Efficient Group-Sparse Transceiver Design for Multiuser MIMO Relaying in C-RAN

Ayoub Saab

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

October 2018

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

Abstract

This thesis addresses the design of multiuser multiple-input multiple-output (MIMO) amplify-and-forward (AF) relaying within a cloud radio access network (C-RAN) from an energy-efficient perspective. The aim is to jointly select remote radio heads (RRH) and optimize their transceiver, each represented by an AF matrix, in order to assist the communication between multiple source-destination pairs. We formulate the design problem as an *interference leakage* minimization subject to per-relay power constraints, while imposing a set of linear constraints to preserve the desired signals at the destinations. To obtain an energy-efficient relaying solution, the objective function is penalized with a regularization term which promotes *group-sparsity* among the resultant relaying weights. A low-complexity iterative algorithm based on the alternating direction method of multipliers (ADMM) is then proposed to solve the regularized problem, which yields closed-form solutions at each iteration. The closed-form solution leads to a thresholding operation that enables the selection of a relay subset, thus yielding solutions for the set of AF matrices that are *exactly* group-sparse. Simulation results demonstrate the explicit benefits of the proposed algorithm, which results in notably lower power consumption and computational complexity than conventional relaying design methods.

Sommaire

Cette thèse porte sur la conception de relais type amplification-et-transfert (AF) à plusieurs entrées et plusieurs sorties (MIMO) dans un réseau d'accès radio "cloud" (C-RAN), du point de vue de l'efficacité énergétique. L'objectif est de sélectionner conjointement des têtes-radio-distantes (RRH) et d'optimiser leur émetteur-récepteur, chacune représentée par une matrice AF, afin de faciliter la communication entre plusieurs paires source-destination. Nous formulons le problème de conception sous la forme d'une minimisation de fuite d'interférences soumise à des contraintes de puissance par relais, tout en imposant un ensemble de contraintes linéaires pour préserver les signaux souhaités chez les destinations. Pour obtenir une solution de relais économe en énergie, la fonction d'objectif est pénalisée par un terme de régularisation qui favorise la parcimonie de groupe parmi les poids de relais résultants. Un algorithme itératif de faible complexité basé sur la méthode des multiplicateurs de direction alternée (ADMM) est alors proposé pour résoudre le problème régularisé, ce qui donne des solutions explicites à chaque itération. Cela conduit à une opération de seuillage qui permet de sélectionner un sous-ensemble de relais, et par suite à un ensemble de matrices AF qui obéissent strictement à la parcimonie de groupe. Les résultats de simulation démontrent les avantages explicites de l'algorithme proposé, ce qui se traduit par une consommation d'énergie et une complexité de calcul nettement inférieures à celles des méthodes de conception de relais classiques.

Acknowledgments

My journey as a graduate student has proven to be a true rollercoaster, filled with ups and downs and lessons to learn from the many teachers I met on and off campus.

I would like to begin by acknowledging my thesis supervisors Prof. Ioannis Psaromiligkos and Prof. Benoît Champagne for their continuous moral and financial support throughout my studies. This thesis was accomplished due to their insights and guidance. Beyond research, I thoroughly enjoyed the discussions I've had with Prof. Ioannis as he unwittingly taught me lessons about kindness and compassion. I sincerely appreciate Prof. Champagne's support and his great patience throughout the rough times I've faced during my last year at McGill. His advice allowed me to move forward, and I am proud to be one of his students.

In addition, I would like to acknowledge the partial funding I received for this work from the Natural Sciences and Engineering Research Council (NSERC) and InterDigital Canada through the Collaborative Research and Development Grant. The technical comments from Afshin Haghighat (Technical staff, InterDigital) have been greatly appreciated. Special thanks go to my friend Nadim Badra (MEng. McGill) for his help in the literature review and to Jiaxin Yang (PhD McGill) for his technical inputs. This work would have never been possible without their assistance.

Many thanks to my friends in Montréal: Nadim, Leo, David, Moe, Ardavan, Jad, Alain, Fouad, Amro, Ayman, Charbel, Emile, Lynn, Rola, Christopher, André-Paul, my football teammates... My salutations go to my friends and family in Lebanon and in the States. I thank my angel Dana for her love and my childhood friend Rody for reminding me about God. Last but not least, my deepest love goes to my parents Ward and Hassane, whose sacrifices I could never repay. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

List of Acronyms

1G	First generation
3G	Third generation
3GPP	Third Generation Partnership Project
4G	Fourth generation
4-QAM	4-point quadrature amplitude modulation
5G	Fifth generation
ADMM	Alternating direction method of multipliers
AF	Amplify-and-forward
AWGN	Additive white Gaussian noise
BBU	Baseband unit
BPSK	Base phase-shift keying
BS	Base station
CF	Compress-and-forward
CSI	Channel state information
CSIR	Channel state information at the receiver
CSIT	Channel state information at the transmitter
C-RAN	Cloud-radio access network
D2D	Device-to-device

DCF	Decompress-and-forward
DF	Decode-and-forward
DL	Downlink
EGC	Equal gain combining
KKT	Karush-Kuhn-Tucker
LTE	Long Term Evolution
MIMO	Multiple-input multiple-output
MRC	Maximum ratio combining
QoS	Quality-of-service
RAN	Radio access network
RRH	Remote radio head
SINR	Signal-to-interference-plus-noise ratio
SIR	Signal-to-inferference ratio
SNR	Signal-to-noise ratio
UE	User equipment
UL	Uplink
V2V	Vehicle-to-vehicle

Chapter 1

Introduction

In this chapter, we expose some important challenges facing the development of future mobile wireless communication systems and discuss the emergence of relay-assisted networks and small cell networks as potential solutions to several important problems arising in this context. We then review the pertinent literature on system optimization under these setups and summarize the main contributions of the thesis.

1.1 Network densification: addressing challenges in future mobile networks

The performance of wireless networks has dramatically improved in the past 30 years, as transmission rates have risen by a thousand-fold from the first generation (1G) of wireless networks to the fourth generation (4G) [1]. Despite this feat, current networks are facing a great challenge as they will not be able to supply the projected data volumes [2]. In fact, the proliferation of wireless devices with advanced capabilities has resulted in an unprecedented increase in both cell density and, more importantly, throughput per user.

Likewise, the emergence of new communication paradigms, such as vehicle-to-vehicle (V2V) and device-to-device (D2D), poses significant challenges on the design and eventual implementation of the next, fifth generation (5G) networks. In addition, to reduce end-to-end latency, power consumption, processing complexity and cost, 5G is required to enable novel (and potentially conflicting) applications, making disruptive technologies and architectures indispensable for its genesis [3].

In general, wireless networks must be designed to combat scarcity of radio resources and channel impairments, as these two factors are the main drawbacks of such networks. For a given radio resource, channel coding/modulation is one way to increase throughput, thereby achieving higher *spectral efficiency*. However, the gain achieved by these techniques is upper-bounded by the theoretical link capacity, so new ways of increasing throughput are desirable. Multiple-input multiple-output (MIMO) architectures are one solution: deploying multiple antennas at the transmitter and/or receiver nodes to exploit the *spatial* dimension enhances a network's capacity by alleviating the detrimental channel effects, such as multipath fading [4]. MIMO improves the reliability of a communication link through spatial diversity and is considered a major breakthrough in radio transceiver design.

On the network level, small cell networks complementing traditional radio access networks (RAN) have been recognized as a potential solution to increase spectral efficiency and energy efficiency [5]. Specifically, centralized cloud-processing coupled with a dense deployment of low-complexity access points has been the discussion of many works. In this new architecture, termed as cloud-RAN (C-RAN), the traditional base station (BS) functionalities are apportioned between the centralized data center and the access points. The former handles baseband signal processing functions and is thus termed the baseband unit (BBU) pool. The latter are responsible for data transmission/reception and analog-to-digital conversion and are termed remote radio heads (RRH) due to their distributed

nature. The BBU and RRHs typically exchange channel state information (CSI) and user equipment (UE) traffic data via low-latency and high-bandwidth optical transport links.

C-RAN offers various benefits as discussed in many recent papers. For instance, [6] argues for its economic gains versus current architectures by noting that macro cell BS deployments carry significant operational and capital expenditures. Another example is [7], wherein the authors discuss the benefits of cloud-based centralization (e.g., flexibility, resource pooling, load management, and interference mitigation) while addressing the associated signal processing challenges imposed by the 3GPP Long Term Evolution (LTE) standard. For the above reasons, and in view that multi-cell joint signal processing is considered a key technique for interference coordination in future wireless networks [8], C-RAN has gained considerable attention. Nonetheless, it raises issues from a network energy perspective since the power consumption associated with the deployment of a large number of RRHs can become excessive when considering both local power consumption at the RRHs (power amplifier, signal processing, etc.) and the optical transport link consumption [9]. An equally important aspect to be considered in C-RAN is the limited capacity for the optical links between the BBU and RRHs, as the performance gains of centralized processing come at the cost of significant signaling overhead [10].

In addition to MIMO and small cells, wireless relaying is considered as a cost-effective technique for capacity enhancement and coverage extension in current cellular standards, e.g., 3GPP LTE and LTE-Advanced [11]. Among the relaying transmission techniques, amplify-and-forward (AF) relaying – also known as non-regenerative relaying – is particularly promising. In addition to its low complexity, it can attain significant gains over direct transmission when coupled with MIMO along with adequate CSI knowledge [12]. Wireless relaying under the umbrella of a centralized network (such as C-RAN) will be the focus of this thesis, where we aim to improve the signal quality at the destination UEs using the

AF-relaying strategy. Moreover, we will consider energy efficiency by tackling the problem of relay-selection and optimizing the transmission under relay-power constraints.

1.2 Literature Survey

The optimization of the C-RAN archictecture has been investigated under different performance metrics, with the network power minimization being a prominent approach [13, 14, 15]. In the case of a downlink (DL) C-RAN multicast scenario, the authors in [13] develop a greedy RRH selection algorithm under worst-case signal-to-interference-plus-noise (SINR) constraints for each UE and design beamformers for the active RRH set. To limit computational cost, two efficient algorithms are derived based on a convex relaxation of the original formulation. For the purpose of group selection, both algorithms assign priorities to each RRH via a heuristic metric that utilizes prior information about the system. Under the same network setup, the effects of imperfect CSI are subsequently taken into account in [14]. In [15], where a densely deployed network of RRHs under C-RAN control is used to cooperatively support multiple UEs, a unified framework is proposed for the joint DL and uplink (UL) UE-RRH association and beamforming design based on energy minimization. In particular, a virtual DL problem is constructed for the UL, making the problem formulation amenable to group-sparse optimization as in [13, 14] and to relaxed-integer programming [16].

In contrast to the above works, [17, 18, 19, 20, 21] address the problem of UE-RRH association and beamforming design from the perspective of finite channel capacity. In [17], the problem of minimizing user data transmission required for DL coordinated multipoint is considered, subject to SINR and per-BS power constraints. To solve the resultant combinatorial optimization problem, two algorithms are proposed, although the capacity

constraints are only handled implicitly. The works in [18, 19, 20, 21] explicity incorporate the capacity constraints in the optimization, where [18] adopts a utility maximization approach in a DL C-RAN with user-centric clustering and [19] opts for maximizing the minimum SINR for all UEs subject to per-RRH fronthaul capacity constraints. While these two works ignore the case when the optimization is infeasible, [20, 21] take it into consideration by studying the problem of joint UE admission control and network power minimization subject to fronthaul constraints and SINR requirements.

In contrast to previously cited works which assume wired fronthaul links of infinite or fixed capacities, the authors in [22, 23, 24] consider wireless fronthaul links and study the weighted sum-rate maximization problem under different C-RAN setups. For the DL C-RAN [22], iterative algorithms are proposed under the assumptions of, firstly, single-cell processing where the RRHs operate as decode-and-forward (DF) relays and, secondly, cooperative processing where the RRHs operate as decompress-and-forward (DCF) relays. In particular, it is shown in this work that the cooperative processing outperforms single-cell processing. Likewise, the work in [23] shows that, in an UL C-RAN setup, a compress-and-forward (CF) cooperative scheme outperforms the DF-based per-cell reception. In order to satisfy the low latency requirement between the BBU and the RRHs, [24] adopts the multicast beamforming strategy over fronthaul links to deliver each user's message to a cluster of RRHs selected according to the user-centric clustering scheme, which then adopts the joint beamforming technique to cooperatively transmit the signal to the target users.

Most of the prior contributions focus on coordinated beamforming in both DL or UL C-RAN setups, although the exploitation of RRHs as relays to improve network capacity and coverage has remained hitherto largely undisclosed.[1] The authors in [27] address the design

[1] See, e.g., [25, 26], and references therein for a comprehensive review of system optimization for conventional relaying networks.

of a multiuser AF relaying subnetwork within a C-RAN network from an energy-efficient perspective, where the problem is cast as one of network energy minimization via joint RRH-selection and relaying matrix optimization, subject to a set of mean-square error-based quality-of-service (QoS) constraints. An iterative solution is proposed based on the concept of the re-weighted l_1-norm, along with a block-coordinate descent type algorithm. However, the resultant iterative algorithm exhibits relatively high computational complexity. Other works involving relays under C-RAN deal with capacity derivations. In [28], upper and lower bounds on the capacity of a DL symmetric C-RAN with multiple relays and a single receiver are obtained. Reference [29] consider a distributed massive MIMO system with multiple relays serving multiple sources jointly via a constrained fronthaul. The relays employ uniform quantization to meet the bandwidth constraints on the fronthaul, and numerical results reveal the outer bound of this scheme and the inevitable gap compared to the ideal C-RAN with infinite bandwidth.

1.3 Thesis contribution

In this work, inspired by [27], we investigate the problem of joint relay selection and transceiver optimization in a multiuser AF relaying subnetwork within C-RAN, but with the additional goal of reducing computational complexity. Specifically, we consider a subnetwork where multiple source-destination pairs communicate concurrently with the assistance of multiple cooperative RRHs connected to the BBU. In contrast to the more traditional network power minimization, we formulate the design problem as a regularized sparsity-induced interference leakage constrained minimization. To this end, we seek an efficient algorithm for the design of AF matrices which abide, firstly, by norm-based constraints dictating the per-relay antenna power budgets and, secondly, by linear constraints capable

of preserving the desired signals at each destination.

The problem is then converted into a form that is suitable for the application of the alternating direction method of multipliers (ADMM) [30]. Interestingly, a simple closed-form solution can be derived for each one of the main ADMM steps, leading to a very low-complexity iterative algorithm for relay selection and transceiver optimization. The closed-form solution leads to a thresholding operation that enables the selection of an energy efficient subset of RRHs, thus yielding solutions for the set of AF matrices that are exactly group-sparse. Simulation results show that the proposed algorithm can yield a satisfactory QoS level at all destinations with only a subset of active RRHs. In addition, the processing time of the proposed algorithm is significantly reduced as compared to benchmark algorithms relying on external optimization solvers such as SeDuMi and SDPT3.

The rest of the thesis is organized as follows. Chapter 2 presents background material on convex optimization, MIMO processing, and relaying principles. Chapters 3 specifies the system model and problem statement. In Chapter 4 the constrained and regularized interference leakage minimization problem is formulated, followed by the development of the low-complexity ADMM-based algorithm. Simulation results are presented and discussed in Chapter 5. Finally, we conclude the thesis in Chapter 6.

Notation: Vectors and matrices are represented by bold lower case and upper case letters, respectively. The transpose, Hermitian transpose, conjugate, and trace operators are respectively denoted by $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, and $\mathrm{Tr}(\cdot)$. The l_p-norm is denoted by $\|.\|_p$. The identity matrix of order N is denoted by $\mathbf{I}_N$. $\mathbb{R}$ and $\mathbb{C}$ represent the set of real numbers and complex numbers, respectively.

Chapter 2

Background

In this chapter, the necessary mathematical background on convex optimization is first presented in Section 2.1 based on material taken from references [31, 32]. Convex optimization is of particular interest for applications since finding a local optimum is sufficient for global optimality. Thereafter, Sections 2.2 and 2.3 review MIMO processing and relaying principles, respectively. The material for these sections have been adopted from the references [4, 33, 34].

2.1 Convex Optimization

2.1.1 Convex problems, duality and the KKT conditions

A set $\mathcal{C} \subseteq \mathbb{R}^n$ is convex if the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$. Formally, for two points $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and any α where $0 \leq \alpha \leq 1$ we have $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{C}$. A function $f(\mathbf{x})$ is convex if its domain $\mathbf{dom}(f)$ is convex and $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$ and $\alpha \in [0, 1]$. When $f(\mathbf{x})$ is differentiable, then the following

inequality holds:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f) \tag{2.1}$$

where $\nabla f(\mathbf{x})$ denotes the gradient of $f(.)$ at the point $\mathbf{x}$. More generally, and by definition, a subgradient $\mathbf{g}$ of a convex function f at $\mathbf{x} \in \mathbf{dom}(f)$ satisfies

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \ \forall \mathbf{y} \in \mathbf{dom}(f) \tag{2.2}$$

The subdifferential $\partial f(\mathbf{x})$ of $f(.)$ at $\mathbf{x}$ is the set of all subgradients:

$$\partial f(\mathbf{x}) = \{\mathbf{g} | \mathbf{g}^T(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}), \ \forall \mathbf{y} \in \mathbf{dom}(f)\} \tag{2.3}$$

Consider the optimization problem in the variable $\mathbf{x} \in \mathbb{R}^n$:

$$\min_{\mathbf{x}} \quad f_0(\mathbf{x}) \tag{2.4a}$$

$$\text{s.t.} \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m \tag{2.4b}$$

$$h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p \tag{2.4c}$$

If $f_i(\mathbf{x})$ are convex $\forall i$ and $h_j(\mathbf{x})$ are affine[1] $\forall j$, then the above is said to be a convex optimization problem, with a domain given by $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}(f_i) \cap \bigcap_{j=1}^{p} \mathbf{dom}(h_j)$. The feasible (or constraint) set of the problem is $\mathcal{F} = \{\mathbf{x} \in \mathcal{D} | f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \quad \text{and} \quad h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p\}$. A problem is said to be feasible if $\mathcal{F}$ is a non-empty set; in this case $\mathbf{x}^\star$ is said to be an optimal point if $f(\mathbf{x}^\star) = \inf\{f_0(\mathbf{x}) | \mathbf{x} \in \mathcal{F}\}$. Conversely, the problem is infeasible if $\mathcal{F}$ is empty; in this case, $\inf\{f_0(\mathbf{x}) | \mathcal{F}\} = \inf\{f_0(\mathbf{x}) | \mathbf{x} \in \emptyset\} = +\infty$, by convention of infimum over the empty set $\emptyset$.

[1] We say a function $A : \mathbb{R}^n \to \mathbb{R}$ is affine if there exists a linear function $L : \mathbb{R}^n \to \mathbb{R}$ and a scalar $b \in \mathbb{R}$ such that $A(\mathbf{x}) = L(\mathbf{x}) + b$, for all $\mathbf{x} \in \mathbb{R}^n$.

At each feasible point $\tilde{\mathbf{x}}$ in $\mathcal{F}$, the constraints in (2.4b)-(2.4c) are satisfied and we thus have $\sum_{i=1}^{m} u_i f_i(\tilde{\mathbf{x}}) \leq 0$ and $\sum_{j=1}^{p} v_j h_j(\tilde{\mathbf{x}}) = 0$ where $\mathbf{u} = [u_1, \ldots, u_m] \in \mathbb{R}^m, \quad u_i \geq 0, \quad i = 1, \ldots, m$ and $\mathbf{v} = [v_1, \ldots, v_p] \in \mathbb{R}^p$. By constructing the so-called Lagrange function as

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i f_i(\mathbf{x}) + \sum_{j=1}^{p} v_j h_j(\mathbf{x}), \tag{2.5}$$

it follows from the previous observation that

$$\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \leq f_0(\tilde{\mathbf{x}}) \tag{2.6}$$

$$\implies \min_{\mathbf{x} \in \mathcal{C}} \ \mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) \leq \min_{\mathbf{x} \in \mathcal{C}} f_0(\mathbf{x}) = f(\mathbf{x}^\star) \tag{2.7}$$

$$\implies \min_{\mathbf{x}} \ \mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) \leq \min_{\mathbf{x} \in \mathcal{C}} \ \mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) \leq f(\mathbf{x}^\star) \tag{2.8}$$

The *dual function* defined as $g(\mathbf{u}, \mathbf{v}) \triangleq \min_{\mathbf{x}} \ \mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v})$ forms a lower bound on the optimal solution $f(\mathbf{x}^\star)$ for any feasible *dual variables* $(\mathbf{u}, \mathbf{v})$. Consequently, an alternative way of finding the solution of (2.4) is the maximization of $g(\mathbf{u}, \mathbf{v})$ subject to $u_i \geq 0, \quad i = 1, \ldots, m$. This leads to the construction of the *dual problem*, which is a concave maximization problem regardless of the convexity of the primal problem. The optimal solution to the dual problem is denoted as $g(\mathbf{u}^\star, \mathbf{v}^\star)$.

Ideally, one would like to obtain *strong duality*, that is, the dual optimal solution $g(\mathbf{u}^\star, \mathbf{v}^\star) = f(\mathbf{x}^\star)$. With no *a priori* assumption about the convexity of the problem, if strong duality holds the following Karush-Kuhn-Tucker (KKT) conditions are necessarily satisfied at the optimal point $(\mathbf{x}^\star, \mathbf{u}^\star, \mathbf{v}^\star)$:

- Primal feasibility: $f_i(\mathbf{x}^\star) \leq 0, \ \text{for } i = 1, \ldots, m \text{ and } h_j(\mathbf{x}^\star) = 0, \quad j = 1, \ldots, p$
- Dual feasibility: $u_i^\star \geq 0, \ \text{for } i = 1, \ldots, m$

- Complementary slackness: $u_i^\star f_i(\mathbf{x}^\star) = 0,$ for $i = 1, \ldots, m$
- Stationarity: $\mathbf{0} \in \partial\mathcal{L}(\mathbf{x}^\star, \mathbf{u}^\star, \mathbf{v}^\star)$

where $\partial\mathcal{L}(\mathbf{x}^\star, \mathbf{u}^\star, \mathbf{v}^\star)$ is the subdifferential of $\mathcal{L}$ at $(\mathbf{x}^\star, \mathbf{u}^\star, \mathbf{v}^\star)$.

Slater's condition states that strong duality holds when the primal problem is convex and strictly feasible, i.e., there exists $\mathbf{x}$ in the relative interior of $\mathcal{D}$ such that $f_i(\mathbf{x}) < 0, \quad i = 1, \ldots, m \quad$ and $\quad h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p.$ For convex problems, hence, satisfying the KKT conditions is sufficient for optimality for any triplet $(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}, \tilde{\mathbf{v}})$.

2.1.2 Final Remarks

While the presentation of convex optimization was given for real variables, it can be readily extended into the complex domain by considering the real and imaginary parts as forming a large vector. That way, optimziation over $\mathbb{C}^n$ becomes optimization over $\mathbb{R}^{2n}$.

2.1.3 An overview of the ADMM

Consider the convex optimization problem in the variables $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{z} \in \mathbb{C}^n$:

$$\min_{\mathbf{x},\mathbf{z}} \quad F(\mathbf{x}) + G(\mathbf{z}) \tag{2.9a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{C}_1, \ \mathbf{z} \in \mathcal{C}_2 \tag{2.9b}$$

$$\mathbf{z} = \mathbf{A}\mathbf{x} \tag{2.9c}$$

where $F(\cdot) : \mathbb{C}^m \to \mathbb{R}$ and $G(\cdot) : \mathbb{C}^n \to \mathbb{R}$ are convex functions, $\mathbf{A}$ is an $n \times m$ matrix, and $\mathcal{C}_1 \subseteq \mathbb{C}^m$ and $\mathcal{C}_2 \subseteq \mathbb{C}^n$ are non-empty convex sets. The solution to the above problem may be achieved by considering an equivalent problem obtained by adding to the objective function (2.9a) a penalty term $\frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2$ where $\rho > 0$ is a penalty parameter [35]. The

ADMM is an iterative procedure for solving this equivalent problem. Assigning a Lagrange multiplier $\mathbf{y}$ to the equality constraint (2.9c) and letting $\mathbf{x}^{(j)}$, $\mathbf{z}^{(j)}$, and $\mathbf{y}^{(j)}$ denote the values of $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{y}$ at the j^{th} iteration, the ADMM consists of the following steps:

$$\mathbf{x}^{(j+1)} = \arg\min_{\mathbf{x}\in\mathcal{C}_1} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^{(j)}, \mathbf{y}^{(j)}) \tag{2.10}$$

$$\mathbf{z}^{(j+1)} = \arg\min_{\mathbf{z}\in\mathcal{C}_2} \mathcal{L}_\rho(\mathbf{x}^{(j+1)}, \mathbf{z}, \mathbf{y}^{(j)}) \tag{2.11}$$

$$\mathbf{y}^{(j+1)} = \mathbf{y}^{(j)} + \frac{\rho}{2}(\mathbf{A}\mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)}) \tag{2.12}$$

where the *augmented* Lagrangian function is given by

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = F(\mathbf{x}) + G(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \mathbf{y}^H(\mathbf{A}\mathbf{x} - \mathbf{z}) + (\mathbf{A}\mathbf{x} - \mathbf{z})^H\mathbf{y}. \tag{2.13}$$

The ADMM performs a dual subgradient ascent in (2.12) where the first two steps (2.10)-(2.11) are necessary to compute the subgradient. Since (2.10) and (2.11) are single-variable problems in $\mathbf{x}$ and $\mathbf{z}$ respectively, it is readily observed that the algorithm decouples the minimization problem. In some cases, this decoupling enables us to leverage the structures of (2.10) and (2.11) to obtain closed-form solutions, as we shall illustrate in Chapter 4. In the case of real variables, the step size in the subgradient ascent (2.12) would be set to ρ, this choice being motivated by the fact that it allows for a partial satisfaction of the KKT optimality conditions [30]. However, in the case of complex variables, which is the situation of interest in this thesis, the step size must be set to $\frac{\rho}{2}$, as further discussed below.

Step-size selection and convergence of ADMM

In anticipation of our development in Chapter 3, let us assume that the convex problem we're attempting to solve may be expressed as

$$\min_{\mathbf{x},\mathbf{z}} \quad f(\mathbf{x}) + G(\mathbf{z}) \tag{2.14}$$

$$\text{s.t.} \quad \mathbf{\Phi}^H\mathbf{x} = \mathbf{c}, \tag{2.15}$$

$$\mathbf{x} = \mathbf{z}, \tag{2.16}$$

$$\mathbf{z}_i{}^H\mathbf{z}_i \leq P_i, \ i = 1, \ldots, L \tag{2.17}$$

where we consider that $G(\mathbf{z}) = \sum_{i=1}^{L} g(\mathbf{z}_i)$ where $\mathbf{z}$ is partitioned as $\mathbf{z} = [\mathbf{z}_1^T \ldots \mathbf{z}_L^T]^T$ with $\mathbf{z}_i \in \mathbb{C}^{N_i^2}$. The Lagrangian associated with this problem is given as

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{z},\mathbf{y},\boldsymbol{\nu},\boldsymbol{\mu}) =& f(\mathbf{x}) + \sum_{i=1}^{L} g(\mathbf{z}_i) - \mathbf{y}^H(\mathbf{z}-\mathbf{x}) - (\mathbf{z}-\mathbf{x})^H\mathbf{y} - \boldsymbol{\nu}^H(\mathbf{\Phi}^H\mathbf{x}-\mathbf{c}) - (\mathbf{x}^H\mathbf{\Phi}-\mathbf{c}^H)\boldsymbol{\nu} \\ &+ \sum_{i=1}^{L} \mu_i(\mathbf{z}_i^H\mathbf{z}_i - P_i)\end{aligned}$$

with $\mathbf{y}$, $\boldsymbol{\nu}$, and μ_i being the Lagrange *dual variables* associated with the constraints $\mathbf{x} = \mathbf{z}$, $\mathbf{\Phi}^H\mathbf{x} = \mathbf{c}$, and $\mathbf{z}_i{}^H\mathbf{z}_i \leq P_i$, respectively. The optimal point $(\mathbf{x}^\star, \mathbf{z}^\star, \mathbf{y}^\star, \boldsymbol{\nu}^\star, \boldsymbol{\mu}^\star)$ must satisfy the KKT conditions:

Primal feasiblity:

$$\mathbf{\Phi}^H\mathbf{x}^\star = \mathbf{c} \tag{2.18}$$

$$\|\mathbf{z}_i^\star\|_2^2 \leq P_i, \ \text{for } i = 1, \ldots, L \tag{2.19}$$

$$\mathbf{x}^\star = \mathbf{z}^\star \tag{2.20}$$

Dual feasibility:

$$\mu_i^\star \geq 0, \text{ for } i = 1, \ldots, L \tag{2.21}$$

Complementary slackness:

$$\mu_i^\star(\|\mathbf{z}_i^\star\|_2^2 - P_i) = 0, \text{ for } i = 1, \ldots, L \tag{2.22}$$

Stationary w.r.t the Lagrangian:

$$\mathbf{0} \in \partial f(\mathbf{x}^\star) + \mathbf{y}^\star - \mathbf{\Phi}\boldsymbol{\nu}^\star \tag{2.23}$$

$$\mathbf{0} \in \partial g(\mathbf{z}_i^\star) - \mathbf{y}_i^\star + \mu_i^\star \mathbf{z}_i^\star, \text{ for } i = 1, \ldots, L \tag{2.24}$$

Our solution methodology consists of augmenting the problem and using the ADMM. At iteration $k+1$, the first step of ADMM involves finding

$$\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{\Phi}^H \mathbf{x} = \mathbf{c}} f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{z}^{(k)} - \mathbf{x}\|_2^2 + \mathbf{y}^{(k)H}\mathbf{x} + \mathbf{x}^H\mathbf{y}^{(k)}$$

We solve this problem by finding the values $\mathbf{x}^{(k+1)}$ and $\boldsymbol{\nu}^{(k+1)}$ which satisfy its KKT conditions:

$$\mathbf{\Phi}^H\mathbf{x}^{(k+1)} = \mathbf{c} \tag{2.25}$$

$$\mathbf{0} = \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{y}^{(k)} + \frac{\rho}{2}(\mathbf{x}^{(k+1)} - \mathbf{z}^{(k)}) - \mathbf{\Phi}\boldsymbol{\nu}^{(k+1)} \tag{2.26}$$

Consequently, the optimality condition given by (2.18) is satisfied at iteration $k+1$ from (2.25).

In the second step of ADMM, the algorithm solves L parallel subproblems. At iteration

$k+1$, we have for the i^{th} subproblem that the solution point satisfies:

$$\begin{aligned} \|\mathbf{z}_i^{(k+1)}\|_2^2 &\leq P_i, \text{ for } i = 1, \ldots, L \\ \mu_i^{(k+1)} &\geq 0, \text{ for } i = 1, \ldots, L \\ \mu_i^{(k+1)}(\|\mathbf{z}_i^{(k+1)}\|_2^2 - P_i) &= 0, \text{ for } i = 1, \ldots, L \end{aligned}$$

As a result, we can see that the optimality conditions (2.19), (2.21), and (2.22) are satisfied. In addition, the solution point at the second step satisfies:

$$\mathbf{0} \in \partial g(\mathbf{z}_i^{(k+1)}) - \mathbf{y}_i^{(k)} - \frac{\rho}{2}(\mathbf{x}_i^{(k+1)} - \mathbf{z}_i^{(k+1)}) + \mu_i^{(k+1)}\mathbf{z}_i^{(k+1)} \tag{2.27}$$

Now by setting $\mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} + \frac{\rho}{2}(\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)})$ in the third step of ADMM, we find from (2.27) that

$$\mathbf{0} \in \partial g(\mathbf{z}_i^{(k+1)}) - \mathbf{y}_i^{(k+1)} + \mu_i^{(k+1)}\mathbf{z}_i^{(k+1)}, \text{ for } i = 1, \ldots, L \tag{2.28}$$

That is, by choosing $\frac{\rho}{2}$ as the step-size, we achieve the optimality condition (2.24). Additionally, we note that as a result of the step-size selection, (2.26) can be expressed as

$$\mathbf{0} = \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{y}^{(k)} + \frac{\rho}{2}(\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)}) + \frac{\rho}{2}(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}) - \mathbf{\Phi}\boldsymbol{\nu}^{(k+1)} \tag{2.29}$$

$$\mathbf{0} = \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{y}^{(k+1)} + \frac{\rho}{2}(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}) - \mathbf{\Phi}\boldsymbol{\nu}^{(k+1)} \tag{2.30}$$

$$-\frac{\rho}{2}(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}) = \nabla f(\mathbf{x}^{(k+1)}) + \mathbf{y}^{(k+1)} - \mathbf{\Phi}\boldsymbol{\nu}^{(k+1)} \tag{2.31}$$

Termination Criterion

The conditions given by (2.20) and (2.23) still need to be satisfied for optimality. To achieve (2.20), we must have $\mathbf{r}^{(k+1)} \triangleq \mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)} \longrightarrow \mathbf{0}$ as $k \longrightarrow \infty$, where $r^{(k+1)}$ is called the *primal residual*. Similarly, the *dual residual* given by $\mathbf{s}^{(k+1)} \triangleq -\frac{\rho}{2}(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)})$ from (2.31) must go to $\mathbf{0}$ as $k \longrightarrow \infty$ to achieve (2.23). Accordingly, a stopping criterion for ADMM which has been often adopted in the literature [30] can be given in terms of these residuals. Specifically, the algorithm is assumed to have converged when

$$\|\mathbf{r}^{(j+1)}\|_2 \leq \epsilon^{pri} \tag{2.32}$$

$$\|\mathbf{s}^{(j+1)}\|_2 \leq \epsilon^{dual} \tag{2.33}$$

where ϵ^{pri} and ϵ^{dual} are non-negative (yet relatively small) feasibility tolerances for the primal and dual residuals, respectively. This is why the step-size selection is crucial for convergence.

2.2 MIMO Processing

A multiple-input multiple-output (MIMO) system can be generally defined as a communication system consisting of a transmitter equipped with N_t antennas, a radio propagation channel, and a receiver equipped with N_r antennas, as illustrated in Figure 2.1. For a narrowband (flat-fading) time-invariant MIMO wireless channel, the received signal vector can be represented as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.34}$$

where $\mathbf{x} \in \mathbb{C}^{N_t}$ is the transmitted signal, $\mathbf{y} \in \mathbb{C}^{N_r}$ is the received signal, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix and $\mathbf{n} \in \mathbb{C}^{N_r}$ is a complex noise vector. The input signal $\mathbf{x}$ is modeled as a

random vector[2] with zero mean and covariance matrix $\mathbf{R}_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^H]$. The (i, j)-th element of matrix $\mathbf{H}$, denoted as h_{ij} is the complex channel gain from transmit antenna j to receive antenna i. We shall assume that the noise vector $\mathbf{n}$ abides by a complex circular Gaussian multi-variate distribution with zero mean and covariance matrix $\sigma^2\mathbf{I}_{N_r}$. For simplicity, given a transmit power P, we consider an equivalent model with a noise power σ^2 of unity and transmit power $P/\sigma^2 = \rho$, where ρ can be interpreted as the average SNR per receive antenna under channel unity gain. This power constraint implies that the input symbols satisfy

$$\sum_{i=1}^{N_r} \mathrm{E}[|x_i|^2] = \rho, \tag{2.35}$$

or equivalently that $\mathrm{Tr}(\mathbf{R}_{\mathbf{x}}) = \rho$.

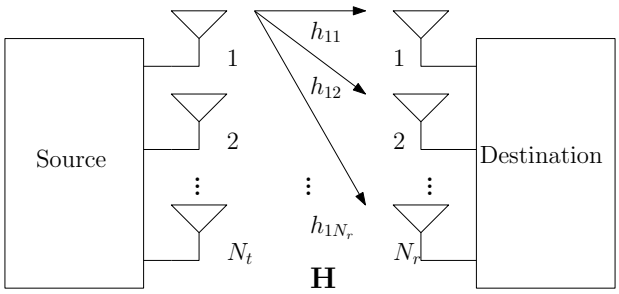

Fig. 2.1: MIMO system model with N_t transmit antennas and N_r receive antennas.

Depending on the scenario being examined, the knowledge of the channel matrix $\mathbf{H}$ at the transmitter and receiver may or may not be available. We refer to the former as channel state information at the transmitter (CSIT) and to the latter as channel state information at the receiver (CSIR).

When a destination is equipped with multiple antennas, processing of the received signals results in so-called spatial diversity. Spatial diversity allows the improvement of the

[2]In practice, the entries of $\mathbf{x}$ are taken from a discrete constellation, such as BPSK or M-QAM [36].

total quality of the received signals from different antennas, depending on the signal processing approach implemented at the receiver. In general, there are two ways of processing signals from multiple diversity branches: selection diversity and combining diversity.

1. Selection diversity is the method whereby the "best" signal copy is selected and processed (demodulated and decoded), while all other copies are discarded. Different criteria are available for what constitutes the "best" signal.

2. Combining diversity is the method whereby all copies of the signal are combined (before or after demoluation), and the combined signal is decoded. Again, different approaches are available for the combination of the signals.

In both cases, the processing algorithms assume that the different signal copies undergo statistically independent fading. The performance gain resulting from the use of multiple antennas is due to two effects: *diversity gain* and *beamforming gain*. Diversity gain reflects the fact that it is improbable that several antenna elements are in a fading dip simultaneously; thus the probabiliy of very low signal levels is decreased by the use of multiple antennas. Beamforming gain reflects the fact that (for combining diversity) the combiner performs an averaging over the noise at different antennas. Indeed, due to the assumed uncorrelated nature of the noise signals at the different antennas, even if the signal levels at all antenna elements are identical, the combiner ouput SNR is larger than the SNR at a single-antenna element. Below, we discuss in further details the processing approaches that are typically employed for selection and combining diversity.

2.2.1 Selection Diversity

In selection diversity, the receiver monitors the SNR of the received signal from each diversity branch, and selects only the received signal corresponding to the highest SNR for

detection. If the bit-error-rate (BER) is predominantly determined by the noise power level, selection based on the received-signal is the best among all selection diversity methods, since this maximimizes the SNR. However, if the BER is determined by the level of co-channel interference, this method may fail. For example, the branch with the highest signal-strength may be due to a high level of interference on the corresponding radio link, and thus resulting in a low SINR level and poor BER performance.

Other selection diversity approaches are BER-driven. In these approaches, a training sequence (known at the receiver) is transmitted and demodulated at the receiver, which selects the antenna whose associated BER is lowest, by comparing the demodulated signal with the known training sequence.

2.2.2 Combining Diversity

Combining diversity leads to better performance than selection diveristy, as all available information is exploited. However, the receiver is more complex than spatial diversity (i.e., antenna selection) receivers. In combining-type receivers, each signal copy is multiplied by a complex weight and then added up. Each complex weight w_n^* can be thought of as consisting of a phase correction, plus a weight for the amplitude:

- Phase correction causes the signal amplitudes to add up, while on the other hand, noise is added incoherently so that noise powers add up (as opposed to amplitudes).
- For amplitude weighting, two methods are widely used: *Maximum ratio combining* (MRC) and *Equal gain combining* (EGC).

Maximum Ratio Combining (MRC)

Maximum ratio combining removes the phase distortion introduced by each branch (or antenna) so that signals can be combined coherently and selects the weighting factor amplitude proportionally to the branch channel gain. Indeed, since a larger branch amplitude yields higher SNR, more weight should be put on the corresponding received signal (with better quality).

Assuming a time-invariant and flat-fading propagation channel, the only disturbance in the received signal is additive white gaussian noise (AWGN). Under these assumptions, each channel realization may be viewed as a time-invariant filter with impulse response

$$h_n(\tau) = \alpha_n \delta(\tau)$$

where $\alpha_n \in \mathbb{C}$ is the instantaneous gain of diversity branch n and $\delta(\tau)$ represents an ideal impulse or Dirac delta function at time τ. The received signals at the different branches are multiplied with corresponding weight w_n^* and added up, so that the SNR after processing is equal to

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\left|\sum_{n=1}^{N} w_n^* \alpha_n\right|^2}{P \sum_{n=1}^{N} |w_n|^2} \tag{2.36}$$

where P is the noise power per branch, assumed to be the same for all branches. According to the Cauchy-Schwartz inequality, we find that

$$\left|\sum_{n=1}^{N} w_n^* \alpha_n\right|^2 \leq \sum_{n=1}^{N} |w_n|^2 \sum_{n=1}^{N} |\alpha_n|^2 \tag{2.37}$$

with equality if and only if $w_n = \alpha_n$. Thus, the SNR is maximized in MRC by choosing the weights as $w_{\mathrm{MRC}} = \alpha_n$.

MRC achieves the best transmission performance at the cost of receiver complexity, i.e., the receiver needs to know (or estimate by some appropriate means) both the branch amplitude and phase.

Equal Gain Combining (EGC)

In EGC, the signals from N diversity branches are weighted equally and coherently combined. The receiver complexity is reduced in comparison with the MRC-type combining, since only the phase distortion introduced by each branch needs to be known (or estimated) so that signals can be combined coherently. The performances of EGC is better than selection-diversity but worse than MRC.

Wireless communications are subject to non-line-of-sight (NLOS) propagation, producing random fluctuations in the received signal level (also known as fading). MIMO technology has been known to be a reliable tool against fading and has been very useful in NLOS environments. In the next subsections, we will discuss the advantages of using multiple antennas at the transmitter and/or the receiver; which mainly take the form of array gain, diversity gain and spatial multiplexing gain.

2.2.3 Array gain

Array gain is defined as the increase in the average receive SNR due to coherent combining of signals at the transmitter or receiver. For instance, consider the case of MRC at the receiver, and assume for simplicity that the channel gains are identical, i.e., $\alpha_n \equiv \alpha$ for all n. Hence, the corresponding input SNR on each branch is $\mathrm{SNR}_{\mathrm{in}} = |\alpha|^2/P$ while the output SNR simplifies to $\mathrm{SNR}_{\mathrm{out}} = N|\alpha|^2/P$. Consequently, the array gain is given by

$$\mathrm{AG} = \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}} = N \tag{2.38}$$

We note that to achieve an array gain, the CSI is generally required. CSIR is often available through transmission of pilots symbols. The challenge lies in obtaining or estimating the CSIT.

Array gain can be attained with transmitter space diversity through scaling the antenna transmit powers relative to the channel gains. On the receiver side, the different signal paths received by multiple antennas can be combined in different manners –differing in complexity and achievable array gain – to obtain a signal which then is processed through a demodulator. Under certain general modeling conditions, the combiner outputs the original transmitted signal scaled by a random complex amplitude term, resulting in a random SNR. The distribution of this SNR depends on the number of diversity paths, the fading distribution of each path and the combining technique. MRC is often the most favorable scheme since the SNR of the combiner output is the sum of the input SNRs on each branch. Therefore, as exemplified above, the combiner output SNR and the array gain increase linearly with the number of diversity branches.

2.2.4 Diversity gain

Sending the same data over independent paths provides diversity gain. The latter is formally defined as the change in the slope of the error probability in the limit of high SNR, resulting from diversity combining at the receiver. A scheme is said to achieve a diversity gain d if the average error probability as a function of SNR, i.e. $P_e(\mathrm{SNR})$, satisfies

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} = -d \tag{2.39}$$

In a uniform scattering environment, the fading amplitudes corresponding to each antenna become approximately independent when the antenna separation exceeds about half

the wavelength [33]. This in turn makes it possible to achieve maximum diversity gain. Similar to the SNR distribution in connection with the discussion of array gain, the probability of error in (2.39) generally depends on the number of diversity paths, the fading distribution of each path and the combining technique. From the above mentioned combining techniques, MRC was shown in [33] to produce full diversity order which is equal to the number of antennas at the receiver.

2.2.5 Spatial Multiplexing gain

A MIMO system provides the benefit of increasing the capacity, almost linearly with the number of antennas. This increase in capacity which is referred to as the spatial multiplexing gain, is obtained at no additional power or bandwidth. The basic approach to achieve this gain is to transmit independent data symbols from individual antennas and allow the receiver to separate the different streams through signal processing. Particularly, N_t independent symbols are transmitted per symbol period; i.e. the spatial code rate is N_t. Several encoding options at the transmitter can be employed to realize spatial multiplexing. These include so-called horizontal encoding (HE), vertical encoding (VE) and combinations thereof. These three schemes, described in [37], all achieve N_t as the spatial code rate and a diversity gain of N_r. This allows us to view a MIMO channel as $\min\{N_t, N_r\}$ parallel spatial channels with $\min\{N_t, N_r\}$ degrees of freedom. The channel capacity, therefore, increases with the SNR as $\min\{N_t, N_r\} \log \mathrm{SNR}$.

2.2.6 Diversity and multiplexing tradeoff

The tradeoff between complexity, data rate and probability of error has been examined extensively in the literature. It has been deduced that multiple antennas can be used for multiplexing and diversity gains at the same time. Assuming CSIT is not available, full

diversity gain and full multiplexing gain can be achieved simultaneously by encoding diagonally across antennas in block fading channels, with asymptotically large block lengths. One good example is the Diagonal Bell Labs Space Time (D-BLAST) where the data stream is first parallel-encoded and the codewords are rotated across antennas so that they are transmitted by all N_t antennas.

Transmission schemes based on D-BLAST can achieve full diversity under the condition that stream rotated temporal coding is capacity achieving. Due to the fact that the receiver decodes each diagonal code signal independently, the receiver complexity is linear in N_t. Nonetheless, the wasted space-time dimensions with an incorrect choice of frame size results in a loss of efficiency. However, if we ignore the wasted space-time dimensions along the diagonals, the D-BLAST system can achieve maximum capacity with outage.

2.2.7 Capacity schemes

The channel capacity is the maximum data rate that can be transmitted over the channel with arbitrarily small error probability. The capacity is very much dependent on the knowledge of the channel gain matrix or its distribution at the transmitter or receiver.

Static Channels

For static channels, a good estimate of $\mathbf{H}$ can be obtained fairly easily at the receiver, so we assume the availability of CSIR throughout this section. Under this assumption, the capacity is given in terms of the mutual information $I(\mathbf{x};\mathbf{y})$ between the channel input vector $\mathbf{x}$ and output vector $\mathbf{y}$ as

$$C = \max_{p(\mathbf{x})} I(\mathbf{x};\mathbf{y}) \tag{2.40}$$

where $p(\mathbf{x})$ is the probability distribution of the input. It turns out that the optimal distribution $p(\mathbf{x})$ that maximizes the capacity is for $\mathbf{x}$ to be zero-mean, circularly symmetric complex Gaussian, subject to a power constraint, i.e. $\mathrm{Tr}(\mathbf{R_x}) = \rho$.

$$C = \max_{\mathbf{R_x}:\mathrm{Tr}(\mathbf{R_x})=\rho} B \log_2 \det[\mathbf{I}_{N_r} + \mathbf{H}\mathbf{R_x}\mathbf{H}^H]. \tag{2.41}$$

where B is the channel bandwidth.

Channel known at the Transmitter

When the fixed channel is known at both the transmitter and receiver, using transmit precoding and receiver shaping as obtained from the singular value decompositon of $\mathbf{H}$, allows a simple characterization of the MIMO channel capacity. Specifically, the capacity equals the sum of capacities on each of the independent parallel channels with the transmit power optimally allocated between the latter. The capacity in this case is given by

$$C = \max_{\sum_i \rho_i \leq \rho} \sum_{i=1}^{\nu} B \log_2(1 + \sigma_i^2 \rho_i) \tag{2.42}$$

where ν is the number of nonzero singular values σ_i of $\mathbf{H}$. Since the MIMO channel decomposes into ν parallel channels, we say that it has ν degrees of freedom. Considering that $\rho = P/\sigma^2$, the capacity can also be expressed in terms of the power allocation P_i to the i-th parallel channel as

$$C = \max_{\sum_i P_i \leq P} \sum_{i=1}^{\nu} B \log_2(1 + \frac{P_i \gamma_i}{P}) \tag{2.43}$$

where $\gamma_i = \sigma_i^2 P/\sigma^2$ is the SNR associated with the i-th channel at full power. The power-allocation scheme which maximizes this capacity is the water-filling scheme, yielding a

capacity

$$C = \sum_{i:\gamma_i \geq \gamma_0} B \log_2(\frac{\gamma_i}{\gamma_0}) \tag{2.44}$$

for some cutoff value γ_0.

Fading Channels

Suppose now that the channel gain matrix experiences flat fading, so the gains h_{ij} vary with time. As in the case of the static channel, the capacity depends on what is known about the channel matrix at the transmitter and receiver. With perfect CSIR and CSIT the transmitter can adapt to the channel fading. In this case, the capacity equals the average over all the channel matrix realizations with optimal power allocation. This average capacity over the channel realization is also referred to as the ergodic channel capacity.

Channel Known at the transmitter

With CSIT and CSIR, the transmitter optimizes its transmission strategy for each fading channel realization, as in the case of a static channel. Here, there are two possibilities for allocating power from the perspective of ergodic capacity. A short-term power constraint assumes that the power associated with each channel realization must be equal to the average power constraint $\bar{P}$. In this case the ergodic capacity becomes

$$C = \mathrm{E}_{\mathbf{H}} \left[\max_{P_i: \sum_i P_i \leq \bar{P}} \sum_i B \log_2(1 + \frac{P_i \gamma_i}{\bar{P}}) \right] \tag{2.45}$$

where $\mathrm{E}_{\mathbf{H}}$ denotes expectation over the channel realizations. A less restrictive constraint is a long-term power constraint, where we can use different powers $P_{\mathbf{H}}$ for different channel realizations $\mathbf{H}$ subject to a constraint on the average power over all channel realizations,

i.e. $\mathrm{E}_{\mathbf{H}}[P_{\mathbf{H}}] \leq \bar{P}$. The ergodic capacity under this assumption is given by

$$C = \max_{P_{\mathbf{H}}:\mathrm{E}_{\mathbf{H}}[P_{\mathbf{H}}]\leq\bar{P}} \mathrm{E}_{\mathbf{H}} \left[\max_{P_i:\sum_i P_i \leq P_{\mathbf{H}}} \sum_i B \log_2(1 + \frac{P_i \gamma_i}{P_{\mathbf{H}}}) \right] \tag{2.46}$$

for $\gamma_i = \sigma_i^2 P_{\mathbf{H}}/\sigma^2$. The short-term power constraint gives rise to a one-dimensional water-filling in space across the antennas, whereas the long-term power constraint allows for a two-dimensional water-filling across both space and time. The expectation with respect to $\mathbf{H}$ in (2.45)-(2.46) can be evaluated based on the distribution of the singular values σ_i of the matrix $\mathbf{H}$.

2.3 Relaying principles

There are two key objectives associated with relay-assisted signaling in wireless communication systems such as cellular networks. First, relay-assisted communication should extend user coverage, whereby more mobile users can achieve their minimum SINR requirements over a larger area. Second, relay-assisted communication should increase system capacity, whereby each mobile user that is being serviced can benefit from higher data rates.

The performance of a wireless relay network depends on two major factors: network configuration and relaying strategy. Network configuration refers to the number of source, destination and relay nodes, as well as their number of antennas. For a particular confiuration, system performance depends on the method of information transmission at the relays. Relaying strategies can be can be non-regenerative such as amplify-and-forward (AF), or regenerative such as decode-and-forward (DF) [38].

Fig. 2.2: N-hop relay channel with $N-1$ single-antenna relay nodes (R) and single-antenna source (S) and destination (D).

2.3.1 Amplify-and-forward (AF) Relaying

Amplify-and-forward is a strategy whereby the relay amplifies the signal from the source and transmits it to the destination, subject to a power constraint. The linear processing applied to the signal is usually represented by scalar multiplication for single-antenna relays, or matrix multipliciation for multi-antenna relays.

We consider the simplified configuration in Fig. 2.2 to describe the system model of the AF relaying strategy, where a source exchanges information with a destination through $N-1$ relays, where all the nodes are single-antenna. We also assume that useful communication paths are only available between adjacent nodes. The signal received by the i-th relay is $y_i = h_i x_{i-1} + n_i$, for $i = 1, \ldots, N-1$, where x_{i-1} is the signal (here, a complex scalar symbol) transmitted by the $(i-1)$-th relay, h_i is the i-th hop channel and n_i is the i-th relay AWGN noise[3]. The signal transmitted by the i-th relay in the AF mode is $x_i = \sqrt{K_i} y_i$, where K_i is its scalar power gain. Thus, the input-output relationship of the whole N-hop AF relay channel can be expressed as

$$r = \prod_{k=1}^{N} h_k s + \sum_{k=1}^{N} \Big(\prod_{l=k+1}^{N} h_l \Big) n_k \tag{2.47}$$

where $s = x_0$ and $r = y_N$ are the transmitted (source) and received (destination) signal vectors. We note that the channel gains in (2.47) are scalar for the simplified model in

[3]With obvious modifications for the $S - R_1$ and $R_{N-1} - D$ links

Fig. 2.2; however, for multi-antennas or MIMO relays, similar relationships can be derived where the channels are represented by matrices instead. The non-regenerative strategy is the simplest one, but has the drawback of amplifying the noise at each relay station and forwarding it to the destination as observed from the above equation.

2.3.2 Decode-and-forward (DF) Relaying

Decode-and-forward is a strategy whereby the relay has to first decode the source symbol and then retransmit it to the destination. The complexity of the DF strategy is increased in comparison with the AF strategy because of the additional message estimation process involved. The performance of this method is thus constrained by the success of the decoding phase. In this case, the input-output relationship for the N-hop DF relay channel can be summarized by the following chain of equations:

$$\begin{aligned} y_i &= h_i x_{i-1} + n_i, \quad i = 1, \ldots, N \\ x_i &= \phi(x_{i-1}) \end{aligned} \tag{2.48}$$

where $\phi(.)$ denotes the decoding function and x_N is the decoded signal at the destination. We note from these equations that under high SNR conditions, perfect transmission of the source signal to the destination is possible with the DF approach.

2.3.3 Comparison of AF and DF Relaying

Let C_{DF} and C_{AF} denote the capacities achieved by the DF and AF schemes, respectively. As stated in [39], the capacity gain $\Delta C = C_{DF} - C_{AF}$ of DF relaying over AF relaying in a N-hop relay channel as in Fig. 2.2, is bounded as:

$$0 \leq \Delta C \leq \log_2 N \tag{2.49}$$

for any channel realizations h_i, $i = 1, \ldots, N$. When C_{DF} is finite, the lower bound is achieved if and only if the destination and all the relays but one are noiseless. The upper bound is achieved if and only if all the hops are equally strong and the SNR is sufficiently high.

The advantage of the DF relaying over AF relaying can also be cast in terms of an SNR gain. Define the SNR gain G of DF relaying from the following expression:

$$C_{AF}(G\gamma_0) = C_{DF}(\gamma_0) \tag{2.50}$$

where γ_0 is the source SNR. Gain G represents that additional SNR (when measured in dB) required for the AF relaying strategy to achieve the same capacity as DF relaying. Based on (2.49), the SNR gain of DF relaying over AF relaying for the N-hop relay channel in Fig. 2.2 is bounded as:

$$1 \leq G \leq N \tag{2.51}$$

In practical applications with a few relay stations or with all the links being equally impaired by noise, AF is often preferred to DF since the small SNR gain of DF is obtained at the price of increased processing complexity.

Chapter 3

System Model and Problem Formulation

In this chapter we describe the system model under study, followed by the problem formulation. We consider a multi-user relaying sub-network consisting of L RRHs and K pairs of source and destination UEs, as depicted in Fig. 3.1. It is assumed that each source UE is paired with a *single* destination UE, both modeled as single-antenna nodes due to their limited processing capabilities and low power budgets. By contrast, RRH-l for $l \in \mathcal{L} = \{1, 2, \ldots, L\}$ is equipped with $N_l \geq 1$ antennas. All RRHs are connected to a central node, namely the BBU pool, whose role is to select and activate a proper subset of RRHs and design AF relaying matrices for the active RRHs. Communication is performed in a two-hop half-duplex mode, where a narrowband flat-fading model is assumed for the radio channels between the UEs and RRHs. The CSI is assumed to be known and constant in a given transmission block. In addition, the direct source-destination UE channel links are severely attenuated, thus unavailable.

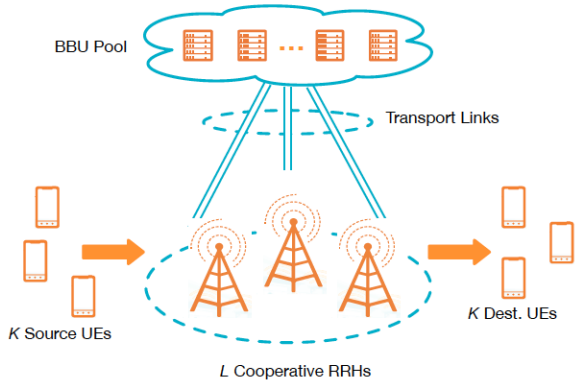

Fig. 3.1: Multi-user sub-network where communication between the source-destination UE pairs is assisted by cooperative MIMO relays via the BBU Pool.

3.1 First hop

During the first hop at the n^{th} time instance, each UE-k for $k \in \mathcal{K} = \{1, 2, \ldots, K\}$ transmits its information symbol $s_k(n)$, modeled as a zero-mean complex random variable with a power of $\mathrm{E}\{|s_k(n)|^2\} = p_k$. Before reaching the l^{th} relay, all transmitted symbols $s_k(n)$ are affected by fading, as represented by multiplication with $\mathbf{h}_{lk} \in \mathbb{C}^{N_l \times 1}$, which is the channel vector between the k^{th} source UE and the l^{th} RRH. The latter receives the sum of the channel-processed information symbols corrupted by additive noise $\mathbf{n}_l(n)$. The received signal at the l-th relay is thus given by

$$\mathbf{r}_l(n) = \sum_{k=1}^{K} \mathbf{h}_{lk} s_k(n) + \mathbf{n}_l(n) \tag{3.1}$$

where $\mathbf{n}_l(n)$ is temporally and spatially white noise, with zero mean and a covariance matrix of $\Sigma_l = \sigma_l^2 \mathbf{I}_{N_l}$.[1] To assist in transmission, RRH-l applies a linear transformation to $\mathbf{r}_l(n)$

[1]Naturally, noise and data are assumed to be statistically uncorrelated, i.e., $\mathrm{E}\{s_k(n)\mathbf{n}_l(n)\} = \mathbf{0}\ \forall k, l$. Further, information symbols between different source UEs are also uncorrelated.

represented by the matrix $\mathbf{B}_l \in \mathbb{C}^{N_l \times N_l}$. The design of $\mathbf{B}_l$ is constrained by an average antenna power budget P_l, expressed as $\mathrm{E}\{\|\mathbf{B}_l(n)\mathbf{r}_l(n)\|_2^2\} \leq P_l$ which is equivalent to

$$\mathrm{Tr}\left(\mathbf{B}_l\big(\sum_{k=1}^{K} p_k \mathbf{h}_{lk}\mathbf{h}_{lk}^H + \Sigma_l\big){\mathbf{B}_l}^H\right) \leq P_l. \tag{3.2}$$

3.2 Second hop

During the second transmission hop, the k^{th} destination UE receives a signal with three components: the desired signal $S_k(n)$, the interference leakage $I_k(n)$, and the temporally white noise $n_k(n)$. We express this as

$$\begin{aligned} d_k(n) &= \sum_{l=1}^{L} \mathbf{g}_{kl}^H \mathbf{B}_l \mathbf{r}_l(n) + n_k(n) \\ &= S_k(n) + I_k(n) + n_k(n), \end{aligned} \tag{3.3}$$

where $\mathbf{g}_{kl}^H \in \mathbb{C}^{1 \times N_l}$ denotes the channel vector between the k^{th} destination UE with the l^{th} RRH, and the signal components are given by

$$S_k(n) = \sum_{l=1}^{L} \mathbf{g}_{kl}^H \mathbf{B}_l \mathbf{h}_{lk} s_k(n) \tag{3.4}$$

$$I_k(n) = \sum_{\substack{j=1,\\ j\neq k}}^{K} \sum_{l=1}^{L} {\mathbf{g}_{kl}}^H \mathbf{B}_l \mathbf{h}_{lj} s_j(n) + \sum_{l=1}^{L} {\mathbf{g}_{kl}}^H \mathbf{B}_l \mathbf{n}_l(n). \tag{3.5}$$

3.3 Problem Formulation

The objective of this work is to enhance the quality of the received signal $d_k(n)$ at each destination UE while considering the power budget, as specified by (3.2). Motivated by the interference alignment techniques [40], an effective means of achieving this objective is

to minimize the *total interference leakage* at all destination UEs, that is $\mathrm{E}\{\sum_{k=1}^{K}|I_k(n)|^2\}$, while enforcing a set of linear constraints meant to preserve the integrity of the desired symbols $s_k(n)$. Based on (3.4), these constraints are given by

$$\sum_{l=1}^{L}\mathbf{g}_{kl}^{H}\mathbf{B}_l\mathbf{h}_{lk}=c_k,\quad \forall k\in\mathcal{K} \tag{3.6}$$

where the c_k's are predefined positive constants. Hence, we state the problem as

$$\min_{\{\mathbf{B}_l\}_{l\in\mathcal{L}}}\ \mathrm{E}\left\{\sum_{k=1}^{K}|I_k(n)|^2\right\}\qquad \text{subject to} \tag{3.7a}$$

$$\mathrm{Tr}\left(\mathbf{B}_l\Big(\sum_{k=1}^{K}p_k\mathbf{h}_{lk}\mathbf{h}_{lk}^{H}+\Sigma_l\Big){\mathbf{B}_l}^{H}\right)\leq P_l,\quad \forall l\in\mathcal{L} \tag{3.7b}$$

$$\sum_{l=1}^{L}\mathbf{g}_{kl}^{H}\mathbf{B}_l\mathbf{h}_{lk}=c_k,\quad \forall k\in\mathcal{K} \tag{3.7c}$$

In the following two subsections, we first express the above into a more convenient formulation, and then we incorporate a measure of group-sparsity into the formulation to encourage the deactivation of a subset of RRHs and thus obtain an energy-efficient solution.

3.3.1 Interference leakage minimization

We begin by replacing each AF matrix by its vectorized version $\mathbf{b}_l=\mathrm{vec}\{\mathbf{B}_l\}$ as obtained by stacking the columns of $\mathbf{B}_l$ one on top of the other. The resulting vectors are collected into a global vector $\mathbf{b}\triangleq[\mathbf{b}_1^T,\cdots,\mathbf{b}_L^T]^T$. To obtain a more convenient form than (3.7), we examine the closed-form expression for the total interference leakage. We first summarize the assumptions made so far.

Assumptions

A1. CSI $\mathbf{g}_{kl}$ and $\mathbf{h}_{lk}$ is known $\forall(k,l) \in \mathcal{K} \times \mathcal{L}$

A2. Information symbols $s_k(n)$: 0-mean, power $\mathrm{E}\{|s_k(n)|^2\} = p_k$

A3. Noise $\mathbf{n}_l(n)$: temporally and spatially white with 0-mean, $\sigma_l^2 \mathbf{I}_{N_l}$ cov. matrix

A4. Noise and data: statistically uncorrelated, $\mathrm{E}\left\{s_k(n)\mathbf{n}_l^H(n)\right\} = \mathbf{0}$, $\forall(k,l) \in \mathcal{K} \times \mathcal{L}$

A5. Information symbols between different UEs: statistically uncorrelated

A6. Noise at different UEs: statistically uncorrelated

The total interference leakage is equal to $\mathrm{E}\{\sum_{k=1}^{K} |I_k(n)|^2\} = \sum_{k=1}^{K} \mathrm{E}\{|I_k(n)|^2\}$. Now, $\mathrm{E}\{|I_k(n)|^2\} = \mathrm{E}\{I_k(n) I_k(n)^*\}$. Omitting the dependence on time n for clarity, the former can be expressed as:

$$\mathrm{E}\left\{\left(\sum_{j\neq k}\sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj} s_j + \sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{n}_l\right).\left(\sum_{j'\neq k}\sum_{l'=1}^{L} \mathbf{g}_{kl'}{}^H \mathbf{B}_{l'} \mathbf{h}_{l'j'} s_{j'} + \sum_{l'=1}^{L} \mathbf{g}_{kl'}{}^H \mathbf{B}_{l'} \mathbf{n}_{l'}\right)^*\right\}$$
$$= \mathrm{E}\left\{\left(\textcircled{1} + \textcircled{2}\right).\left(\textcircled{3} + \textcircled{4}\right)^*\right\}$$
$$= \mathrm{E}\left\{\textcircled{1} \times \textcircled{3}^*\right\} + \mathrm{E}\left\{\textcircled{1} \times \textcircled{4}^*\right\} + \mathrm{E}\left\{\textcircled{2} \times \textcircled{3}^*\right\} + \mathrm{E}\left\{\textcircled{2} \times \textcircled{4}^*\right\}$$

From assumptions A1 and A4, we have $\mathrm{E}\left\{\textcircled{1} \times \textcircled{4}^*\right\} = 0$ and $\mathrm{E}\left\{\textcircled{2} \times \textcircled{3}^*\right\} =$

0. Therefore, we are left with two terms to compute, the first of which is equal to:

$$\begin{aligned}\mathrm{E}\left\{\textcircled{1}\times\textcircled{3}^{*}\right\} &= \mathrm{E}\left\{\sum_{j\neq k}\sum_{l=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{h}_{lj}s_{j}\times\left(\sum_{j'\neq k}\sum_{l'=1}^{L}\mathbf{g}_{kl'}{}^{H}\mathbf{B}_{l'}\mathbf{h}_{l'j'}s_{j'}\right)^{*}\right\}\\ &= \mathrm{E}\left\{\sum_{j\neq k}\sum_{j'\neq k}\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{h}_{lj}\mathbf{h}_{l'j'}^{H}\mathbf{B}_{l'}^{H}\mathbf{g}_{kl'}s_{j}s_{j'}{}^{*}\right\}\end{aligned}$$

From assumptions A1 and A5, we have

$$\begin{aligned}\mathrm{E}\left\{\textcircled{1}\times\textcircled{3}^{*}\right\} &= \sum_{j\neq k}\sum_{j'\neq k}\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{h}_{lj}\mathbf{h}_{l'j'}^{H}\mathbf{B}_{l'}^{H}\mathbf{g}_{kl'}\,\mathrm{E}\left\{s_{j}s_{j'}{}^{*}\right\}\\ &= \sum_{j\neq k}\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{h}_{lj}\mathbf{h}_{l'j}^{H}\mathbf{B}_{l'}^{H}\mathbf{g}_{kl'}p_{j}\\ &= \sum_{j\neq k}p_{j}\left|\sum_{l=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{h}_{lj}\right|^{2}\end{aligned}$$

As for the second term, A1 and A6 give

$$\begin{aligned}\mathrm{E}\left\{\textcircled{2}\times\textcircled{4}^{*}\right\} &= \mathrm{E}\left\{\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{n}_{l}\mathbf{n}_{l'}^{H}\mathbf{B}_{l'}^{H}\mathbf{g}_{kl'}\right\}\\ &= \sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\,\mathrm{E}\left\{\mathbf{n}_{l}\mathbf{n}_{l'}^{H}\right\}\mathbf{B}_{l'}^{H}\mathbf{g}_{kl'}\\ &= \sum_{l=1}^{L}\sigma_{l}^{2}\mathbf{g}_{kl}{}^{H}\mathbf{B}_{l}\mathbf{B}_{l}^{H}\mathbf{g}_{kl}\\ &= \sum_{l=1}^{L}\sigma_{l}^{2}\left\|\mathbf{B}_{l}^{H}\mathbf{g}_{kl}\right\|_{2}^{2}\end{aligned}$$

The total interference leakage may be thus given by

$$\mathrm{E}\{\sum_{k=1}^{K}|I_k(n)|^2\} = \sum_{k=1}^{K}\left(\underbrace{\sum_{j\neq k} p_j \left|\sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj}\right|^2}_{\text{power of } j^{th} \text{ interferer}} + \sum_{l=1}^{L} \underbrace{\sigma_l^2 \left\|\mathbf{B}_l^H \mathbf{g}_{kl}\right\|_2^2}_{\text{power of } l^{th} \text{ relay's forwarded noise}}\right) \tag{3.8}$$

From the Kronecker product property $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, the first term inside the summation in (3.8) can be expressed as

$$\sum_{j\neq k} p_j \left|\sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj}\right|^2 = \sum_{j\neq k} p_j \left|\sum_{l=1}^{L} \mathrm{vec}(\mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj})\right|^2 = \sum_{j\neq k} p_j \left|\sum_{l=1}^{L} (\mathbf{h}_{lj}^* \otimes \mathbf{g}_{kl})^H \mathrm{vec}(\mathbf{B}_l)\right|^2.$$

Defining $\boldsymbol{\delta}_k^{(j)} \triangleq [(\mathbf{h}_{1j}^* \otimes \mathbf{g}_{k1})^T \ (\mathbf{h}_{2j}^* \otimes \mathbf{g}_{k2})^T \ \ldots \ (\mathbf{h}_{Lj}^* \otimes \mathbf{g}_{kL})^T]^T \in \mathbb{C}^{\sum_l N_l^2 \times 1}$, we find that the above is equal to

$$\sum_{j\neq k} p_j \left|\boldsymbol{\delta}_k^{(j)H} \mathbf{b}\right|^2 = \sum_{j\neq k} p_j \left|\mathbf{b}^H \boldsymbol{\delta}_k^{(j)}\right|^2 = \mathbf{b}^H \boldsymbol{\Delta}_k \mathbf{b},$$

with $\boldsymbol{\Delta}_k \triangleq \sum_{j\neq k} p_j \boldsymbol{\delta}_k^{(j)} \boldsymbol{\delta}_k^{(j)H}$. The second term inside the summation (3.8) can be written as

$$\sum_{l=1}^{L} \sigma_l^2 \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{I}_{N_l} \mathbf{B}_l^H \mathbf{g}_{kl} = \sum_{l=1}^{L} \sigma_l^2 \mathbf{g}_{kl}{}^H \mathbf{B}_l \left(\sum_{i=1}^{N_l} \mathbf{e}_i \mathbf{e}_i^H\right) \mathbf{B}_l^H \mathbf{g}_{kl} = \sum_{l=1}^{L} \sum_{i=1}^{N_l} \sigma_l^2 \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{e}_i \mathbf{e}_i^H \mathbf{B}_l^H \mathbf{g}_{kl}$$

where $\mathbf{e}_i$ is the 'basis' vector of length N_l composed of 1 at index i and of 0's elsewhere.

Applying the same Kronecker product property as before, the above is expressed as

$$\begin{aligned}\sum_{l=1}^{L}\sigma_l^2{\mathbf{g}_{kl}}^H\mathbf{B}_l\mathbf{I}_{N_l}\mathbf{B}_l^H\mathbf{g}_{kl} &= \sum_{l=1}^{L}\sum_{i=1}^{N_l}\sigma_l^2\Big((\mathbf{e}_i\otimes\mathbf{g}_{kl})^H\mathbf{b}_l\Big)\Big((\mathbf{e}_i\otimes\mathbf{g}_{kl})^H\mathbf{b}_l\Big)^*\\ &= \sum_{l=1}^{L}\sum_{i=1}^{N_l}\sigma_l^2\Big|(\mathbf{e}_i\otimes\mathbf{g}_{kl})^H\mathbf{b}_l\Big|^2\\ &= \sum_{l=1}^{L}\sigma_l^2\mathbf{b}_l^H\sum_{i=1}^{N_l}(\mathbf{e}_i\otimes\mathbf{g}_{kl})(\mathbf{e}_i\otimes\mathbf{g}_{kl})^H\mathbf{b}_l\\ &= \sum_{l=1}^{L}\mathbf{b}_l^H\mathbf{G}_{kl}\mathbf{b}_l\\ &= \mathbf{b}^H\mathbf{G}_k\mathbf{b}\end{aligned}$$

with $\mathbf{G}_{kl}\triangleq\sigma_l^2\sum_{i=1}^{N_l}(\mathbf{e}_i\otimes\mathbf{g}_{kl})(\mathbf{e}_i\otimes\mathbf{g}_{kl})^H$ and $\mathbf{G}_k\triangleq\text{blkdiag}\{\mathbf{G}_{k1},\mathbf{G}_{k2},\ldots,\mathbf{G}_{kL}\}$. Combining the two results, the total interference leakage reduces to a quadratic term in $\mathbf{b}$:

$$\sum_{k=1}^{K}\mathrm{E}\left\{|I_k(n)|^2\right\}=\sum_{k=1}^{K}\left(\mathbf{b}^H\mathbf{\Delta}_k\mathbf{b}+\mathbf{b}^H\mathbf{G}_k\mathbf{b}\right)=\mathbf{b}^H\mathbf{\Theta}\mathbf{b}$$

where $\mathbf{\Theta}\triangleq\sum_{k=1}^{K}\mathbf{\Delta}_k+\mathbf{G}_k$.

Finally, we apply the Kronecker property $\text{Tr}\left(\mathbf{ABA}^H\right)=\text{vec}(\mathbf{A})^H(\mathbf{B}\otimes\mathbf{I})\,\text{vec}(\mathbf{A})$ to (3.7b) to obtain the compact formulation

$$\min_{\mathbf{b}}\ \mathbf{b}^H\mathbf{\Theta}\mathbf{b} \tag{3.9a}$$

$$\text{s.t.}\quad {\mathbf{b}_l}^H\mathbf{\Psi}_l\mathbf{b}_l\le P_l,\quad\forall l\in\mathcal{L} \tag{3.9b}$$

$$\mathbf{\Phi}^H\mathbf{b}=\mathbf{c}, \tag{3.9c}$$

where

$$\mathbf{\Psi}_l \triangleq \Big(\sum_{k=1}^{K} p_k \mathbf{h}_{lk} \mathbf{h}_{lk}^H + \Sigma_l \Big) \otimes \mathbf{I}_{N_l}, \tag{3.10}$$

$$\mathbf{c} \triangleq [c_1, c_2, \cdots, c_K]^T, \tag{3.11}$$

and the matrix

$$\mathbf{\Phi} = [\boldsymbol{\phi}_{l,k}] \in \mathbb{C}^{\sum_{l=1}^{L} N_l^2 \times K} \tag{3.12}$$

is partitioned into $L \times K$ blocks, each given by $\boldsymbol{\phi}_{l,k} = \mathbf{h}_{lk}^* \otimes \mathbf{g}_{kl}$.

We emphasize that at this stage of the formulation, nothing prevents any of the RRHs from participating in the relay-assisted transmission, potentially leading to a situation where all RRHs are activated, i.e., $\|\mathbf{b}_l\|_2^2 > 0 \ \forall l \in \mathcal{L}$. In the following subsection, we modify (3.9) by adding a regularization term to reduce the number of active RRHs while meeting an acceptable QoS level.

3.3.2 Relay Selection via Group-Sparsity

Whenever RRH-l is inactive, we have $\|\mathbf{b}_l\|_2 = 0$. Consequently, having a small set of active RRHs implies that the solution vector $\mathbf{b}$ is group-sparse, or equivalently, the vector $\mathcal{B} \triangleq \big[\|\mathbf{b}_1\|_2, \cdots, \|\mathbf{b}_L\|_2\big]^T$ containing the norms of the L RRHs consists of a small number of non-zero elements. This property (sparsity) is captured by the cardinality operator denoted by $\|\cdot\|_0$.

Since our aim is to obtain an energy efficient solution, a group-sparse solution is desired. A straightforward way to promote this property during optimization is to include in the objective function (3.9a) a penalty term based on $\|\mathcal{B}\|_0$. However, this renders the objective function non-convex and non-smooth. To resolve this issue, the $\|\cdot\|_0$ operator is replaced

by its closest convex relaxation, namely the l_1-norm, which in the case of $\mathcal{B}$ is

$$\left\|\mathcal{B}\right\|_1 = \sum_{l=1}^{L} \left\|\mathbf{b}_l\right\|_2, \tag{3.13}$$

and is termed as the mixed $l_{1,2}$-norm. This approach is widely employed for enforcing group-sparsity and was first introduced by [41]. Using (3.13), we may penalize the objective function (3.9a) by the regularization term $\lambda \sum_{l=1}^{L} \|\mathbf{b}_l\|_2$ with $\lambda > 0$ being a tuning parameter that controls the degree of group-sparisty. A different weight λ_l may also be used for each RRH-l.

To simplify the subsequent analysis, we define $\mathbf{\Psi} \triangleq \texttt{blkdiag}\{\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_L\}$ and $\mathbf{x} \triangleq \mathbf{\Psi}^{1/2}\mathbf{b} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_L^T]^T$ with $\mathbf{x}_l = \mathbf{\Psi}_l^{1/2}\mathbf{b}_l$. The sample covariance matrix $\mathbf{\Psi}_l$ in (3.10) is non-singular, which implies that whether $\|\mathbf{x}_l\|_2 > 0$ or $\|\mathbf{x}_l\|_2 = 0$ depends entirely on whether $\|\mathbf{b}_l\|_2 > 0$ or $\|\mathbf{b}_l\|_2 = 0$. So both vectors $\mathbf{x}$ and $\mathbf{b}$ share the same group-sparsity structure. Based on this observation, (3.9) may be replaced by the following regularized problem

$$\boxed{\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{x}^H \breve{\mathbf{\Theta}} \mathbf{x} + \sum_{l=1}^{L} \lambda_l \|\mathbf{x}_l\|_2 \\ \text{s.t.} \ & \mathbf{x}_l^H \mathbf{x}_l \leq P_l, \quad \forall l \in \mathcal{L} \\ & \breve{\mathbf{\Phi}}^H \mathbf{x} = \mathbf{c}, \end{aligned}} \tag{3.14}$$

where $\breve{\mathbf{\Theta}} = \mathbf{\Psi}^{-\frac{1}{2}} \mathbf{\Theta} \mathbf{\Psi}^{-\frac{1}{2}}$ and $\breve{\mathbf{\Phi}} = \mathbf{\Psi}^{-\frac{1}{2}} \mathbf{\Phi}$.

The above may be transformed into a second-order cone program (SOCP) using the techniques introduced in [42, Section 2.2] and can thus be solved to global optimality using a standard optimization package via the interior point method [31].

Chapter 4

Proposed Solution

In what follows, we exploit the structure of the norm-regularized relaying optimization problem and develop an algorithm for solving it using the ADMM. We show that each step of the algorithm admits a closed-form solution, which can significantly reduce its computational complexity.

4.1 ADMM-based Low-Complexity Algorithm

The problem in (3.14) may be written in a form amenable to the ADMM. By first introducing a copy of $\mathbf{x}$, namely $\mathbf{z}$, via the linear constraint

$$\mathbf{x} = \mathbf{z}, \tag{4.1}$$

and by constructing the constraint sets

$$\mathcal{C}_1 : \breve{\mathbf{\Phi}}^H \mathbf{x} = \mathbf{c} \tag{4.2}$$

$$\mathcal{C}_2 : \mathbf{z}_l^H \mathbf{z}_l \leq P_l, \quad \forall l \in \mathcal{L}, \tag{4.3}$$

(3.14) may be re-expressed as

$$\min_{\mathbf{x},\mathbf{z}} \ \mathbf{x}^H \breve{\mathbf{\Theta}} \mathbf{x} + \sum_{l=1}^{L} \lambda_l \|\mathbf{z}_l\|_2 \tag{4.4a}$$

$$\text{s.t.} \ \ \mathbf{x} \in \mathcal{C}_1, \ \ \mathbf{z} \in \mathcal{C}_2 \tag{4.4b}$$

$$\mathbf{x} = \mathbf{z} \tag{4.4c}$$

The ADMM algorithm aims to iteratively minimize the augmented Lagrangian $\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y})$ given by

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = \mathbf{x}^H \breve{\mathbf{\Theta}} \mathbf{x} + \frac{\rho}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 + \sum_{l=1}^{L} \lambda_l \|\mathbf{z}_l\|_2 - \big(\mathbf{z} - \mathbf{x}\big)^H \mathbf{y} - \mathbf{y}^H \big(\mathbf{z} - \mathbf{x}\big), \tag{4.5}$$

where $\mathbf{y}$ denotes the Lagrange multiplier associated with (4.1) and $\rho > 0$ which remains constant during the ADMM iterations. We note that an optimal solution $(\mathbf{x}^\star, \mathbf{z}^\star, \mathbf{y}^\star)$ that minimizes $\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y})$ must satisfy $\mathbf{x}^\star = \mathbf{z}^\star$ since $\mathbf{z}$ is a synthesized copy of $\mathbf{x}$. Therefore, solving for (4.4) becomes equivalent to solving the following problem,

$$\min_{\mathbf{x},\mathbf{z}} \ \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) \tag{4.6a}$$

$$\text{s.t.} \ \ \mathbf{x} \in \mathcal{C}_1, \ \ \mathbf{z} \in \mathcal{C}_2 \tag{4.6b}$$

$$\mathbf{x} = \mathbf{z} \tag{4.6c}$$

The basic idea behind ADMM is to solve the above problem with respect to $\mathbf{x}$ and $\mathbf{z}$ separately in an alternating manner, i.e., one variable at a time with the other fixed. After each round of update of $\mathbf{x}$ and $\mathbf{z}$, the dual variable $\mathbf{y}$ is updated to ensure that $\mathbf{x}$ and $\mathbf{z}$ become closer to each other. In effect, the above optimization problem can now be

decoupled into three separate steps, all of which, interestingly, admit a simple closed-form solution, as detailed below. For notational simplicity, we temporarily drop the ADMM iteration index for each variable.

4.1.1 x update

The subproblem solving for $\mathbf{x}$ can be expressed as $\min_{\mathbf{x}\in\mathcal{C}_1} \mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{y})$, with $\mathbf{z}$ and $\mathbf{y}$ fixed. Using (4.5), solving for $\mathbf{x}$ can be expressed as the following linearly-constrained quadratic program after neglecting all terms that are independent of $\mathbf{x}$:

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \ \mathbf{x}^H \breve{\mathbf{\Theta}} \mathbf{x} + \frac{\rho}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + \mathbf{y}^H\mathbf{x} + \mathbf{x}^H\mathbf{y} \tag{4.7a}$$

$$\text{s.t.} \quad \breve{\mathbf{\Phi}}^H \mathbf{x} = \mathbf{c}. \tag{4.7b}$$

It is observed that the objective function (4.7a) is strictly convex in $\mathbf{x}$ and the Slater's constraint qualification holds, i.e., (4.7) is strictly feasible. Hence, the KKT sufficient conditions hold for the optimal solution $\mathbf{x}^\star$ together with some optimal dual variable $\boldsymbol{\nu}^\star$. Defining $\mathbf{Q} \triangleq \breve{\mathbf{\Theta}} + \frac{\rho}{2}\mathbf{I}_{\sum N_l^2}$, the augmented Lagrangian associated with the equality constraint (4.7b) is

$$\mathcal{L}_\rho(\mathbf{x},\boldsymbol{\nu}) = \mathbf{x}^H\mathbf{Q}\mathbf{x} + \mathbf{x}^H\left(\mathbf{y}-\frac{\rho}{2}\mathbf{z}\right) + \left(\mathbf{y}-\frac{\rho}{2}\mathbf{z}\right)^H\mathbf{x} - \boldsymbol{\nu}^H(\breve{\mathbf{\Phi}}^H\mathbf{x}-\mathbf{c}) - (\mathbf{x}^H\breve{\mathbf{\Phi}}-\mathbf{c}^H)\boldsymbol{\nu} \tag{4.8}$$

From the KKT conditions, the optimal points $(\mathbf{x}^\star,\boldsymbol{\nu}^\star)$ are such that the constraint (4.7b) is satisfied and the partial derivative of the Lagrangian (w.r.t. $\mathbf{x}^H$) is equal to zero, i.e.,

$$\nabla\mathcal{L}_\rho(\mathbf{x}^\star,\boldsymbol{\nu}^\star) = 0 \tag{4.9}$$

$$\mathbf{Q}\mathbf{x}^\star + \mathbf{y} - \frac{\rho}{2}\mathbf{z} - \breve{\mathbf{\Phi}}\boldsymbol{\nu}^\star = 0, \tag{4.10}$$

yielding

$$\mathbf{Q}\,\mathbf{x}^{\star} = \frac{\rho}{2}\mathbf{z} - \mathbf{y} + \breve{\mathbf{\Phi}}\boldsymbol{\nu}^{\star} \tag{4.11}$$

$$\breve{\mathbf{\Phi}}^{H}\,\mathbf{x}^{\star} = \mathbf{c} \tag{4.12}$$

Re-arranging (4.11), we obtain

$$\mathbf{x}^{\star} = \mathbf{Q}^{-1}\left(\frac{\rho}{2}\mathbf{z} - \mathbf{y} + \breve{\mathbf{\Phi}}\boldsymbol{\nu}^{\star}\right). \tag{4.13}$$

To determine the value of the optimal Lagrange multiplier $\boldsymbol{\nu}^{\star}$, we use (4.13) in equation (4.12). After some matrix manipulations, $\boldsymbol{\nu}^{\star}$ is given by

$$\boldsymbol{\nu}^{\star} = \breve{\mathbf{Q}}^{-1}\left(\mathbf{c} - \breve{\mathbf{\Phi}}^{H}\mathbf{Q}^{-1}(\frac{\rho}{2}\mathbf{z} - \mathbf{y})\right) \tag{4.14}$$

where $\breve{\mathbf{Q}} \triangleq \breve{\mathbf{\Phi}}^{H}\mathbf{Q}^{-1}\breve{\mathbf{\Phi}}$. Then, substituting (4.14) back into (4.13), the following closed-form solution is obtained

$$\mathbf{x}^{\star} = \mathbf{Q}^{-1}\Big(\big(\mathbf{I} - \breve{\mathbf{\Phi}}\breve{\mathbf{Q}}^{-1}\breve{\mathbf{\Phi}}^{H}\mathbf{Q}^{-1}\big)\big(\frac{\rho}{2}\mathbf{z} - \mathbf{y}\big) + \breve{\mathbf{\Phi}}\breve{\mathbf{Q}}^{-1}\mathbf{c}\Big). \tag{4.15}$$

4.1.2 z update

Similarly, the subproblem of solving for $\mathbf{z}$ can be written as $\min_{\mathbf{z}\in\mathcal{C}_2}\mathcal{L}_{\rho}\left(\mathbf{x},\mathbf{z},\mathbf{y}\right)$ with $\mathbf{x}$ and $\mathbf{y}$ fixed. Observing that the first term in (4.5) is independent of $\mathbf{z}$, we write

$$\mathbf{z}^{\star} = \arg\min_{\mathbf{z}\in\mathcal{C}_2}\sum_{l=1}^{L}\lambda_l\|\mathbf{z}_l\|_2 + \frac{\rho}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 - \mathbf{y}^{H}\mathbf{z} - \mathbf{z}^{H}\mathbf{y}. \tag{4.16}$$

Decoupling (4.16) over each $\mathbf{z}_l$, we obtain L parallel subproblems each expressed by

$$\mathbf{z}_l^\star = \arg \min_{\|\mathbf{z}_l\|_2^2 \leq P_l} \lambda_l \|\mathbf{z}_l\|_2 + \frac{\rho}{2}\|\mathbf{z}_l - \mathbf{x}_l\|_2^2 - \mathbf{y}_l^H \mathbf{z}_l - \mathbf{z}_l^H \mathbf{y}_l, \tag{4.17}$$

where the original Lagrange multiplier may be decomposed into L such multipliers, i.e., $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_L^T]^T$. Problem (4.17) is convex and hence solvable by a standard optimization package but our objective is to find a closed-form solution. First, we derive the following lemma:

Lemma 1. *Consider the convex minimization problem in the variable* $\mathbf{x}$*:*

$$\min_{\|\mathbf{x}\|_2^2 \leq P} \lambda \|\mathbf{x}\|_2 + \frac{\rho}{2}\mathbf{x}^H\mathbf{x} - \mathbf{x}^H\mathbf{a} - \mathbf{a}^H\mathbf{x} \tag{4.18}$$

The global minimizer $\mathbf{x}^\star$ *admits a closed-form solution and is given by*

$$\mathbf{x}^\star = \frac{\mathbf{a}}{\|\mathbf{a}\|_2 \left(\frac{\rho}{2} + \eta^\star\right)} [\|\mathbf{a}\|_2 - \lambda]_+ \tag{4.19}$$

where $[c]_+ \triangleq \max\{0, c\}$ *is the soft-thresholding operation and the optimal dual variable* $\eta^\star$ *associated with the quadratic constraint* $\|\mathbf{x}\|_2^2 \leq P$ *is given by*

$$\eta^\star = \left[\frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2}\right]_+ \tag{4.20}$$

Proof: Consider the minimization problem in the variable $\mathbf{x}$:

$$\min_{\|\mathbf{x}\|_2^2 \leq P} f(\mathbf{x}) = \lambda \|\mathbf{x}\|_2 + \frac{\rho}{2}\mathbf{x}^H\mathbf{x} - \mathbf{x}^H\mathbf{a} - \mathbf{a}^H\mathbf{x} \tag{4.21}$$

where $\lambda > 0$, $\rho > 0$, and $\mathbf{a} \neq \mathbf{0}$. The problem is convex and strictly feasible so by Slater's

condition, strong duality holds. As a result, if $\mathbf{x}^\star$ and $\mu^\star$ are optimal primal and dual solutions, they must satisfy the KKT conditions:

$$\text{Primal and dual feasibility:} \quad \|\mathbf{x}^\star\|_2^2 \leq P, \quad \mu^\star \geq 0 \tag{4.22}$$

$$\text{Complementary slacknesss:} \quad \mu^\star(\|\mathbf{x}^\star\|_2^2 - P) = 0 \tag{4.23}$$

$$\begin{aligned} \text{Stationarity:} \quad & \mathbf{0} \in \partial\left(f(\mathbf{x}^\star) + \mu^\star(\|\mathbf{x}^\star\|_2^2 - P)\right) \\ \iff & \mathbf{0} \in \partial h(\mathbf{x}^\star) + \frac{\rho}{2}\mathbf{x}^\star - \mathbf{a} + \mu^\star\mathbf{x}^\star \\ \iff & \partial h(\mathbf{x}^\star) = -\frac{\rho}{2}\mathbf{x}^\star + \mathbf{a} - \mu^\star\mathbf{x}^\star \end{aligned} \tag{4.24}$$

where $\partial h(\mathbf{x}^\star)$ is the subdifferential of $h(\mathbf{x})$ at $\mathbf{x}^\star$ and $h(\mathbf{x}) = \lambda\|\mathbf{x}\|_2$. The l_2-norm is convex and $\lambda > 0$, so $\partial h(\mathbf{x}) = \lambda\partial\left(\|\mathbf{x}\|_2\right)$ with

$$\partial(\|\mathbf{x}\|_2) = \begin{cases} \left\{\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right\} & \text{if } \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{g} \mid \|\mathbf{g}\|_2 \leq 1\} & \text{if } \mathbf{x} = \mathbf{0}. \end{cases} \tag{4.25}$$

To find the optimal solutions we distinguish between two cases: (i) $\|\mathbf{a}\|_2 - \lambda > 0$, and (ii) $\|\mathbf{a}\|_2 - \lambda \leq 0$.

If $\|\mathbf{a}\|_2 - \lambda > 0$, we show by contradiction that $\mathbf{x}^\star \neq \mathbf{0}$. Assuming $\mathbf{x}^\star = \mathbf{0}$, we find from (4.24) that $\lambda\mathbf{g} = \mathbf{a}$ with $\|\mathbf{g}\|_2 \leq 1$. This implies that $1 - \frac{\|\mathbf{a}\|_2}{\lambda} \geq 0$ or $\|\mathbf{a}\|_2 - \lambda \leq 0$, which is a contradiction. Since $\mathbf{x}^\star \neq \mathbf{0}$, the stationarity condition yields

$$\left(\frac{\lambda}{\|\mathbf{x}^\star\|_2} + \frac{\rho}{2} + \mu^\star\right)\mathbf{x}^\star = \mathbf{a}. \tag{4.26}$$

In order for (4.26) to hold, $\mathbf{x}^\star$ must take the form of $\mathbf{x}^\star = z\mathbf{a}$ for some $z > 0$. Using this

observation in (4.26) and re-arranging, we find that

$$z = \frac{\|\mathbf{a}\|_2 - \lambda}{\|\mathbf{a}\|_2 \left(\frac{\rho}{2} + \mu^\star\right)} > 0. \tag{4.27}$$

Thus, $\mathbf{x}^\star \neq \mathbf{0}$ in turn implies that $\|\mathbf{a}\|_2 - \lambda > 0$ and the two statements are equivalent (since one implies the other and vice versa). The optimal dual variable $\mu^\star$ must satisfy the remaining KKT conditions. Primal feasibility gives

$$\mu^\star \geq \frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2}, \tag{4.28}$$

dual feasibility and complementary slackness imply

$$\mu^\star = \begin{cases} \frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2} & \text{if } \frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2} > 0 \\ 0 & \text{if } \frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2} \leq 0 \end{cases}. \tag{4.29}$$

Now if $\|\mathbf{a}\|_2 - \lambda \leq 0$, then $\mathbf{x}^\star = \mathbf{0}$ and these conditions satisfy stationarity and primal feasibility, respectively. Taking $\mu^\star = 0$ ensures complementary slackness and dual feasibility.

In conclusion, the optimal variables can be succinctly expressed by

$$\mathbf{x}^\star = \frac{\mathbf{a}}{\|\mathbf{a}\|_2 \left(\frac{\rho}{2} + \mu^\star\right)} [\|\mathbf{a}\|_2 - \lambda]_+ \tag{4.30}$$

$$\mu^\star = \left[\frac{\|\mathbf{a}\|_2 - \lambda}{\sqrt{P}} - \frac{\rho}{2}\right]_+. \tag{4.31}$$

where $[c]_+ \triangleq \max\{0, c\}$. This completes the proof. ■

As a direct result of Lemma 1, the optimal solution to (4.17) is given by

$$\mathbf{z}_l^\star = \frac{\mathbf{a}_l}{\|\mathbf{a}_l\|_2 \left(\frac{\rho}{2} + \eta_l\right)} [\|\mathbf{a}_l\|_2 - \lambda_l]_+ \tag{4.32}$$

where η_l is given by

$$\eta_l = \left[\frac{\|\mathbf{a}_l\|_2 - \lambda_l}{\sqrt{P_l}} - \frac{\rho}{2} \right]_+ \tag{4.33}$$

and $\mathbf{a}_l = \frac{\rho}{2}\mathbf{x}_l + \mathbf{y}_l$.

In summary, both $\mathbf{x}$ and $\mathbf{z}$ are obtained in closed-form at each iteration with the aid of (4.15) and (4.32). In addition, the $\mathbf{z}$-minimization step can be carried out in a parallel fashion. The ADMM-based algorithm is now summarized in Algorithm 1, where the primal and dual residuals are defined as follows,

$$\mathbf{r}^{(j+1)} = \mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)} \tag{4.34}$$

$$\mathbf{s}^{(j+1)} = -\frac{\rho}{2}(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)}). \tag{4.35}$$

4.2 An improved Two-Stage ADMM implementation

The ADMM-based algorithm is capable of selecting a subset of active RRHs. However, similar to the LASSO problem in compressive sensing literature, the addition of the norm-based regularization term in the objective function may lead to worse relaying performance, which can be improved by solving for the optimal relaying AF matrices one more time for

Algorithm 1 ADMM for solving (3.14)

1: **Initialization**: primal variable $\mathbf{z}^{(0)}$ (arbitrary non-zero vector); dual variable $\mathbf{y}^{(0)} = \mathbf{0}$; set ADMM iteration index $j = 0$;
2: **repeat**
3: Update $\mathbf{x}^{(j+1)}$ using (4.15)
4: Update $\mathbf{z}_l^{(j+1)}$ for all $l \in \mathcal{L}$ using (4.32), where $\mathbf{a}_l^{(j+1)} = \frac{\rho}{2}\mathbf{x}_l^{(j+1)} + \mathbf{y}_l^{(j)}$
5: Update the Lagrange multiplier

$$\mathbf{y}^{(j+1)} = \mathbf{y}^{(j)} + \frac{\rho}{2}\left(\mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)}\right)$$

6: $j \leftarrow j + 1$;
7: **until** $\|\mathbf{r}^{(j+1)}\|_2 \leq \epsilon^{pri}$ and $\|\mathbf{s}^{(j+1)}\|_2 \leq \epsilon^{dual}$

those active RRHs selected from the previous step. This problem can be formulated as

$$\min_{\breve{\mathbf{x}}} \ \breve{\mathbf{x}}^H \breve{\mathbf{\Theta}}_{\mathbf{R}} \breve{\mathbf{x}} \tag{4.36a}$$

$$\text{s.t. } \breve{\mathbf{x}}_l{}^H \breve{\mathbf{x}}_l \leq P_l, \ \forall l \in \mathcal{A} \tag{4.36b}$$

$$\breve{\mathbf{\Phi}}_{\mathbf{R}}^H \breve{\mathbf{x}} = \mathbf{c} \tag{4.36c}$$

where $\breve{\mathbf{x}}$ now only consists of weights from active RRHs and $\mathcal{A}$ denotes the subset of active RRHs. The latter can be determined from the outputs $\mathbf{z}_l$ of Algorithm 1, i.e.,

$$\mathcal{A} = \{l \in \mathcal{L} : \|\mathbf{z}_l\|_2 > 0\}. \tag{4.37}$$

Note that $\breve{\mathbf{\Theta}}_{\mathbf{R}}$ and $\breve{\mathbf{\Phi}}_{\mathbf{R}}$ are reduced versions of the original matrices given in (4.2) and (4.4), where elements related to the inactive RRHs are deleted. The two-stage ADMM implementation is now summarized as follows:

1. Solve (3.14) using Algorithm 1 and select the subset of active RRHs $\mathcal{A}$ based on (4.37);

2. Solve (4.36) for active RRHs using ADMM, as described in Algorithm 2.

We note that the development of Algorithm 2 parallels the development of Algorithm 1. Indeed, the former is obtained by setting $\lambda = 0$ in the substeps of ADMM.

Algorithm 2 ADMM for solving (4.36)

1: **Initialization**: primal variable $\mathbf{z}^{(0)}$ initialized from the output of Algorithm 1; dual variable $\mathbf{y}^{(0)} = \mathbf{0}$; set ADMM iteration index $j = 0$;
2: **repeat**
3: Update $\mathbf{x}^{(j+1)}$

$$\mathbf{x} = \mathbf{Q}^{-1}\Big(\big(\mathbf{I} - \breve{\mathbf{\Phi}}\breve{\mathbf{Q}}^{-1}\breve{\mathbf{\Phi}}^{H}\mathbf{Q}^{-1}\big)\big(\frac{\rho}{2}\mathbf{z}^{(j)} - \mathbf{y}^{(j)}\big) + \breve{\mathbf{\Phi}}\breve{\mathbf{Q}}^{-1}\mathbf{c}\Big)$$

4: Update $\mathbf{z}_l^{(j+1)}$ for all $l \in \mathcal{A}$

$$\mathbf{a}_l^{(j+1)} = \frac{\rho}{2}\,\mathbf{x}_l^{(j+1)} + \mathbf{y}_l^{(j)}$$

$$\eta_l^{(j+1)} = \max(0, \frac{\|\mathbf{a}_l^{(j+1)}\|_2}{\sqrt{P}} - \frac{\rho}{2})$$

$$\mathbf{z}_l^{(j+1)} = \frac{\mathbf{a}_l^{(j+1)}}{\frac{\rho}{2} + \eta_l}, \quad \text{for} l \in \mathcal{A}$$

5: Update the Lagrange multiplier

$$\mathbf{y}^{(j+1)} = \mathbf{y}^{(j)} + \frac{\rho}{2}\left(\mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)}\right)$$

6: $j \leftarrow j + 1$;
7: **until** $\|\mathbf{r}^{(j+1)}\|_2 \leq \epsilon^{pri}$ and $\|\mathbf{s}^{(j+1)}\|_2 \leq \epsilon^{dual}$

Chapter 5

Simulation Results and Discussion

5.1 System and Model Parameters

In all our simulations, we consider a relaying sub-network consisting of $K = 6$ source-destination pairs and $L = 6$ AF RRHs. For simplicity, we use the same number of antennas and power budget for all RRHs. Specifically, we choose $N_l = 6$ antennas and $P_l = 2W$ for all l. The distortionless constraints are set to unity, i.e., $c_k = 1 \quad \forall k \in \mathcal{K}$ and the source symbols s_k are generated from a 4-QAM constellation with a power $p_k = 1, \quad \forall k \in \mathcal{K}$. We adopt a narrowband flat-fading Rayleigh channel model with coefficients generated as independent and identically distributed (i.i.d.) zero-mean complex circular Gaussian variables with unit variance. The noise variances at the RRHs, σ_l^2, are set according to the desired input relay signal-to-noise ratios (SNR) defined as

$$\gamma_l = \frac{P_l}{\sigma_l^2 \times N_l} \tag{5.1}$$

Finally, the noise variance at all the destination nodes is set to 10^{-1}.

As for the ADMM termination criterion, we adopt the scheme in [30], where the primal

and dual residual tolerances are given by

$$\epsilon^{pri} = \sqrt{\bar{N}}\epsilon^{abs} + \epsilon^{rel}\max\{\|\mathbf{x}^{(j+1)}\|_2, \|\mathbf{z}^{(j+1)}\|_2\} \tag{5.2}$$

$$\epsilon^{dual} = \sqrt{\bar{N}}\epsilon^{abs} + \epsilon^{rel}\{\|\mathbf{y}^{(j+1)}\|_2\} \tag{5.3}$$

where ϵ^{abs} is an absolute tolerance and ϵ^{rel} is a relative tolerance and $\bar{N} = \sum_{l=1}^{L} N_l^2$.

Finally, all the simulation results are averaged over 100 independent realizations.

5.2 Results and Discussions

5.2.1 Convergence behaviour of Algorithm 1

Figures 5.1, 5.2 demonstrate the convergence behaviour of Algorithm 1 for two specific channel realizations. It can be observed from the top-left figures that in both cases the algorithm converges within a small number of iterations with the adopted scheme for the residual tolerances (above). The top-right graphs of Figure 5.1 and 5.2 display the norm of the relaying weights, i.e., $\|\mathbf{z}_l\|_2$. In Figure 5.1, the norm for one RRH goes to zero and in Figure 5.2, the norm for two RRHs goes to zero. In effect, this means that for these specific realizations, the algorithm yields a subset of five active RRHs and four active RRHs, respectively. In addition, we can observe from the bottom-left and bottom-right graphs in both figures that the algorithm indeed converges, as the residual error between $\mathbf{x}$ and $\mathbf{z}$ approaches zero and the obtained solution for $\mathbf{x}$ satisfies the linear distortionless constraints in (4.2).

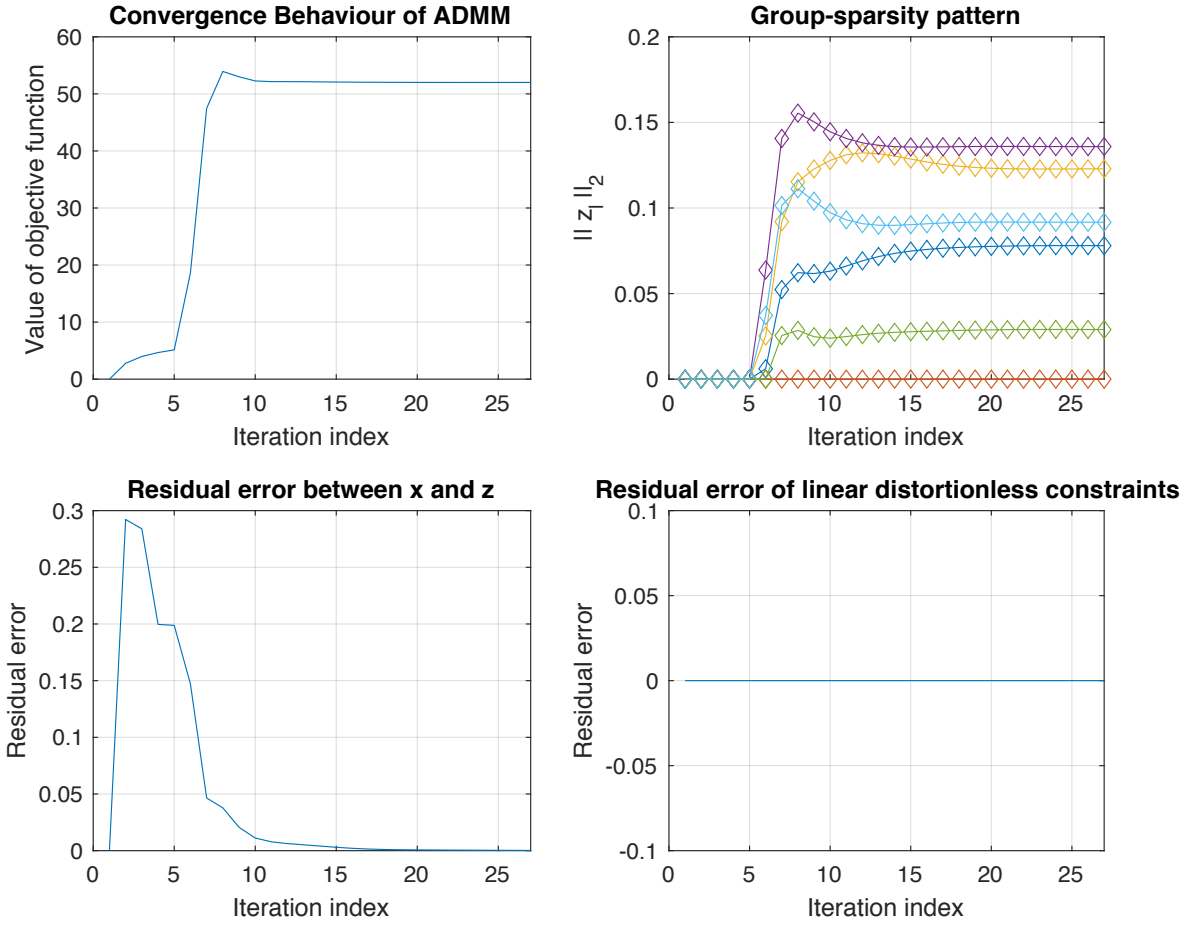

Fig. 5.1: Convergence behavior of ADMM algorithm with $\gamma_l = 15$ dB, $\lambda = 100$, $\epsilon^{abs} = 0$, and $\epsilon^{rel} = 10^{-3}$.

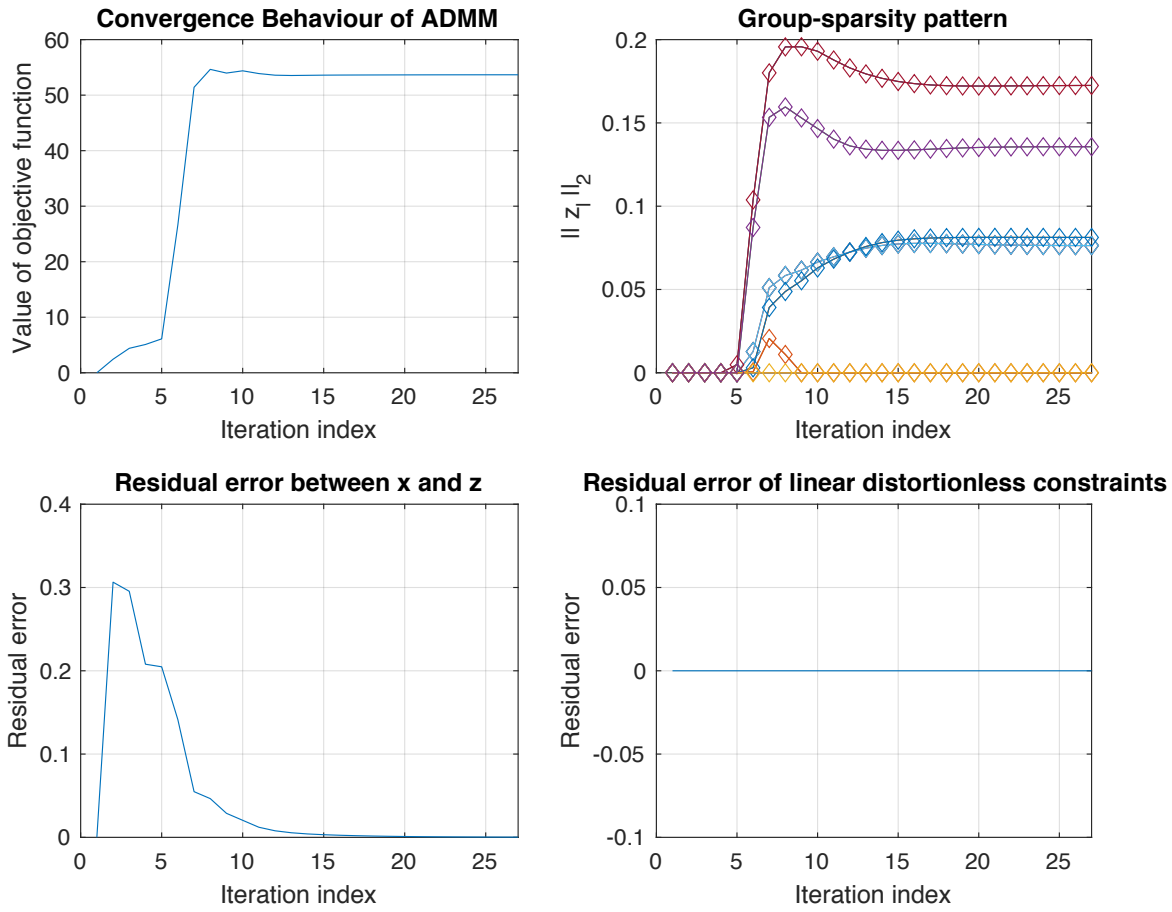

Fig. 5.2: Convergence behavior of ADMM algorithm with $\gamma_l = 15$ dB, $\lambda = 100$, $\epsilon^{abs} = 0$, and $\epsilon^{rel} = 10^{-3}$.

5.2.2 Two-step ADMM results

We shift our attention to the signal-to-interference ratio (SIR) performance at the destination nodes, which for destination k is defined as

$$SIR_k = \frac{\mathrm{E}\{|S_k(n)|^2\}}{\mathrm{E}\{|I_k(n)^2\}} \tag{5.4}$$

$$= \frac{p_k|\sum_{l=1}^{L} \mathbf{g}_{kl}^H \mathbf{B}_l \mathbf{h}_{lk}|^2}{\sum_{j\neq k} p_j \left|\sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj}\right|^2 + \sum_{l=1}^{L} \sigma_l^2 \|\mathbf{B}_l^H \mathbf{g}_{kl}\|_2^2} \tag{5.5}$$

$$= \frac{p_k|c_k|^2}{\sum_{j\neq k} p_j \left|\sum_{l=1}^{L} \mathbf{g}_{kl}{}^H \mathbf{B}_l \mathbf{h}_{lj}\right|^2 + \sum_{l=1}^{L} \sigma_l^2 \|\mathbf{B}_l^H \mathbf{g}_{kl}\|_2^2}. \tag{5.6}$$

The focus on SIR makes eliminates the dependence on the noise variance at the destination nodes. In Figure 5.3, we evaluate the achieved average SIR defined as $\frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{E}\{|S_k(n)|^2\}}{\mathrm{E}\{|I_k(n)^2\}}$ as a function of the relay input SNR. We vary the regularization parameter λ to examine the

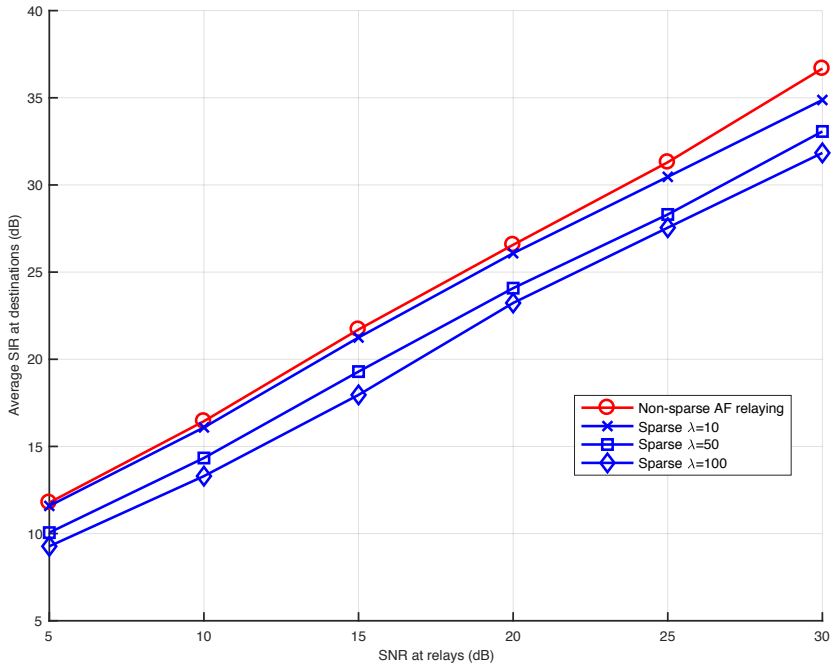

Fig. 5.3: Average SIR (in dB) versus SNR at the relays (in dB) at different sparsity levels

tradeoff between the achievable network performance and the energy efficiency in terms of the number of active relays and the total network energy consumption. The red curve ($\lambda = 0$) indicates the baseline performance obtained via leakage-minimization without any relay selection. Since all RRHs are involved in the relay transmission, the non-sparse AF relaying yields the highest SIR at the destinations, as expected. There's a linear relationship between the plotted variables where – by linear regression – we find that the relationship satisfies

$$SIR_{dB} = a \times SNR_{dB} + b \tag{5.7}$$

where

$$a = 0.9936$$

$$b = 6.6927.$$

The correlation coefficient is $r = 0.9998$. Since $a \approx 1$, we can model the relationship as $SIR_{dB} = SNR_{dB} + 6.7$. Accordingly, the curve demonstrates around 6.7 dB gain at the destination UEs without relay selection. When relay selection is incorporated into the problem, we observe that the associated curve shifts down from the baseline curve. This tradeoff is expected, since turning off some relays will result in some loss in SIR. For instance, for $\lambda = 100$, there's approximately a loss of 4 dB in SIR.

An alternative way of presenting the results is demonstrated in Figure 5.4. Again, we

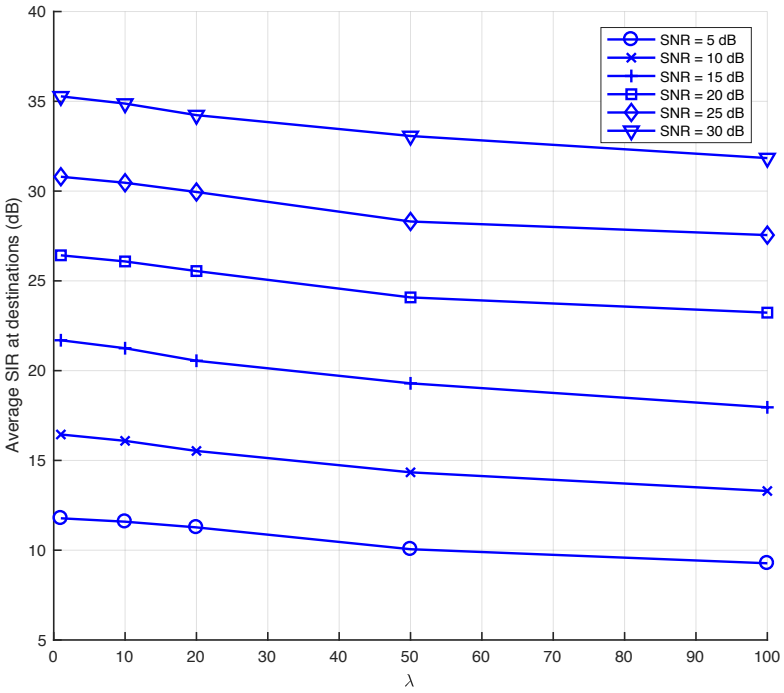

Fig. 5.4: Average destination SIR (in dB) versus the regularization parameter λ at different relay-SNR levels

observe that the SIR decreases with the increase on the weight placed on group-sparsity, i.e., λ. We also note that when $\lambda = 0$, there's around 6 dB gain in SIR.

To gain more intuition behind the performance reduction observed in Figure 5.3, we show the average number of active RRHs and their total transmission power with varying λ in Figure 5.5. It is observed that for all the relay input SNR levels, on average less than 5.5 RRHs are active for the case of $\lambda = 20$ while less than 4.5 RRHs become active for the case of $\lambda = 100$. The figure illustrates that the number of active relays is essentially controlled by λ irrespective of the SNR level at the relay. Intuitively, the higher the λ (or the weight placed on group-sparsity) the higher the number of inactive relays. The RRH individual power is defined as $P_l = \frac{1}{\eta_l} P_{t,l} + P_{c,l}$, where η_l denotes the efficiency of the power amplifier, e.g., $\eta_l = 50\%$, $P_{t,l}$ denotes the RRH transmission power and $P_{c,l}$ denotes the frounthaul link power consumption [43], which can be saved when the l-th RRH is switched off. The proposed group-sparse algorithm yields a 17% reduction in the sum RRH power. Based on the results, it becomes evident that the proposed solution can improve the network energy efficiency while still providing a satisfactory level of QoS for all the end-users.

5.2.3 Penalty parameter variation

The algorithm's speed of convergence is controlled by the choice of the penalty parameter ρ. Accordingly, the complexity of the ADMM-based selection is measured by varying this parameter, computing the CPU-time taken (in seconds) and making a comparison with the CPU-time required by two optimization packages: SeDuMi and SDPT3.

From Table 5.1, we note that a value of ρ in the range of $[2\lambda, 10\lambda]$ achieves the best convergence speed. The table demonstrates that the ADMM-based selection converges faster than both SeDuMi and SDPT3 when the proper value of the penalty parameter ρ is selected. As a result, one may use the ADMM-based algorithm as a fast selection

mechanism.

	Regularization parameter λ						
	0.1	1	5	20	100	250	500
SeDuMi	0.2210	0.2366	0.2244	0.2116	0.2105	0.2081	0.2078
SDPT3 4.0	0.3943	0.3966	0.3633	0.3792	0.3926	0.4077	0.4178
ADMM, $\rho = 0.1\lambda$	0.5976	0.7979	0.9390	1.0278	1.1046	1.1724	1.2460
ADMM, $\rho = 0.5\lambda$	0.1250	0.1473	0.1809	0.2103	0.2300	0.2514	0.2468
ADMM, $\rho = \lambda$	0.0650	0.0850	0.0911	0.1128	0.1292	0.1281	0.1392
ADMM, $\rho = 2\lambda$	0.0344	0.0459	0.0489	0.0627	0.0685	0.0707	0.0772
ADMM, $\rho = 10\lambda$	0.0317	0.0370	0.0379	0.0486	0.0705	0.0788	0.0837
ADMM, $\rho = 100\lambda$	0.2282	0.2667	0.3108	0.4195	0.6743	0.7306	0.6826

Table 5.1: CPU time (seconds) for $\epsilon_{abs} = 0, \epsilon_{rel} = 10^{-3}$

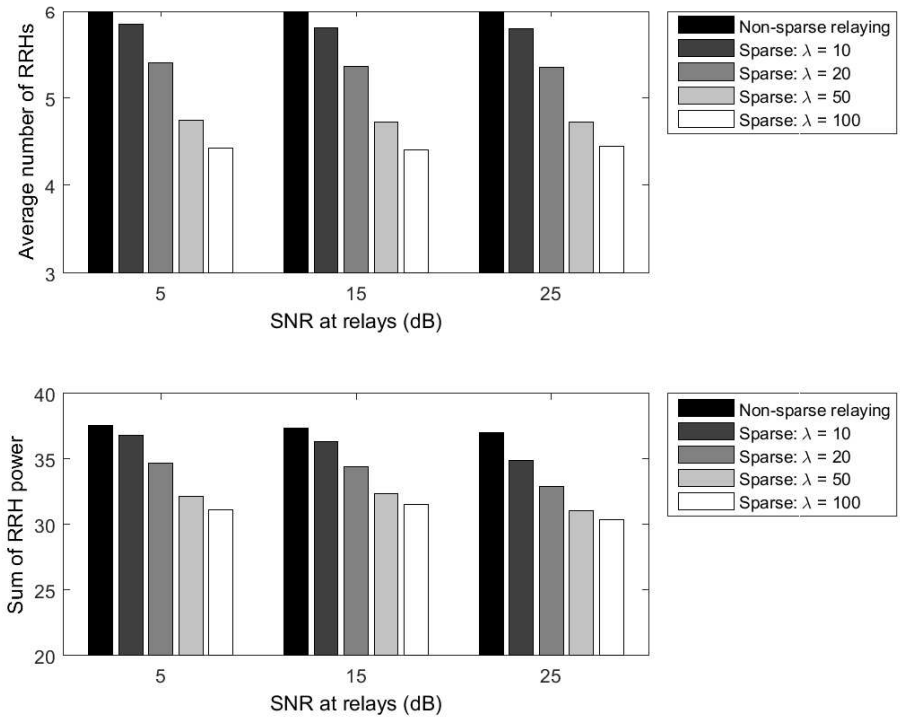

Fig. 5.5: Top figure: average number of active RRHs versus the regularization parameter λ. Bottom figure: total RRH transmission power versus the regularization parameter λ. The front-haul power is set to $P_{c,l} = 5.6$ for all RRHs.

Chapter 6

Conclusions and Future Work

In this thesis, a low-complexity joint RRH selection and relay AF transceiver optimization algorithm was proposed for a multiuser relaying network within a C-RAN. First, we presented the necessary mathematical background on optimization, MIMO processing and relaying in Chapter 2. We introduced convex problems, the subgradient and discussed duality and the KKT optimality conditions. Additionally, we presented the ADMM algorithm and its stopping critertion; an algorithm which has been essential in this thesis due to its fast convergence property.

In Chapter 3, we outlined the system model along with the problem formulation. Assuming known CSI at the BBU, the approach is based on choosing the AF matrices of the relays that minimize the total interference leakage at the destinations while satisfying RRH transmit power constraints along with a set of linear signal preservation constraints. Since we seek an energy-efficient solution, a regularization term representing the group-sparsity pattern associated with the relay AF matrices was added to the objective function. Finally, we were able to cast our formulation as a convex optimization problem by vectorizing the matrices and manipulating the expressions.

In Chapter 4, the AF matrices of all the relays are jointly optimized via a two-stage approach using the ADMM algorithm. Stage 1 – outlined in Algorithm 1 – serves as a quick selection mechanism that yields a subset of active relays $\mathcal{A}$, while stage 2 solves for the AF relay matrices by minimizing the interference leakage based on the previous results, as outlined in Algorithm 2. Specifically, the optimization in stage 2 includes explicit constraints on the activity of the RRHs. It is worthwile to note that the ADMM yields closed-form solutions at each iteration of Algorithm 1 and results in a solution vector which is *exactly* group-sparse, thanks to a thresholding operation that appears in the derivation of the closed-form solutions. In addition, we note that one of the steps of the ADMM-based algorithm may be run in a parallel fashion, further reducing computational complexity.

Simulation results in Chapter 5 demonstrate the explicit benefits of the proposed algorithm, which results in significantly lower power consumption and computational complexity than conventional relaying design methods.

As future work, the robustness of our approach to imperfect CSI can be tested via an appropriate statistical model. In practice, the CSI at the BBU pool will be imperfect, which may originate from a variety of sources. For instance, in frequency-division duplex (FDD) systems, the CSI imperfection may originate from downlink training based channel estimation and uplink limited feedback. It could also be due to the hardware deficiencies, partial CSI acquisition, and delays in CSI acquisition [14]. In addition, the limited fronthaul capacity on the optical transport links may be incorportated into the problem formulation. This would include including capacity constraints in the optimization formulation, and since capacity is a function of the SNR, this would introduce non-linear constraints. The challenge would be to 'linearize' the constraints in order to make the optimization amenable to the application of the ADMM algorithm. Otherwise, it would be challenging to obtain closed-form solutions, as derived in this thesis.

References

[1] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 1st ed., 2011.

[2] "Ericsson mobility report - on the pulse of the network society," *Ericsson*, 2015.

[3] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, pp. 74–80, Feb 2014.

[4] S. G. Glisic, *Relay-Assisted Wireless Networks*. Wiley Telecom, 2011.

[5] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. W. Mark, and X. Shen, "Cloud assisted HetNets toward 5G wireless networks," *IEEE Communications Magazine*, vol. 53, pp. 59–65, June 2015.

[6] *C-RAN White Paper: The Road Towards Green RAN*, Dec. 2013.

[7] D. Wubben, P. Rost, J. S. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN," *IEEE Signal Processing Magazine*, vol. 31, pp. 35–44, Nov 2014.

[8] B. Soret, K. I. Pedersen, N. T. K. Jørgensen, and V. Fernández-López, "Interference coordination for dense wireless networks," *IEEE Communications Magazine*, vol. 53, pp. 102–109, Jan. 2015.

[9] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander, "Impact of backhauling power consumption on the deployment of heterogeneous mobile networks," in *Proceedings of the 2011 IEEE Global Telecommunications Conference* -, pp. 1–5, Dec 2011.

[10] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks – A technology overview," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 405–426, 1st qtr. 2015.

[11] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, "Relay architecures for 3GPP LTE-advanced," *EURASIP Journal on Wireless Communication and Networking*, vol. 2009, Article ID 618787, pp. 1–14, 2009.

[12] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2593–2604, June 2007.

[13] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 2809–2823, May 2014.

[14] Y. Shi, J. Zhang, and K. B. Letaief, "Robust group sparse beamforming for multicast green cloud-RAN with imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 63, pp. 4647–4659, Sept 2015.

[15] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 494–508, Jan 2015.

[16] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for comp transmissions using mixed integer conic programming," *IEEE Transactions on Signal Processing*, vol. 61, pp. 3972–3987, Aug 2013.

[17] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2762–2775, June 2013.

[18] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[19] L. Liang and Z. Rui, "Downlink SINR balancing in C-RAN under limited fronthaul capacity," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3506–3510, Mar 2016.

[20] V. N. Ha and L. B. Le, "Joint coordinated beamforming and admission control for fronthaul constrained cloud-rans," in *Proceedings of the 2014 IEEE Global Communications Conference*, pp. 4054–4059, Dec 2014.

[21] A. Abdelnasser and E. Hossain, "Resource allocation for an ofdma cloud-ran of small cells underlaying a macrocell," *IEEE Transactions on Mobile Computing*, vol. 15, pp. 2837–2850, Nov 2016.

[22] S. H. Park, K. J. Lee, C. Song, and I. Lee, "Joint design of fronthaul and access links for C-RAN with wireless fronthauling," *IEEE Signal Processing Letters*, vol. 23, pp. 1657–1661, Nov 2016.

[23] S. H. Park, K. J. Lee, C. Song, and I. Lee, "Compressed cooperative reception for the uplink of c-ran with wireless fronthaul," in *Proceedings of the 2017 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 211–215, Aug 2017.

[24] B. Hu, C. Hua, J. Zhang, C. Chen, and X. Guan, "Joint fronthaul multicast beamforming and user-centric clustering in downlink C-RANs," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5395–5409, Aug 2017.

[25] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4837–4851, Dec 2009.

[26] C. Xing, S. Ma, and Y. C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Transactions on Signal Processing*, vol. 58, pp. 2273–2283, April 2010.

[27] J. Yang, B. Champagne, Y. Zou, and L. Hanzo, "Centralized energy-efficient multiuser multiantenna relaying in next-generation radio access networks," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 7913–7924, Sept 2017.

[28] S. S. Bidokhti, G. Kramer, and S. S. Shitz, "Capacity bounds on the downlink of symmetric, multi-relay, single receiver C-RAN networks," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 2058–2062, June 2017.

[29] Q. Huang and A. Burr, "The capacity of cloud-ran: Outer bound with quantisation and constrained fronthaul load," in *Proceedings of the 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 497–501, Nov 2016.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, Jan. 2011.

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

[32] R. Rockafellar, *Convex Analysis*. Princeton landmarks in mathematics and physics, Princeton University Press, 1970.

[33] A. Goldsmith and K. (Firm), *Wireless Communications*. Cambridge University Press, 2005.

[34] A. F. Molisch, *Diversity*. Wiley-IEEE Press, 2011.

[35] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA: Athena Scientific, 1997.

[36] J. Proakis, *Digital Communications*. Electrical engineering series, McGraw-Hill, 2001.

[37] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, pp. 198–218, Feb 2004.

[38] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, September 1979.

[39] G. Levin and S. Loyka, "Amplify-and-forward versus decode-and-forward relaying: which is better?," in *Proceedings of the 22nd International Zurich Seminar on Communications (IZS)*, Eidgenössische Technische Hochschule Zürich, 2012.

[40] 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*, vol. 59, pp. 5499–5508, nov 2011.

[41] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[42] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Applications*, vol. 284, pp. 193–228, Nov. 1998.

[43] Y. Shi, J. Cheng, J. Zhang, B. Bai, W. Chen, and K. B. Letaief, "Smoothedl_p-minimization for green cloud-ran with user admission control," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 1022–1036, April 2016.