *Operations Research*, 8(1):101–111, 1960.

[GBD60b] Philip Wolfe George B. Dantzig. Decomposition principle for linear programs. *Operations Research*, 8(1):101–111, 1960.

[GBKS15] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund. Subspace estimation and decomposition for hybrid analog-digital millimetre-wave mimo systems. In *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 395–399, June 2015.

[GCJ11] K. Gomadam, V. R Cadambe, and S. A Jafar. A distributed numerical approach to interference alignment and applications to wireless interference networks. *IEEE Transactions on Information Theory*, 57(6):3309–3322, June 2011.

[GHH+10] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and Wei Yu. Multi-cell MIMO cooperative networks: A new look at interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, 2010.

[GKBS13] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund. Interference alignment via controlled perturbations. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 3996–4001, Dec 2013.

[GKBS15] H. Ghauch, Taejoon Kim, M. Bengtsson, and M. Skoglund. Distributed low-overhead schemes for multi-stream MIMO interference channels. *IEEE Transactions on Signal Processing*, 63(7):1737–1749, April 2015.

[GKBS16a] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund. Subspace estimation and decomposition for large millimeter-wave mimo systems. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):528–542, April 2016.

[GKBS16b] H. Ghauch, Taejoon Kim, M. Bengtsson, and M. Skoglund. Separability and sum-rate maximization in MIMO interfering networks. *IEEE Transactions on Signal and Information Processing over Networks, Submitted , available at http://arxiv.org/pdf/1606.08589.pdf*, 2016.

[GMBS15] H. Ghauch, R. Mochaourab, M. Bengtsson, and M. Skoglund. Distributed precoding and user selection in mimo interfering networks. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, pages 461–464, Dec 2015.

[GP11] H.G. Ghauch and C.B. Papadias. Interference alignment: A one-sided approach. In *2011 IEEE Global Communications Conference (GLOBECOM 2011)*, pages 1 –5, December 2011.

[GRI+16] H. Ghauch, M. M. U. Rahman, S. Imtiaz, J. Gross, M. Skoglund, and C. Qvarfordt. Performance bounds for antenna domain systems. *IEEE Transactions on Signal Processing, Submitted , available at http://arxiv.org/pdf/1606.08401.pdf*, 2016.

[GRIG16] H. Ghauch, M. Rahman, S. Imtiaz, and J. Gross. Coordination and antenna domain formation in Cloud-RAN systems. *IEEE International Communications Conference ICC*, 2016.

[GVL96] Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[HDMF10] T. Hrycak, S. Das, G. Matz, and H.G. Feichtinger. Low complexity equalization for doubly selective channels modeled by a basis expansion. *IEEE Transactions on Signal Processing*, 58(11):5706–5719, Nov 2010.

[HKG+14] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang. Millimeter wave mimo channel tracking systems. In *2014 IEEE Globecom Workshops (GC Workshops)*, pages 416–421, Dec 2014.

[HKL+13] Sooyoung Hur, Taejoon Kim, D.J. Love, J.V. Krogmeier, T.A. Thomas, and A. Ghosh. Millimeter wave beamforming for wireless backhaul and access in small cell networks. *Communications, IEEE Transactions on*, 61(10):4391–4403, October 2013.

[HRLP16] M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang. A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing. *IEEE Signal Processing Magazine*, 33(1):57–77, Jan 2016.

[HXRL13] Mingyi Hong, Zi Xu, M. Razaviyayn, and Zhi-Quan Luo. Joint user grouping and linear virtual beamforming: Complexity, algorithms and approximation bounds. *IEEE Journal on Selected Areas in Communications,*, 31(10):2013–2027, October 2013.

[KFVY06] M.K. Karakayali, G.J. Foschini, R. Valenzuela, and R.D. Yates. On the maximum common rate achievable in a coordinated network. In *Communications, 2006. ICC '06. IEEE International Conference on*, volume 9, pages 4333–4338, June 2006.

[KG11] S. J. Kim and G. B. Giannakis. Optimal resource allocation for MIMO ad hoc cognitive radio networks. *IEEE Transactions on Information Theory*, 57(5):3117–3131, May 2011.

[KTJ12] J. Kaleva, A. TÃ¶lli, and M. Juntti. Weighted sum rate maximization for interfering broadcast channel via successive convex approximation. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 3838–3843, Dec 2012.

[KTJ13] P. Komulainen, A. Tölli, and M. Juntti. Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems. *IEEE Transactions on Signal Processing*, 61(9):2204–2218, May 2013.

[LB15] Thomas Lipp and Stephen Boyd. Variations and extension of the convex–concave procedure. *Optimization and Engineering*, pages 1–25, 2015.

[LD05] Marco Lubbecke and Jacques Desrosiers. Selected topics in column generation. *Oper. Res.*, 53(6):1007–1023, 2005.

[LH03] D.J. Love and R.W. Heath. Equal gain transmission in multiple-input multiple-output wireless systems. *Communications, IEEE Transactions on*, 51(7):1102–1110, July 2003.

[LHA13] A. Lozano, R.W. Heath, and J.G. Andrews. Fundamental limits of cooperation. *IEEE Transactions on Information Theory*, 59(9):5213–5226, Sept 2013.

[LHMJL13] Namyoon Lee, R.W. Heath, D. Morales-Jimenez, and A. Lozano. Base station cooperation with dynamic clustering in super-dense cloud-RAN. In *Globecom Workshops (GC Wkshps), 2013 IEEE*, pages 784–788, Dec 2013.

[LMS+10] Zhi-Quan Luo, Wing-Kin Ma, A.M.-C. So, Yinyu Ye, and Shuzhong Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine,*, 27(3):20–34, May 2010.

[LSC+12] Daewon Lee, Hanbyul Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana. Coordinated multipoint transmission and reception in lte-advanced: deployment scenarios and operational challenges. *IEEE Communications Magazine*, 50(2):148–155, February 2012.

[LY73] David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*. Springer Publishing Company, Incorporated, 1973.

[MAMK08] M.A. Maddah-Ali, A.S. Motahari, and A.K. Khandani. Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, 54(8):3457 –3470, August 2008.

[MBGB15] R. Mochaourab, R. Brandt, H. Ghauch, and M. Bengtsson. Overhead-aware distributed csi selection in the mimo interference channel. In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 1038–1042, Aug 2015.

[MET14] METIS D6.2. Initial report on horizontal topics, first results and 5G system concept. March 2014.

[MET15] METIS D3.3. Final performance results and consolidated view on the most promising multi-node/multi-antenna transmission technologies. Feb 2015.

[MNM11] P. Mohapatra, K. E. Nissar, and C.R. Murthy. Interference alignment algorithms for the k user constant MIMO interference channel. *IEEE Transactions on Signal Processing*, 59(11):5499–5508, Nov 2011.

[MRRGPH15] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, and R.W. Heath. Dictionary-free hybrid precoders and combiners for mmwave mimo

systems. In *Signal Processing Advances in Wireless Communications (SPAWC), 2015 IEEE 16th International Workshop on*, pages 151–155, June 2015.

[NBH10] J. Nsenga, A. Bourdoux, and F. Horlin. Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–6, May 2010.

[NLN14] D. H. N. Nguyen and T. Le-Ngoc. Sum-rate maximization in the multicell MIMO multiple-access channel with interference coordination. *IEEE Transactions on Wireless Communications*, 13(1):36–48, January 2014.

[NSGS10] F. Negro, S.P. Shenoy, I. Ghauri, and D.T.M. Slock. On the MIMO interference channel. In *Information Theory and Applications Workshop (ITA), 2010*, pages 1–9, Jan 2010.

[OBB+14] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren. Scenarios for 5g mobile and wireless communications: the vision of the metis project. *IEEE Communications Magazine*, 52(5):26–35, May 2014.

[OMI+03] K. Ohata, K. Maruhashi, M. Ito, S. Kishimoto, K. Ikuina, T. Hashiguchi, K. Ikeda, and N. Takahashi. 1.25 Gbps wireless Gigabit ethernet link at 60 GHz-band. In *Radio Frequency Integrated Circuits (RFIC) Symposium, 2003 IEEE*, pages 509–512, June 2003.

[OMM+00] K. Ohata, K. Maruhashi, Jun-ichi Matsuda, M. Ito, W. Domon, and S. Yamazaki. A 500Mbps 60GHz-band transceiver for IEEE 1394 wireless home networks. In *Microwave Conference, 2000. 30th European*, pages 1–4, Oct 2000.

[pet09] Interference alignment via alternating minimization. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, pages 2445–2448. IEEE, April 2009.

[PH11] S. W Peters and R. W Heath. Cooperative algorithms for MIMO interference channels. *IEEE Transactions on Vehicular Technology*, 60(1):206–218, January 2011.

[PH12] S.W. Peters and R.W. Heath. User partitioning for less overhead in MIMO interference channels. *IEEE Transactions on Wireless Communications*, 11(2):592 –603, February 2012.

[Pri03] R.E. Prieto. A general solution to the maximization of the multidimensional generalized rayleigh quotient used in linear discriminant analysis for signal classification. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, volume 6, pages VI–157–60 vol.6, April 2003.

[RGIG15] M. M. U. Rahman, H. Ghauch, S. Imtiaz, and J. Gross. RRH clustering and transmit precoding for interference-limited 5G CRAN downlink. In *IEEE Globecom Workshops (GC Wkshps)*, pages 1–7, Dec 2015.

[RHL12] Meisam Razaviyayn, Mingyi Hong, and Zhi-Quan Luo. A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization. *SIAM Journal on Optimization*, 23(11), sep 2012.

[RLL11] M. Razaviyayn, G. Lyubeznik, and Zhi-Quan Luo. On the degrees of freedom achievable through interference alignment in a MIMO interference channel. In *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*, pages 511–515, June 2011.

[RLL12] M. Razaviyayn, G. Lyubeznik, and Z. Q. Luo. On the degrees of freedom achievable through interference alignment in a mimo interference channel. *IEEE Transactions on Signal Processing*, 60(2):812–821, Feb 2012.

[RSM+13] T.S. Rappaport, Shu Sun, R. Mayzus, Hang Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, and F. Gutierrez. Millimeter wave mobile communications for 5G cellular: It will work! *Access, IEEE*, 1:335–349, 2013.

[Saa11] Yousef Saad. Numerical Methods for Large Eigenvalue Problems. *Manchester University Press*, (Second Edition):1–337, 2011.

[SDS+05] Jari Salo, Giovanni Del Galdo, Jussi Salmi, Pekka Kyösti, Marko Milojevic, Daniela Laselva, and Christian Schneider. MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996). On-line, January 2005. http://www.tkk.fi/Units/Radio/scm/.

[SGHP10a] I. Santamaria, O. Gonzalez, R. W Heath, and S. W Peters. Maximum sum-rate interference alignment algorithms for MIMO channels. In *2010 IEEE Global Communications Conference (GLOBECOM 2010)*, pages 1–6. IEEE, December 2010.

[SGHP10b] I. Santamaria, O. Gonzalez, R. W Heath, and S. W Peters. Maximum sum-rate interference alignment algorithms for MIMO channels. In *2010 IEEE Global Communications Conference (GLOBECOM 2010)*, pages 1–6. IEEE, December 2010.

[Sor96] Danny C. Sorensen. Implicitly restarted arnoldi/lanczos methods for large scale eigenvalue calculations. Technical report, 1996.

[SRL14] M. Sanjabi, M. Razaviyayn, and Zhi-Quan Luo. Optimal joint base station assignment and beamforming for heterogeneous networks. *IEEE Transactions on Signal Processing*, 62(8):1950–1961, April 2014.

[SRLH11] Qingjiang Shi, Meisam Razaviyayn, Zhi-Quan Luo, and Chen He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, 2011.

[SSB+09] D.A. Schmidt, Changxin Shi, R.A. Berry, M.L. Honig, and W. Utschick. Minimum mean squared error interference alignment. In *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pages 1106 –1110, November 2009.

[SSB+13] D.A Schmidt, Changxin Shi, R.A Berry, M.L. Honig, and W. Utschick. Comparison of distributed beamforming algorithms for MIMO interference networks. *IEEE Transactions on Signal Processing*, 61(13):3476–3489, July 2013.

[SY15a] F. Sohrabi and W. Yu. Hybrid digital and analog beamforming design for large-scale mimo systems. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2929–2933, April 2015.

[SY15b] F. Sohrabi and Wei Yu. Hybrid digital and analog beamforming design for large-scale mimo systems. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 2929–2933, April 2015.

[SY16] F. Sohrabi and W. Yu. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):501–513, April 2016.

[SZ95] K. Schittkowski and C. Zillober. *Stochastic Programming: Numerical Techniques and Engineering Applications*, chapter Sequential Convex Programming Methods, pages 123–141. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.

[SZ01] S. Shamai and B.M. Zaidel. Enhancing the cellular downlink capacity via co-processing at the transmitting end. In *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, volume 3, pages 1745–1749 vol.3, 2001.

[TCZY15] Meixia Tao, Erkai Chen, Hao Zhou, and Wei Yu. Content-centric sparse multicast beamforming for cache-enabled cloud RAN. *CoRR*, abs/1512.06938, 2015.

[TO02] M. Torlak and O. Ozdemir. A Krylov subspace approach to blind channel estimation for CDMA systems. In *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, volume 1, pages 674–678 vol.1, Nov 2002.

[TPA11] Y.M. Tsang, A.S.Y. Poon, and S. Addepalli. Coding the beams: Improving beamforming training in mmwave communication system. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6, Dec 2011.

[Tse01] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

[TV04] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communications*. 2004.

[VvdV10] V. Venkateswaran and A.-J. van der Veen. Analog beamforming in MIMO communications with phase shift networks and online channel estimation. *Signal Processing, IEEE Transactions on*, 58(8):4131–4143, Aug 2010.

[Wat07] David S. Watkins. *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1 edition, 2007.

[WLP+09] Junyi Wang, Zhou Lan, Chang-Woo Pyo, T. Baykas, Chin-Sean Sum, M.A. Rahman, Jing Gao, R. Funada, F. Kojima, H. Harada, and S. Kato. Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems. *IEEE Journal on Selected Areas in Communications*, 27(8):1390–1399, October 2009.

[WTW08] L.P. Withers, R.M. Taylor, and D.M. Warme. Echo-MIMO: A two-way channel training method for matched cooperative beamforming. *IEEE Transactions on Signal Processing*, 56(9):4419–4432, Sept 2008.

[XY13] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.

[YCL14] S. You, L. Chen, and Y. E. Liu. Convex-concave procedure for weighted sum-rate maximization in a mimo interference network. In *2014 IEEE Global Communications Conference*, pages 4060–4065, Dec 2014.

[YGJK10] C. M Yetis, Tiangao Gou, S. A Jafar, and A. H Kayran. On feasibility of interference alignment in MIMO interference networks. *IEEE Transactions on Signal Processing*, 58(9):4771–4782, September 2010.

[YRBC04] Wei Yu, Wonjong Rhee, S. Boyd, and J.M. Cioffi. Iterative waterfilling for gaussian vector multiple-access channels. *IEEE Transactions on Information Theory*, 50(1):145 – 152, January 2004.

[ZCA+09] Jun Zhang, Runhua Chen, J.G. Andrews, A. Ghosh, and R.W. Heath. Networked mimo with clustered linear precoding. *IEEE Transactions on Wireless Communications*, 8(4):1910–1921, April 2009.

[ZMK05] Xinying Zhang, A.F. Molisch, and Sun-Yuan Kung. Variable-phase-shift-based rf-baseband codesign for mimo antenna selection. *IEEE Transactions on Signal Processing*, 53(11):4091–4103, Nov 2005.

[ZQL13] J. Zhao, T. Q. S. Quek, and Z. Lei. Coordinated multipoint transmission with limited backhaul data transfer. *IEEE Transactions on Wireless Communications*, 12(6):2762–2775, June 2013.

[ZTCY15] H. Zhou, M. Tao, E. Chen, and W. Yu. Content-centric multicast beamforming in cache-enabled cloud radio access networks. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2015.

[ZXLS07] Xiayu Zheng, Yao Xie, Jian Li, and Petre Stoica. Mimo transmit beamforming under uniform elemental power constraint. *IEEE Transactions on Signal Processing*, 55(11):5395–5406, Nov 2007.