TECHNIQUES FOR INTERFERENCE ANALYSIS AND SPECTRUM MANAGEMENT OF DIGITAL SUBSCRIBER LINES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mark H. Brady
April 2006
© Copyright by Mark H. Brady 2006
All Rights Reserved
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

John M. Cioffi  Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Thomas M. Cover

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ravi Narasimhan

Approved for the University Committee on Graduate Studies.
Abstract

Dynamic Spectrum Management (DSM) is emerging as a key component in next-generation digital subscriber line (DSL) technology. Because DSL twisted-pair binder channels are chiefly interference-limited, the multiuser performance and compatibility of DSM is of foremost importance. Although these issues are well understood for Static Spectrum Management (SSM) used in state-of-the-art DSL, existing techniques are not amenable to the study of Dynamic Spectrum Management.

This thesis presents novel interference analyses that characterize the performance of two classes of DSM: when multiuser interference is treated as noise, and when multiuser encoding and decoding (“vectored transmission”) is performed. A game-theoretic framework is adopted whereby Nash equilibria of certain strictly-competitive games characterize the “worst” interference scenarios.

When interference is treated as noise, this approach yields a lower bound that is close to performance in the field. Numerical results are presented for two relevant scenarios: an upstream VDSL deployment exhibiting the near-far effect, and an asymmetric DSL remote terminal (RT) deployment with long central office (CO) lines. The results show that the performance improvement of DSM over SSM techniques in these channels can be preserved by appropriate distributed power control, even in worst-case interference environments.

When multiuser encoding and decoding is feasible, results are obtained both for downstream and upstream transmission. For downstream transmission, a worst-case throughput (sum rate) is obtained from the Nash equilibrium of a strictly-competitive game. In upstream transmission, a general weighted sum-rate is similarly considered. Partial uniqueness properties of the Nash equilibria of each game are developed.
Acknowledgement

The past few years at Stanford have been both challenging and rewarding. It seems fitting upon completion of graduate studies to consider the path - and numerous individuals - that have lead to such a juncture.

It has been an honor and pleasure to join Prof. John Cioffi’s research group at Stanford. In serving as my principal academic advisor, Prof. Cioffi’s pioneering experience and vast knowledge of DSL have been an indispensable assistance. Moreover, his supervision and guidance have helped me to work efficaciously while avoiding many roadblocks in the doctoral degree program. I thank Prof. Cioffi for all his efforts on my behalf.

I am also indebted to co-advisor Prof. Thomas Cover. I have been most fortunate to have learned information theory under the tutelage of such a distinguished figure in the field. Prof. Cover’s research results have also been invaluable in my studies; it is hardly an exaggeration to say that many of the technical results herein owe their existence to standing upon shoulders of giants. I thank Prof. Cover for his generous assistance.

I would like to thank Prof. Yinyu Ye and Prof. Ravi Narasimhan for their participation with my oral defense and thesis. Prof. Ye’s rigorous and practical classes on optimization have aided and influenced my research efforts. It has been a pleasure collaborating with Prof. Narasimhan on optimization and wireless problems, as well as learning valuable information about the industry.

I would like to thank the Group’s administrative assistants Pat Oshiro and Joice Debolt for their cheerful and tireless help on countless occasions these past years. The Group will surely continue to benefit from their eminently capable hands.
Stanford has been all the more hospitable on account of the outstanding research community that is fostered. Briefly, I would like to acknowledge Jungsub Byun, Chiang-yu Chen, Seong-Taek Chung, Amal Ekbal, Arpita Ghosh, Chansoo Hwang, Sumanth Jagannathan, Edward Jang, Young-han Kim, Youngjae Kim, Bin Lee, Jung-won Lee, Wooyul Lee, Alessandro Magnani, Moshe Malkin, Mehdi Mohseni, Chris Ng, Jisung Oh, Vahbod Pourahmad, Wonjong Rhee, Kibeom Seong, Kee-Bong Song, Erik Stauffer, Dimitris Toupakaris, Mai Vu, Shu-ping Yeh, Taesang Yoo, David Yu, Wei Yu, and Rui Zhang, and many others with whom it has been a pleasure to work.

Financial support for graduate study has been provided by the Stanford Graduate Fellowship program under the sponsorship of Texas Instruments, as well as the National Science Foundation under contract CNS-0427677.
## Contents

Abstract iv

Acknowledgement v

1 Introduction 1

1.1 Background on Broadband Access 2

1.2 Research Contributions 4

1.3 Outline 5

1.4 Notation and Mathematical Preliminaries 6

1.5 Convex Optimization 9

2 The Worst-Case Interference 12

2.1 Technical Background 12

2.2 System Model 14

2.2.1 Channel Model 14

2.2.2 DSL Modem Model 14

2.3 The Worst-Case Interference (WCI) 16

2.3.1 Game-Theoretic Characterization of the WCI 16

2.3.2 Formalization of the WCI Game 17

2.3.3 Interpretation of the WCI Game 18

2.3.4 Derivation of Nash Equilibrium Conditions 20

2.3.5 Structure of the Worst-Case Interference 22

2.3.6 Numerical Computation of the Saddle Point 28

2.4 The Multiuser Worst Case Interference 38
5 Conclusions

5.1 Summary and Inferences ........................................ 98
5.2 Future Research .................................................. 99

A Dual Decomposition Subproblem .............................. 101

A.1 Analytical Solution ............................................. 102
A.2 Analytical Interpretation ...................................... 105

B Saddle Point Error Bounds ................................. 108

B.1 Formulation ...................................................... 108
B.2 Error Bound ...................................................... 109
B.3 Applications ...................................................... 113
  B.3.1 Worst Case Interference .................................. 113
  B.3.2 Multiuser Worst Case Interference ...................... 113

C Nash Equilibria Under an FDM Condition ............. 114

Bibliography ......................................................... 118
List of Tables

2.1 Summary of Interference Channel game $\mathcal{G}$. 
2.2 Summary of Multiuser Worst Case Interference game $\mathcal{H}$. 
4.1 Summary of Multiple Access Channel game $\mathcal{M}$. 
4.2 Summary of Broadcast Channel game $\mathcal{B}$. 

18  
40  
84  
89  
x
List of Figures

2.1 Illustration of spectrum management center (SMC) loop plant environment showing downstream FEXT and NEXT ....................................... 14
2.2 Numerical computation flowchart ................................................................. 29
2.3 Illustration of multiuser WCI with achievable and capacity regions .......... 39

3.1 Binder configuration for upstream VDSL simulations ................................. 65
3.2 Binder configuration for downstream RT ADSL simulations ...................... 66
3.3 Achievable rates in upstream VDSL as a function of victim line length .......... 67
3.4 Achievable rates in upstream VDSL as a function of power back-off ............ 68
3.5 Achievable rates in downstream ADSL as a function of RT power backoff ... 71
3.6 Spectral allocations of players at one operating point ................................ 72

4.1 Gaussian Broadcast Channel with 2 users .................................................... 76
4.2 Gaussian Multiple Access Channel with 2 users ......................................... 76
4.3 Illustration of $\mu$-weighted rate maximization over $C_{MAC}(\Lambda)$ considered in multiple access game $\mathcal{M}$ ................................................................. 85
4.4 Binder configuration for vectored transmission system .............................. 93
Chapter 1

Introduction

The information revolution - seeded by information theorists in the mid 20th century and developed by engineers in recent decades - continues to advance with scant abatement. Today, applications that might never have been anticipated by the pioneering theorists such as wireless local area networks, digital cellular telephony, and internet access are found in everyday life.

This thesis examines one particular manifestation of these developments - digital subscriber line (DSL) - and develops new technologies that improve its performance. Both the theoretical and technological context are germane; recent results in the fields of communication theory and multiuser information theory as well as recent developments in business practices influence which problems are relevant and amenable to solution.

From the perspective of communication theory, the copper twisted-pair channel used by DSL suffers from numerous impairments, including: frequency-selective attenuation, non-stationarity, power constraints, multiuser interference, thermal noise, and alien noise. Significant research over the past decades has effectively surmounted many of these impairments. Presently, multiuser interference effects are of great interest because they are the primary limiting constraint to data rate performance. Techniques for analyzing and mitigating this multiuser interference are developed in this thesis both for current generation and next-generation DSL systems.
CHAPTER 1. INTRODUCTION

From the perspective of information theory, exciting results continue to be developed for "classical" multiuser configurations such as broadcast and multiple-access channels. Such results, in combination with the rich literature on the topic, have enabled many of the technical analyses contained herein.

From the perspective of business, maturing technologies in internet access have expanded the broadband market while following a trend of deeper penetration and novel applications (home banking, personal communication, photography). Moreover, future technologies such as high-definition television (HDTV), video-on-demand, music services, and voice over internet protocol (IP) are incubated by limited high-speed broadband access presently available. Achieving the performance needed to deploy these new applications on a global scale continues to drive research and new business opportunities.

1.1 Background on Broadband Access

Facilitating economic prosperity and a vibrant civil society is a NSF national priority [31] that is being realized by educational and technical initiatives, among them ubiquitous high-speed internet access to reduce the “digital divide” between diverse communities. Each broadband access technique is constrained by technology and economics, between which certain trade-offs exist. Presently, high performance services generally require high capital expenditure and operating costs. This blocks goals such as universal high-speed service and perpetuates the “digital divide”.

A number of different modalities such as DSL, cable, wireless, and fiber are in competition for broadband access; each faces different technical performance and economic challenges. For example, technologies such as fiber-to-the-home (FTTH) may be used to deliver effectively arbitrarily large throughput to any location, but such performance may come with high infrastructure costs. Wireless access may offer lower infrastructure costs to establish basic connectivity, but suffer from both rate and delay performance limitations. Finally, DSL can leverage the existing copper twisted pair network, but currently has performance limited by multiuser crosstalk effects.
The first DSL services were introduced in the early 1990’s as an improvement over low-rate (approximately 160 kbps) Integrated Services Digital Network (ISDN) services [71]. Since that time, numerous standards and forms of DSL (termed xDSL) have substantially improved performance [71] [72]. Recently, very-high bit rate DSL (VDSL) has found broad adoption due to high performance (up to 100 Mbps) over short loops that are present in some localities. Asymmetric DSL (ADSL), which was originally intended for video broadcast, has been widely employed for home internet access due to the download-intensive nature of most applications and the flexibility of ADSL to operate under common loop deployments.

Because of rapid gains over the past two decades, DSL technologies are reaching fundamental limits that prevent further performance gains under current approaches. Two primary innovations have been proposed to ameliorate this problem:

- **Dynamic Spectrum Management (DSM)** is emerging as a key component in next-generation DSL standards. In DSM, spectrum is allocated adaptively in response to channel and interference conditions, allowing mitigation of interference and best use of the channel. Because multiuser interference is the primary limiting factor to DSL performance, the potential for rate improvement by exploiting the structure of the interference is substantial.

- **Multiuser “vectored” encoding and decoding** affords fundamental information-theoretic performance improvements over current single-user techniques. Although such “vectored” transmission can require significant changes to modem hardware, the advantages afforded by techniques such as interference pre-subtraction and cancelation are compelling.

The namesakes of **Dynamic Spectrum Management** and **Static Spectrum Management (SSM)** arise from the manner in which modems’ operating parameters are specified. A key parameter is the allocation of spectrum and the constraints placed on a modem’s emissions in the frequency domain. In SSM, masks (upper bounds) are imposed on transmit power spectral densities (PSDs) with the aim of limiting the amount of crosstalk induced in other lines sharing the same binder group [1]. Because these SSM masks are fixed for all loop configurations, they can often be far from
optimal or even prudent spectrum usage in typical deployments. DSM attempts to remedy this issue by dynamically adapting frequency masks. The situation for other modem parameters (coding, interleaving depth, power backoff, etc.) is analogous.

Within the SSM framework, techniques have been developed to characterize the crosstalk interference from DSL modems. Standardized tests for “spectral compatibility” [1] assess “new technology” by defining PSD masks and examining the impact to standardized systems using an empirical crosstalk model based on the 99th percentile worst-case crosstalk strength. Such analyses are useful when a reasonable estimate of spectrum of all users can be assumed à priori. However, if spectrum is instead allocated dynamically, not only is this knowledge is not available à priori, but other users’ spectrum may not even be known even during operation; hence the existing techniques are not applicable.

Spectral compatibility between different operators using DSM is a primary concern because new pathologies may arise with adaptive operation. For example, modems may be forced to adapt to changing interference observed as other modems in the binder vary their spectrum usage. Moreover, it is not unreasonable to suspect that each competing service provider would perform DSM in a greedy, if not malicious, fashion at the expense of spectral compatibility.

New paradigms are therefore needed to assess DSM and/or “vectored” multiuser performance in next-generation systems.

1.2 Research Contributions

The study of interference arising from DSL modems employing dynamic spectrum management has not attracted significant research interest and thus has heretofore been poorly understood both in theory and in practice. In addressing this subject, the research topics of this thesis may be delineated into three primary areas:

1. The interference properties of non-vectored DSL systems employing dynamic spectrum management are analyzed in a novel game-theoretic framework. Properties of the Nash equilibria of a certain strictly-competitive game are developed
and shown to correspond to relevant situations encountered in DSL systems. Based on this analysis, a novel bound is constructed that guarantees data rates for each DSL user in such a system.

2. The preceding bound is employed to characterize the performance of the well-known iterative waterfilling (IW) algorithm. It is shown that stable operating points of the IW algorithm attain certain minimum performances by the users. This fact is important in practice because the IW algorithm may have very many stable operating points - with varying levels of performance at each of them.

3. The interference properties of vectored DSL systems using dynamic spectrum management are studied from two novel game-theoretic frameworks corresponding to upstream and downstream transmission. Analogously to the non-vectored case, bounds on the system throughput (downstream) and any arbitrary weighting of the user rates (upstream) are constructed from Nash equilibria of these games. Such bounds may be applied to the design of next-generation systems and standards for vectored DSL.

1.3 Outline

This thesis is organized as follows. The remainder of this Chapter provides mathematical background: Section 1.4 defines notation and Section 1.5 provides a brief overview of finite-dimensional convex optimization.

In Chapter 2, Dynamic Spectrum Management is analyzed for modems that treat interference as noise. A notion of a worst-case interference for a given “victim” modem is defined and shown to exist. Structural properties of the worst interference are developed for relevant special cases. Two different algorithms are derived to compute this worst-case interference; both are proven to converge.

In Chapter 3, numerical examples of the worst-case interference analysis are presented. An emphasis is placed on loop deployments that are known to have unfavorable interference properties in actual DSL systems. Numerical results are given for
both ADSL and VDSL. It is shown that in these settings, the worst case interference produces a lower bound that is numerically close to the performance of the iterative waterfilling algorithm.

In Chapter 4, Dynamic Spectrum Management is analyzed in next-generation modems that may perform central-office-side multiuser encoding and decoding. Due to the asymmetry of the physical configuration, upstream and downstream transmission are analyzed separately. Lower bounds to achievable rates when experiencing hostile interference are computed for both cases.

Chapter 5 draws conclusions from the preceding results and suggests future directions for research in the area.

1.4 Notation and Mathematical Preliminaries

The notation employed herein is generally following that of [8] and [7]. Sequences are denoted as \( \{(z)_n\} \), where the parenthetical subscript notation \( (z)_n \) is used to denote the \( n \)th term in the sequence \( \{(z)_n\} \) (to avoid ambiguity with use of subscripts for vector indexing). The notation\(^1\) \( A \subset B \) denotes that the set \( A \) is a subset of \( B \).

Vectors are written in boldface, e.g. \( \mathbf{x} \), and are assumed to be column vectors unless otherwise noted. The notation \( \mathbf{x}_n \) denotes element \( n \) of the vector \( \mathbf{x} \). The notation \( \mathbf{x}_{-n} \) denotes the vector formed from \( \mathbf{x} \) by replacing the \( n \)th element with 0. \( \mathbf{0} \) denotes the zero vector, while \( \mathbf{1} \) denotes a vector of where each element is equal to 1. Define the vector \( \mathbf{e}_k \) to have all elements equal to 0 except the \( k \)th, which is equal to 1.

Vector \( p \)-norms on \( \mathbb{R}^n \) are defined for \( \mathbf{x} \in \mathbb{R}^n \) as

\[
||\mathbf{x}||_p = \left( \sum_{m=1}^{n} |x_m|^p \right)^{1/p}.
\]  

\(^1\)Some authors use identical notation to denote a proper subset and the notation \( \subseteq \) to denote a subset that may hold with equality. The notation of this work does not use this convention, hence \( A \subset A \) for all sets \( A \).
for all $1 \leq p \in \mathbb{R}$. For $p = \infty$ one takes

$$||x||_\infty = \max_m |x_m|. \quad (1.2)$$

Norms are assumed to be Euclidean $|| \cdot ||_2$ unless otherwise indicated.

A permutation $\pi$ on the finite set $X = \{1, \ldots, N\}$ is a bijective map $\pi : X \mapsto X$. Heuristically, $\pi$ represents a re-ordering of the elements of $X$. Thus if $v \in \mathbb{R}^n$ is a vector and $\pi$ is a permutation on $\{1, \ldots, N\}$, then $v_{\pi(n)}$ (where $n \in X$) denotes the $\pi(n)$-th element of the vector $x$. For a given permutation $\pi$ on $X$, the notation $\pi^{-1}$ denotes the unique bijection $\pi^{-1} : X \mapsto X$ satisfying $\pi^{-1}(\pi(n)) = n$ for all $n \in X$.

Matrix norms are defined in the usual fashion for $A \in \mathbb{R}^{n \times m}$ as

$$||A||_p = \sup_{||x||_p \leq 1} ||Ax||_p. \quad (1.3)$$

When the norm is not explicitly specified, the "spectral" norm\(^2\) may be assumed ($p = 2$).

Throughout this thesis, the standard topology is assumed on $\mathbb{R}^n$ along with customary topological definitions. Thus if $S \subset \mathbb{R}^n$ and $s \in S$, we say $s \in \text{int}(S)$ if there exists an open set $O$ such that $s \in O$ and $O \subset S$. Define the affine hull of $S$ as

$$\text{aff}(S) = \left\{ \theta_1 x_1 + \ldots + \theta_k x_k : x_1, \ldots, x_k \in S, \sum_{j=1}^k \theta_j = 1 \right\}. \quad (1.4)$$

For $s \in S$, we say that $s \in \text{relint}(S)$ if there exists an open set $O$ such that $s \in O$ and $\text{aff}(S) \cap O \subset S$.

Consider $f : D \mapsto \mathbb{R}$ where $D \subset \mathbb{R}^n$. The notation $\nabla f(d)$ denotes the gradient\(^3\) of $f$ at the point $d \in \text{int}(D)$. The following column vector notation is employed to

\(^2\)The spectral norm of $A$ is equal to the squared magnitude of the largest singular value of $A$.

\(^3\)The derivative in the sense of Gateaux.
CHAPTER 1. INTRODUCTION

denote the gradient, when it exists

\[ \nabla f(d) = \begin{bmatrix} \frac{\partial f}{\partial d_1}(d) \\ \vdots \\ \frac{\partial f}{\partial d_n}(d) \end{bmatrix}. \]  

(1.5)

The function \( f \) is said to be differentiable at \( d \) if the gradient (1.5) exists at \( d \). If \( \nabla f(d) \) exists for all \( d \in \mathbb{R}^n \) then \( f \) is said to be differentiable. If \( \nabla f(d) \) exists and is a continuous function of \( d \) at all points \( d \in \mathbb{R}^n \), then \( f \) is said to be smooth. The notation \( \nabla_S f(d) \) denotes the gradient with respect to some subset \( S \) of the variables in \( d \).

The notation \( \nabla^2 f(d) \) refers to the Hessian of \( f \) evaluated at the point \( d \in \text{int}(D) \), when it exists

\[ \nabla^2 f(d) = \begin{bmatrix} \frac{\partial^2 f}{\partial d_1 \partial d_1}(d) & \cdots & \frac{\partial^2 f}{\partial d_1 \partial d_n}(d) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial d_n \partial d_1}(d) & \cdots & \frac{\partial^2 f}{\partial d_n \partial d_n}(d) \end{bmatrix}. \]  

(1.6)

If \( \nabla^2 f(d) \) exists for some \( d \in \text{int}(D) \) then \( f \) is said to be twice differentiable at \( d \). Similarly, if \( \nabla^2 f(d) \) exists for all \( d \in \mathbb{R}^n \) then \( f \) is said to be twice differentiable.

**Definition 1** Given the map \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), we say that the vector \( g \in \mathbb{R}^n \) is a subgradient of \( f \) at \( x \in \mathbb{R}^n \) if \( f(z) \leq f(x) + (z - x)^T g \) for all \( z \in \mathbb{R}^n \).

**Definition 2** Given the map \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), the set of all subgradients of \( f \) at \( x \) is defined as the subdifferential of \( f \) at \( x \) and is denoted \( \partial f(x) \).

A supergradient may be defined analogously.

**Definition 3** Given the map \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), we say that the vector \( g \in \mathbb{R}^n \) is a supergradient of \( f \) at \( x \in \mathbb{R}^n \) if \( f(z) \geq f(x) + (z - x)^T g \) for all \( z \in \mathbb{R}^n \).

It may be readily verified that \( g \) is a supergradient of \( f \) at \( x \) if and only if \( -g \) is a subgradient of \( -f \) at \( x \). The following result [7] shows that the subgradient reduces
exactly to the gradient when a function is Frechet-differentiable\(^4\).

**Theorem 1** The map \( f : \mathbb{R}^n \mapsto \mathbb{R} \) is differentiable at \( x \) in the sense of Frechet with gradient \( \nabla f(x) \) if and only if the subdifferential at \( x \) exists and has unique element \( \nabla f(x) \), that is

\[
\nabla f(x) = \partial f(x).
\] (1.7)

### 1.5 Convex Optimization

Convex optimization plays a prominent role in the technical analyses of this work. This section\(^5\) introduces the topic of convex optimization and summarizes certain useful results. The following notation

\[
\begin{align*}
\min_x & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad a_i^T x = b_i, \quad i = 1, \ldots, q.
\end{align*}
\] (1.8)

where \( f_i : \text{dom } f_i \mapsto \mathbb{R} \), \( \text{dom } f_i \subset \mathbb{R}^n \), \( i = 0, \ldots, m \) are convex functions, denotes a *convex optimization problem* \([8]\). Define the domain \( \mathcal{D} \) of the optimization problem (1.8) as

\[
\mathcal{D} = \bigcap_{i=0}^m \text{dom } f_i.
\] (1.9)

A point \( x \in \mathcal{D} \) satisfying \( f_i(x) \leq 0, i = 1, \ldots, m \) and \( a_i^T x = b_i, i = 1, \ldots, q \) is said to be *feasible*. The set of all feasible points is denoted \( \mathcal{X} \)

\[
\mathcal{X} = \{ x : f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad a_i^T x = b_i, \quad i = 1, \ldots, q \}.
\] (1.10)

\(^4\)Note that differentiability in the sense of Frechet implies differentiability in the sense of Gateaux \([7]\).

\(^5\)Following \([8]\).
The optimal value of (1.8), termed \( p^* \), is defined as

\[
p^* = \inf_x \{ f_0(x) : f_i(x) \leq 0, \ i = 1, \ldots, m, \ a_i^T x = b_i, \ i = 1, \ldots, q \}
\]

\[
= \inf_{x \in X} f_0(x).
\]  

(1.11)

if a feasible point exists, and \( p^* = \infty \) if no feasible point exists. Note that by this extended-real convention, the optimum value of the convex optimization problem may be \( \pm \infty \). The function \( f_0 \) is termed the objective function or simply the objective.

We say that \( x^* \in \mathbb{R}^n \) is optimal\(^6\) for (1.8) if \( x^* \) is feasible and \( f_0(x^*) = p^* \). If \( x \) is feasible, we say that the constraint \( f_i(x) \leq 0 \) is active (respectively inactive) at \( x \) if \( f_i(x) = 0 \) (\( f_i(x) < 0 \)).

Now define the map \( L : \text{dom} \ f_0 \times \mathbb{R}^m \times \mathbb{R}^q \mapsto \mathbb{R} \), termed the Lagrangian, as

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^q \nu_i (a_i^T x - b_i),
\]  

(1.12)

where the newly-defined variables \( \lambda, \nu \) are termed dual variables. Further define the (Lagrange) dual function \( g : \mathbb{R}_+^m \times \mathbb{R}^q \mapsto \mathbb{R} \) where

\[
g(\lambda, \nu) = \inf_{x \in X} L(x, \lambda, \nu) = \inf_{x \in X} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^q \nu_i (a_i^T x - b_i)
\]  

(1.13)

where the value is taken as \(-\infty\) if no greatest lower bound exists. The (Lagrangian)

\(^6\)Although the notation (1.8) contains the symbol \( \min \), observe that the “minimum” need not be achieved in the sense that there need not exist an optimal \( x^* \); moreover, there need not even exist any feasible point. However, the sequel is concerned exclusively with problems where the minimum exists, in which case the notation of convex optimization problems is in exact agreement with mathematical convention.
dual problem is defined\textsuperscript{7} as
\[ \max_{\lambda,\nu} g(\lambda, \nu) \quad (1.14) \]
subject to \( \lambda \succeq 0 \).

Now define
\[ d^* = \sup_{\lambda \succeq 0, \nu} g(\lambda, \nu). \quad (1.15) \]

It can be shown that for any feasible \((\lambda, \nu)\), the value \(g(\lambda, \nu)\) is a lower bound to \(p^*\). In light of this lower bound, one always has \(d^* \leq p^*\). We will say that \textit{strong duality} holds if \(d^* = p^*\). Strong duality can be shown to hold under certain technical conditions. One particularly useful set of such conditions is the following

1. The functions \(f_0, f_1, \ldots, f_m\) are convex.

2. There exists some \(\mathbf{x} \in \text{relint}(D)\) such that \(f_i(\mathbf{x}) < 0, \ i = 1, \ldots, m\) and \(a_j^T \mathbf{x} = b_j, \ j = 1, \ldots, q\). This condition is alternatively termed \textit{Slater’s constraint qualification}.

In fact, these conditions imply not only strong duality, but that there exists some \(\lambda \succeq 0, \nu\) such that \(g(\lambda, \nu) = p^*, \ i.e.,\) the supremum in (1.15) is achieved.

\textsuperscript{7}Note that as in the primal problem, the maximum need not be achieved, and thus the use of the term max differs from mathematical convention. However, in the sequel we shall be concerned exclusively with situations where the maximum is achieved.
Chapter 2

The Worst-Case Interference

This chapter motivates and introduces the Worst Case Interference (WCI) analysis in digital subscriber line (DSL) modems that treat interference as noise. Section 2.1 provides background information on interference channels and spectrum management in state-of-the-art DSL. Section 2.2 defines the channel and system model currently used for discrete multi-tone (DMT) DSL. Section 2.3 motivates and defines the worst-case interference analysis, discusses its properties, and considers two classes of numerical algorithms to solve it.

Section 2.4 gives a generalization of the (single “victim” user) worst-case interference to multiple-user rate bounds. The proofs of a number of Lemmata and Theorems used in the preceding sections are given in Section 2.5.

2.1 Technical Background

DSM algorithms have been proposed for the cases of distributed and centralized modem control scenarios. This chapter considers what has been termed “Level 0-2 DSM” [70], wherein modems may be controlled to manage spectrum and other operating parameters, but not for multiuser encoding and decoding. A centralized DSM center controlling multiple lines offers both higher potential performance and improved management capabilities than purely autonomous operation [48]. Distributed DSM schemes based on the iterative waterfilling (IW) algorithm [72] have been presented.
IW has also been studied from a game-theoretic viewpoint [14] [58]. Numerous algorithms for centralized DSM have been proposed. [10] presents a technique to approximately maximize users’ weighted sum-rate. Approximate rate maximization has also been considered subject to frequency-division and fixed rate-proportion constraints [73]. Optimal [11] and suboptimal [53] algorithms to minimize transmit power have also been studied.

The capacity region of the Gaussian interference channel (IC) is in general unknown, even for the simplest 2-user case [66]. Communication in the presence of hostile interference has been studied from a game-theoretic perspective in numerous applications, e.g. [63] [42]. A simple and relevant IC achievable region is that attained by treating interference as noise; this is a subset of the Han-Kobayashi region, which is the most general achievable region known\(^1\) [40]. For parallel independent interference channels where interference is treated as noise, an achievable region is given in [16]. Capacity results for frequency-selective interference channels satisfying the strong interference condition are also known [13].

An extensive suite of DSL literature on upstream power-backoff techniques to mitigate the “near-far” problem has been developed for static spectrum-management systems [45] [67] [11] [47]. A power-backoff algorithm for DSM systems implementing iterative waterfilling has been proposed [88].

In current DSL standards, upstream and downstream transmission use either distinct frequency bands or shared bands. In the latter case\(^2\), “echo” is created between upstream and downstream transmission [72]. As analog hybrid circuits do not provide sufficient isolation, echo mitigation is essential in practical systems [44]. Numerous echo cancellation structures have been proposed for DSL transceivers [3] [87] [46].

---

1 The Han-Kobayashi achievable scheme requires, in general, multiuser decoding; treating interference as noise does not suffice to achieve all points in the achievable region. “Multiuser interference subtraction” is an interpretation of the operation of such a multiuser decoder.

2 FDM systems may also suffer from limited echo due to frequency bleeding effects, particularly near band edges.
2.2 System Model

2.2.1 Channel Model

A copper twisted-pair DSL binder is modelled as a frequency-selective multiuser Gaussian interference channel [72] [22]. The binder contains a total of $L + 1$ twisted pairs, with one DSL line per twisted pair, as shown in Figure 2.1. The effect of near-end crosstalk (NEXT) and far-end crosstalk (FEXT) interference generated by $L$ “interfering” users that generate crosstalk into one “victim” user is considered. This coupling is illustrated for downstream transmission in Figure 2.1.

2.2.2 DSL Modem Model

Modem Architecture

The standardized [29] discrete multitone\(^3\) (DMT)-based modulation scheme is employed, so that transmission over the frequency-selective channel may be decoupled into $N$ independent subcarriers or tones. Both frequency division multiplexing (FDM) and overlapping bandplans are considered. As overlapping bandplans require echo

\(^3\)DMT is an orthogonal frequency division multiplexing (OFDM) scheme.
cancellation that is imperfect in practice, error that is introduced acts as a form of interference and is of concern. Echo cancellation error is modelled presuming a prevalent echo cancellation structure utilizing a joint time-frequency LMS algorithm \cite{44} is employed\footnote{Other models may be more applicable to different echo cancellation structures.}. Using the terminology of \cite{44}, let $\mu$ denote the LMS adaptive step size parameter. The “excess MSE” for a given tone is modelled \cite[Eqn. 12.74]{86} as proportional to the product of the LMS adaptive step size parameter $\mu$ and the transmit power on that tone. The constant of proportionality is absorbed by defining $\hat{\beta}$ as the ratio of excess MSE to transmitted energy on a given tone.

\textbf{Achievable Rate Region}

This section discusses an achievable rate region for a DSL modem based on the preceding channel and system model. The following analysis applies to both upstream and downstream transmission. For specificity, the following refers to downstream transmission: first, consider the case where echo cancellation is employed. Denote the victim modem’s downstream transmit power on tone $n$, $n \in \{1, \ldots, N\}$ as $x_n$. Let element $l$, $l \in \{1, \ldots, L\}$ of the vector $y^{(n)} \in \mathbb{R}^{2L}_+$ denote the downstream transmit power of interfering modem $l$ on tone $n$. Similarly, let element $l$, $l \in \{L+1, \ldots, 2L\}$ of $y^{(n)}$ denote the upstream transmit power of interfering user $l-L$. Define element $l$, $l \in \{l, \ldots, L\}$ of the row vector $h^{(n)} \in \mathbb{R}^{2L}_+$, as the FEXT power gain from interfering user $l$ on tone $n$ (necessarily, $h^{(n)} \succeq 0$). Similarly, define element $l$, $l \in \{L+1, \ldots, 2L\}$ of $h^{(n)}$ to be the NEXT power gain from interfering user $l-L$. Let element $n$ of $\tilde{h}_n \in \mathbb{R}^{N}_+$ denote the victim line’s insertion gain on tone $n$ ($\tilde{h}_n \geq 0$).

Independent AWGN (thermal noise) with power $\sigma_n^2 > 0$ is present on tone $n$. Let $\hat{\beta}_n$ denote the echo cancellation ratio on tone $n$ as described above. Echo-cancellation error is treated as AWGN. Let $\Gamma$ denote the SNR gap-to-capacity \cite{72}. Then the following bit loading\footnote{The achieved data rate of a given modem is proportional to the number of bits loaded (less overhead); this constant of proportionality is normalized to 1 in the theoretical development.} is achievable on tone $n$ by treating interference as noise \cite{72}

\begin{equation}
   b_n = \log \left( 1 + \frac{\tilde{h}_n x_n}{\Gamma(h^{(n)} y^{(n)} + \hat{\beta} x_n + \sigma_n^2)} \right) . \tag{2.1}
\end{equation}
Observe that if \( \tilde{h}_n = 0 \), then it is necessarily the case that \( b_n = 0 \), implying that tone \( n \) is never loaded. Thus, in the sequel, \( \tilde{h}_n > 0 \) for all \( n \in \{1, \ldots, N\} \) is considered without loss of generality by removing those tones with zero direct gain (\( \tilde{h}_n = 0 \)). Defining \( \alpha_n = \Gamma / \tilde{h}_n \), \( \beta_n = \Gamma \hat{\beta}_n / \tilde{h}_n \), and \( N_n = \Gamma \sigma_n^2 / \tilde{h}_n \), and substituting

\[
b_n = \log \left( 1 + \frac{x_n}{\alpha_n h^{(n)}y^{(n)} + \beta_n x_n + N_n} \right).
\]

Because \( \Gamma \geq 1 \), it follows that \( \alpha_n > 0 \), \( \beta_n \geq 0 \), and \( N_n > 0 \).

### Achievable Rate Region for FDM

When an FDM scheme is employed, NEXT and echo cancellation are eliminated because transmission and reception occur on distinct frequencies\(^6\). As a common configuration in ADSL and VDSL standards [72], this represents the important special case of the preceding model where \( \beta_n = 0 \) (with echo cancellation) and \( h^{(n)} = 0 \) for all \( n \), \( L + 1 \leq l \leq 2L \) (with frequency division multiplexing). Additional technical results will be shown to hold in the FDM setting, as detailed in Section 2.3.

### 2.3 The Worst-Case Interference (WCI)

#### 2.3.1 Game-Theoretic Characterization of the WCI

This section introduces and motivates the concept of the worst-case interference (WCI). Suppose that a “victim” modem desires to keep its data rate at some level. Such a scenario is commonplace as carriers widely offer DSL service at fixed data rates. The objective is to bound the impact that multiuser interference can have on this victim modem, thereby determining whether service may be guaranteed. To this objective, one considers interferences that are the most harmful in the sense of minimizing the achievable rate of a “victim” modem. However, it is not clear what form such interferences might take, nor what response they should merit.

---

\(^6\)Effects arising from implementation issues that may lead to crosstalk between upstream and downstream bands are not explicitly considered.
Examining this problem from the standpoint of game theory leads to substantial insight. Consider a worst-case interference game where one player jointly optimizes the spectrum of all the interfering modems, irrespective of the data rate they achieve in doing so, to cause the most deleterious interference to the victim modem. Thus in this game all the interfering modems act as one player, while the victim modem acts as the other player, with the channel and noise variance known to all. Although such an arrangement may appear pathological, it will be shown numerically that such a situation is quite close to what occurs in certain loop topologies. Neither is assuming such coordination of the interferers unreasonable in practice, because under “Level 2” DSM [70] [48] each collocated carrier may individually coordinate its own lines, or collocated equipment may be centrally controlled by a competing carrier. Channels may be estimated in the field, approximated by standardized models [72], and in the future, potentially published by operators [15] or otherwise inferred from public information.

A Nash equilibrium in this game may be interpreted as characterizing a worst-case interference an optimal response (power allocation policy) to it. The structure of the Nash equilibrium lends insight into the problem as well as suggesting techniques that may be implemented in practical systems.

### 2.3.2 Formalization of the WCI Game

Consider the following two-player game: let Player 1 control the spectrum allocation of victim modem, and Player 2 control the spectrum allocations of all the interfering modems. Referring again to downstream transmission for specificity, let the total (sum) downstream power of the victim modem $\sum_n x_n$ be upper bounded by $P^x$ where $0 < P^x < \infty$. Player 1 is also subject to a positive power constraint $C^x$ on each tone, so that $x \preceq C^x$. Note that this constraint may be made redundant by setting e.g. $C^x \geq 1P^x$. The requirement that $C^x > 0$ is without loss of generality by disregarding all unusable tones $n$ for which $C^x_n = 0$. Similarly for Player 2, consider per-line power constraints $0 \prec P^y \prec \infty$, where the total downstream power of the $l^{th}$ interfering modem $l \in \{1, \ldots, L\}$ is upper bounded by the $l^{th}$ element of $P^y \in \mathbb{R}_{++}^{2L}$ and the total
CHAPTER 2. THE WORST-CASE INTERFERENCE

Player | Controls | Objective | Strategy Set |
-------|----------|-----------|--------------|
1      | Victim modem | max $J$ | $\{x : 0 \leq x \leq M^x, \sum_n x_n \leq P^x\}$ |
2      | Interfering modems | min $J$ | $\{[y^{(1)}, \ldots, y^{(N)}] : 0 \leq y^{(n)} \leq M^{y,(n)}, [y^{(1)}, \ldots, y^{(N)}]1 \leq P^y\}$ |

Table 2.1: Summary of Interference Channel game $G$.

The upstream power of interfering modem $l$ is upper bounded by element $l + L$ of $P^y$. Further, consider positive power constraints $C^{y,(n)} \in \mathbb{R}^{2L}_{++}$ for $n = 1, \ldots, N$ such that $y^{(n)} \preceq C^{y,(n)}$ for each $n$; any such power constraints equal to zero may be equivalently enforced by zeroing respective element(s) of $\{h^{(n)}\}$.

The strategy set of Player 1 is the set of all feasible power allocations for the victim modem, $S_1 = \{x : 0 \leq x \leq C^x, 1^T x \leq P^x\}$, and the strategy set of Player 2 is the set of all feasible power allocations for the interfering modems, $S_2 = \{[y^{(1)}, \ldots, y^{(N)}] : 0 \leq y^{(n)} \leq C^{y,(n)}, n = 1, \ldots, N, [y^{(1)}, \ldots, y^{(N)}]1 \leq P^y\}$. Define $S = S_1 \times S_2$. This is a strictly competitive or zero sum two-player game $(S_1, S_2, J)$ where the objective function $J : S \mapsto \mathbb{R}_+$ is defined to be the achievable data rate of the victim user

$$J(x, [y^{(1)}, \ldots, y^{(N)}]) = \sum_{n=1}^{N} \log \left(1 + \frac{x_n}{\alpha_n h^{(n)} y^{(n)} + \beta_n x_n + N_n}\right) . \quad (2.3)$$

The game $G = (S_1, S_2, J)$ is defined to be the Worst-Case Interference game. A summary of the Worst-Case Interference game is given in Table 2.1.

### 2.3.3 Interpretation of the WCI Game

The previous section has specified the feasible strategies of each player in the game as well as each player’s objective. It remains to be defined in precisely what manner game play proceeds. Two possible scenarios are as follows: first, one could have the victim modem (Player 1) choose a strategy, whereupon the interfering modems (Player 2) could view that strategy and formulate an optimal response to it. If each player acted optimally, their behavior could be computed by solving the following
optimization problem

$$\max_x \min_{\{y(1), \ldots, y(N)\}} \sum_{n=1}^{N} \log \left( 1 + \frac{x_n}{\beta_n x_n + \alpha_n h(n) y_n + \sigma_n} \right)$$

subject to

$$x \succeq 0, x \preceq C x,$$

$$y^{(n)} \succeq 0, y^{(n)} \preceq C^{y,(n)} \quad n = 1, \ldots, N$$

$$1^T x \leq P^x,$$

$$[y^{(1)} \ldots y^{(N)}] 1 \preceq 1 P^y.$$ \hspace{1cm} (2.4)

Alternatively, Player 2 could choose some strategy first, and Player 1 would then choose an optimal response to that strategy. This would result in the following optimization problem

$$\min_{\{y^{(1)}, \ldots, y^{(N)}\}} \max_x \sum_{n=1}^{N} \log \left( 1 + \frac{x_n}{\beta_n x_n + \alpha_n h(n) y_n + \sigma_n} \right)$$

subject to

$$x \succeq 0, x \preceq C x,$$

$$y^{(n)} \succeq 0, y^{(n)} \preceq C^{y,(n)} \quad n = 1, \ldots, N$$

$$1^T x \leq P^x,$$

$$[y^{(1)} \ldots y^{(N)}] 1 \preceq 1 P^y.$$ \hspace{1cm} (2.5)

Among these two scenarios ((2.4) and (2.5)), it is clear that it is never disadvantageous to play second: whosoever moves second can see his opponent’s strategy beforehand, and therefore has additional information as opposed to moving first. This is an interpretation of the mathematical inequality underlying weak duality

$$\sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} f(x, y) \leq \inf_{y \in \mathcal{Y}} \sup_{x \in \mathcal{X}} f(x, y),$$ \hspace{1cm} (2.6)

which holds regardless of assumptions on $\mathcal{X}, \mathcal{Y}$, and $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ \cite{8}. In particular, it holds when the conditions of (2.4) and (2.5) are assumed.

Consider the physical interpretation of the inequality (2.6) holding strictly. This would imply that for any $x$ that the victim modem chooses, the interferers would choose $[y^{(1)}, \ldots, y^{(N)}]$. But then it would be possible, for this given $[y^{(1)}, \ldots, y^{(N)}]$ to increase the victim modem’s rate by the victim reloading some new $x'$. To respond, the interferers could choose a new $[y'^{(1)}, \ldots, y'^{(N)}]$, and so on. In this sense, the game
would be unstable and it is not clear what might be rightfully termed the “worst-case interference”; indeed the “worst-case interference” would depend explicitly on which choice of $x$ was elected.

Now consider the physical interpretation of the (2.6) holding with equality. There is then no advantage to moving second, the interference resulting from Player 2’s optimal strategy can be termed a worst-case interference, and the optimizations (2.4) and (2.5) are equivalent. This scenario, which will be shown to occur in the WCI game, is directly related to the existence of a pure-strategy\(^7\) Nash equilibrium.

### 2.3.4 Derivation of Nash Equilibrium Conditions

A Nash equilibrium in pure strategies in the WCI game $G$ is defined to be any saddle point $(x, [y^{(1)}, \ldots, y^{(N)}]) \in S$ satisfying

\[
J(\tilde{x}, [y^{(1)}, \ldots, y^{(N)}]) \leq J(x, [y^{(1)}, \ldots, y^{(N)}]) \leq J(x, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]),
\]

for all $\tilde{x} \in S_1$, $[\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \in S_2$. The condition (2.8) immediately implies the claim that Player 1’s rate at a Nash equilibrium of $G$ lower bounds the achievable rate with any other feasible interference profile. This bound also extends to other settings: in the noncooperative IW game [14], a (possibly non-unique) Nash equilibrium is known to always exist in pure strategies; condition (2.8) again yields a rate lower bound at every Nash equilibrium of the IW game for the line corresponding to Player 1.

It is now shown that a Nash equilibrium of $G$ always exists because of certain properties of the objective and strategy sets. First, the convex-concave structure of the objective is established.

**Theorem 2** If $\alpha \geq 0$, $\beta \geq 0$, $\gamma > 0$, $h \in \mathbb{R}^{2L}_+$, and $\alpha, \beta, \gamma, h$ are bounded, then the

\[^7\]Pure strategies contrast with mixed, or randomized strategies that could be adopted by the players. In defining Nash equilibria under mixed strategies, the equations (2.7), (2.8) customarily are taken with respect to expectations over the strategies.
function $g : \mathbb{R}_+ \times \mathbb{R}^{2L}_+ \mapsto \mathbb{R}_+$ defined by

$$g(x, y) = \log \left( 1 + \frac{x}{\alpha h^T y + \beta x + \gamma} \right)$$

is strictly concave in $x$ and is convex in $y$.

Proof: It is first shown that $f : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_+$, $f(x, \eta) = \log ((1 + \beta)x + \alpha \eta + \gamma) - \log(\alpha \eta + \beta x + \gamma)$ is convex in $\eta$ and strictly concave in $x$. It is sufficient [8] to show that for all $x \geq 0$, it holds $\frac{\partial^2 f}{\partial \eta^2} \geq 0$ on the interval $(-\epsilon, \infty)$ for some $\epsilon > 0$ and similarly for all $\eta \geq 0$ that $\frac{\partial^2 f}{\partial x^2} < 0$ on the interval $(-\epsilon, \infty)$ for some $\epsilon > 0$. By differentiating and simplifying

\[
\frac{\partial f}{\partial x} = \frac{\alpha \eta + \gamma}{(\alpha \eta + \beta x + \gamma)((\beta + 1)x + \alpha \eta + \gamma)},
\]

\[
\frac{\partial^2 f}{\partial x^2} = \frac{-(\alpha \eta + \beta x + \gamma)^2 (\alpha \eta + (\beta + 1)x + \gamma)^2}{(\alpha \eta + (\beta + 1)x + \gamma)^2} < 0,
\]

\[
\frac{\partial f}{\partial \eta} = -\frac{\alpha x}{(\alpha \eta + \beta x + \gamma)((\beta + 1)x + \alpha \eta + \gamma)},
\]

\[
\frac{\partial^2 f}{\partial \eta^2} = \frac{\alpha^2 (2\alpha \eta + (2\beta + 1)x + 2\gamma)x}{(\alpha \eta + \beta x + \gamma)^2 (\alpha \eta + (\beta + 1)x + \gamma)^2} \geq 0,
\]

where $\epsilon = \gamma/(4\beta(\beta+1))$ in (2.11) and $\epsilon = \gamma/(2\alpha)$ when $\alpha > 0$ and $\epsilon = 1$ when $\alpha = 0$ in (2.13). For all $(x, y) \in \mathbb{R}_+ \times \mathbb{R}^{2L}_+$, it must be that $h^T y \geq 0$. Thus $g(x, y) = f(x, h^T y)$. By the affine mapping composition property [8], it follows that $g(x, y)$ is convex in $y$ and strictly concave in $x$.

Lemma 1 The function $J$ is strictly concave in $x$ for fixed $[y^{(1)}, \ldots, y^{(N)}]$ and convex in $[y^{(1)}, \ldots, y^{(N)}]$ for fixed $x$.

Proof: Because the objective (2.3) is a sum of functions that are strictly concave in $x_n$ and convex in $y^{(n)}$, $J$ is strictly concave in $x$ and convex in $[y^{(1)}, \ldots, y^{(N)}]$ [8]

Theorem 3 The WCI game $G$ has a (possibly non-unique) Nash equilibrium in pure strategies, and a value $^8 R^*$.

^8That $G$ has a value means precisely that the the objective is equal at all Nash equilibria of $G$. 

 CHAPTER 2. THE WORST-CASE INTERFERENCE 21
Proof: Because $S_1 \subset \mathbb{R}^N$ and $S_2 \subset \mathbb{R}^{2LN}$ are closed and bounded, by Heine-Borel they are both compact. Also, the objective is a composition of continuous functions, hence continuous, and $J$ is strictly concave in $x$ and convex in $[y^{(1)}, \ldots, y^{(N)}]$ (Lemma 1). The conditions of [5, Thm. 4.4] are thus satisfied and therefore a pure-strategy saddle point exists. Note that the saddle-point need not be unique, in general. Because a saddle point exists in pure strategies, the game has a value [5, Thm. 4.1], which shall be denoted as $R^*$. Thus

$$\max_{x \in S_1} \min_{[y^{(1)}, \ldots, y^{(N)}] \in S_2} J = \min_{[y^{(1)}, \ldots, y^{(N)}] \in S_2} \max_{x \in S_1} J = R^*. \quad (2.14)$$

2.3.5 Structure of the Worst-Case Interference

The previous section showed that under very general conditions, a Nash equilibrium in $\mathcal{G}$ exists. Because Nash equilibria represent “worst” interferences, it is possible to study the properties of the worst interference by examining the structure of the Nash equilibria of $\mathcal{G}$. This analysis will allow one to answer the following questions:

- Does there exist a unique Nash equilibrium or unique ‘worst case interference’?
- If the Nash equilibrium is not unique, what properties does the worst interference have, and how do they impact the optimal response of Player 1?
- How does the action of Player 1 relate to the iterative waterfilling algorithm?

The first question may be partially resolved by examining the following numerical example

**Example 1** Let $N = 2$, $L = 2$, $h^{(1)} = h^{(2)} = [1 \ 1]$, $P^x = 1$, $P^y = [1 \ 1]^T$, $N_1 = N_2 > 0$, $\alpha_1 = \alpha_2 = 1$, $\Gamma = 1$, and suppose the FDM condition is satisfied and the per-tone power constraints are redundant. Then it may be readily verified by symmetry arguments that for $x = [1/2 \ 1/2]^T$, both $y^{(1)} = [1 \ 0 \ 0 \ 0]^T$, $y^{(2)} = [0 \ 1 \ 0 \ 0]^T$ and $y^{(1)} = y^{(2)} = [1/2 \ 1/2 \ 0 \ 0]^T$ (and convex combinations thereof) form Nash equilibria of $\mathcal{G}$. 
This example shows that far from the Nash equilibrium being unique, there may exist infinitely many Nash equilibria. Given that the Nash equilibrium is not generally unique, it may not be the case that the worst-case interference is unique. The structure of the set of Nash equilibria - and the associated interference profiles - is explored in the following results. Some basic intuition is first established showing that “waterfilling” is Player 1’s optimal strategy in response to the interference induced at a given Nash equilibrium where the FDM condition holds and the individual-tone constraints are inactive.

**Theorem 4** Let \( (\hat{x}, [\hat{y}^{(1)}, \ldots, \hat{y}^{(n)}]) \) be a Nash equilibrium of the WCI game \( G \). If the FDM condition holds for \( G \) and \( C^x_n \geq P^x \) for all \( n \), then the Nash equilibrium strategy of Player 1 (namely \( \hat{x} \)) is given by “waterfilling” against the combined noise and interference \( \alpha_n h^{(n)} \hat{y}^{(n)} + N_n \) from Player 2.

**Proof:** Let \( ([\hat{x}, [\hat{y}^{(1)}, \ldots, \hat{y}^{(n)}]]) \) be any saddle point of \( J \). The condition \( C^x_n \geq P^x \) ensures that the per-tone constraints are trivially satisfied whenever the power constraint \( (P^x) \) is. Evaluating the right-hand side of (2.14), if \( \beta_n = 0 \) (from FDM assumption) then

\[
R^* = \max_{x \in S_1} \sum_{n=1}^{N} \log \left( 1 + \frac{\hat{x}_n}{\alpha_n h^{(n)} \hat{y}^{(n)} + N_n} \right). \tag{2.15}
\]

The optimization problem (2.15) is seen to be precisely the same as single-user rate maximization with parallel Gaussian channels [22], and hence the (modified) water-filling spectrum is optimal and unique (for fixed \( [\hat{y}^{(1)}, \ldots, \hat{y}^{(n)}] \)). In particular, the modified AWGN noise level on tone \( n \) is seen to be \( \alpha_n h^{(n)} \hat{y}^{(n)} + N_n \). This is the same modified noise level used in the rate-adaptive IW algorithm [72].

Additional properties of both Player 1 and Player 2’s Nash equilibrium strategies are now developed. The following theorem shows that Player 1 always has a unique optimal strategy and that Player 2’s optimal strategies all cause the same amount of interference on the tones used by Player 1. This fact is somewhat surprising due to its asymmetry: while Player 1 is restricted to a single optimal strategy, Player 2
might have infinitely many optimal strategies. At the same time, all of Player 2’s optimal strategies cause the same interference to Player 1 on the tones it uses.

**Theorem 5** The Nash equilibrium strategy of Player 1 is unique; that is, there exists some \( \hat{x} \in S_1 \) such that for each \( (x, [y^{(1)}, \ldots, y^{(N)}]) \in P \), it is the case that \( x = \hat{x} \). Moreover, for Player 2, the induced “active” interference at each Nash equilibrium is unique; in particular, \( (x, [y^{(1)}, \ldots, y^{(N)}]) \), \( (\hat{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) implies that \( \alpha_n h^{(n)} y^{(n)} = \alpha_n h^{(n)} \tilde{y}^{(n)} \) for each \( n = 1, \ldots, N \) satisfying \( \hat{x}_n > 0 \).

**Proof:** To show that Player 1’s optimal strategy is identical for all Nash equilibria, consider the saddle points \( (x, [y^{(1)}, \ldots, y^{(N)}]) \in P \) and \( (\hat{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \), which are not necessarily distinct. By Theorem 2 and separability over tones, the objective (2.3) is strictly concave in \( x \) and therefore has a unique maximizer [8], namely \( \hat{x} \), when one fixes \( [y^{(1)}, \ldots, y^{(N)}] = [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \). Observe that \( (x, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) by the exchangeability property of saddle points [5]. Consequently, \( \hat{x} \) is also the unique maximizer of (2.3) for \( [y^{(1)}, \ldots, y^{(N)}] = [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \). This implies that \( x = \hat{x} \). Taking \( \hat{x} = \tilde{x} \) establishes the result.

To show the second claim, define \( I = \{i : \hat{x}_i > 0\} \) where \( \hat{x} \) is the unique Nash equilibrium strategy of Player 1 as per the first claim, and suppose that there exists a nonempty set \( D = \{n \in I : \alpha_n h^{(n)} y^{(n)} \neq \alpha_n h^{(n)} \tilde{y}^{(n)}\} \). Consider \( (\hat{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) and \( (\hat{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \), where \( \tilde{x} = \hat{x} = \tilde{x} \). Define \( S_2 \ni [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] = \frac{1}{2} [y^{(1)}, \ldots, y^{(N)}] + \frac{1}{2} [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \). The function \( g : \mathbb{R}_+^N \mapsto \mathbb{R}_+ \) defined by

\[
g([i_1, \ldots, i_N]) = \sum_{n=1}^N \log \left( 1 + \frac{\hat{x}_n}{i_n + \beta_n \tilde{x}_n + N_n} \right),
\]

is convex in each variable \( i_n \) and strictly convex in each variable \( i_n \) for which \( n \in I \) due to (2.13). By the fact that \( \emptyset \neq D \subset I \) and the convexity properties, it follows that

\[
g([\alpha_n h^{(1)} \tilde{y}^{(1)}, \ldots, \alpha_n h^{(N)} \tilde{y}^{(N)}]) < \frac{1}{2} g([\alpha_n h^{(1)} y^{(1)}, \ldots, \alpha_n h^{(N)} y^{(N)}])
\]

\[
+ \frac{1}{2} g([\alpha_n h^{(1)} \tilde{y}^{(1)}, \ldots, \alpha_n h^{(N)} \tilde{y}^{(N)}]),
\]

(2.17)
and consequently that

\[ J(\hat{x}, [\alpha_n h^{(1)} y^{(1)}(1), \ldots, \alpha_n h^{(N)} y^{(N)}]) < \frac{1}{2} J(\hat{x}, [\alpha_n h^{(1)} y^{(1)}(1), \ldots, \alpha_n h^{(N)} y^{(N)}]) + \frac{1}{2} J(\hat{x}, [\alpha_n h^{(1)} \tilde{y}^{(1)}(1), \ldots, \alpha_n h^{(N)} \tilde{y}^{(N)}]) = R^*, \quad (2.18) \]

which contradicts (2.8). Therefore \( D = \emptyset \).  

As a corollary Theorem 5 implies that the “interference profile” \( \alpha_n h^{(n)} y^{(n)} + \beta_n x_n + N_n \) is invariant on each active tone \( \{n : (x_n > 0)\} \) at every Nash equilibrium. Even though the Nash equilibrium need not be unique, one therefore has a strong sense in which to speak of a worst-case interference profile that is most deleterious to Player 1. This is illustrated in the following example

**Example 2** Let \( N = 3 \), \( L = 2 \), \( h^{(1)} = h^{(2)} = [1 1 0 0 0 0] \), \( h^{(3)} = [0 0 1 0 0 0] \), \( P^x = 1 \), \( P^y = [1 1 1]^T \), \( N_1 = N_2 = 1 \), \( N_3 = 1000 \), \( \alpha_1 = \alpha_2 = \alpha_3 = 1 \), \( \Gamma = 1 \), and suppose the FDM condition is satisfied and the per-tone power constraints are redundant.

These parameters may be interpreted as Player 2 controlling 3 modems, of which 2 crosstalk into tones 1 and 2 while the third crosstalks only into tone 3. Because the AWGN noise power seen by Player 1 on tone 3 (\( N_3 \)) is very high, it is intuitive that Player 1, if acting optimally, would not seek to use tone 3.

Indeed, one can show that \( x = [1/2 1/2 0]^T \), and \( y^{(1)} = [1 0 0 0 0 0]^T \), \( y^{(2)} = [0 1 0 0 0 0]^T \), \( y^{(3)} = [0 0 1 0 0 0]^T \) is a Nash equilibrium. It may also be verified that \( x = [1/2 1/2 0]^T \), \( y^{(1)} = y^{(2)} = [1/2 1/2 0 0 0 0]^T \) is a Nash equilibrium. Heuristically, this shows that because \( N_3 \) is so large, Player 1 will never use tone 3 and hence the behavior of the modem (controlled by Player 2) that crosstalks on tone 3 is irrelevant.

This example verifies the claim of Theorem 5 that the strategy of Player 1 is identical at both Nash equilibria. Furthermore, the interference profile

\[ \alpha_1 h^{(1)} y^{(n)} + \beta_1 x_1 + N_1 = \alpha_2 h^{(n)} y^{(2)} + \beta_2 x_2 + N_2 = 2, \quad (2.19) \]

is equal on both tones 1 and 2 at both Nash equilibria, as claimed. Yet, the interference
on tone 3 is not the same at both equilibria (equal to 1001 in the former example, 1000 in the latter). Thus the restriction of uniqueness of the worst interference profile to tones where $x_n > 0$ is not vacuous.

A condition ensuring the uniqueness of the worst-case interference on all tones is given in the following Theorem.

**Theorem 6** If $h^{(n)} > 0$ for all $n = 1, \ldots, N$ in the WCI game $G$, then for any $(\mathbf{x}, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]) \in P$ and $(\tilde{\mathbf{x}}, [\tilde{\mathbf{y}}^{(1)}, \ldots, \tilde{\mathbf{y}}^{(N)}]) \in P$ it holds $\alpha_n h^{(n)} \mathbf{y}^{(n)} = \alpha_n h^{(n)} \tilde{\mathbf{y}}^{(n)}$ for all $n = 1, \ldots, N$.

**Proof:** It was proven in Theorem 5 that $\mathbf{x} = \tilde{\mathbf{x}}$ and $\alpha_n h^{(n)} \mathbf{y}^{(n)} = \alpha_n h^{(n)} \tilde{\mathbf{y}}^{(n)}$ for all $n$ satisfying $\mathbf{x}_n > 0$. Therefore it remains only to show that $\alpha_n h^{(n)} \mathbf{y}^{(n)} = \alpha_n h^{(n)} \tilde{\mathbf{y}}^{(n)}$ for all $n$ satisfying $\mathbf{x}_n = 0$. In fact, one can prove a stronger version of this statement

$$\alpha_n h^{(n)} \mathbf{y}^{(n)} = \alpha_n h^{(n)} \tilde{\mathbf{y}}^{(n)} = 0,$$

for all $n$ satisfying $\mathbf{x}_n = 0$. This is shown by contradiction. The gist of the counterexample is that Player 2 could inflict strictly more harm by jamming on a tone that Player 1 is using.

Observe that the objective $J(\mathbf{x}, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}])$ is strictly increasing in each variable $x_n$ for any fixed $[\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}] \in S_2$. It claimed that $\sum_{n=1}^{N} \mathbf{x}_n > 0$. Indeed if $\sum_{n=1}^{N} \mathbf{x}_n = 0$ then

$$0 = J(\mathbf{x}, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]) < J\left(\frac{P^x}{N}, 1, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]\right),$$

which again a contradiction of (2.7).

Suppose that contrary to (2.20), $\alpha_n h^{(n')} \mathbf{y}^{(n')} > 0$ and $\mathbf{x}_{n'} = 0$ for some $n'$ where $1 \leq n' \leq N$. Then because $\alpha_{n'} \geq 0$ and $h^{(n')} \geq 0$ there must exist some $l'$, $1 \leq l' \leq 2L$

---

9Note that in DSL practice, despite the fact that the crosstalk may vary strongly as a function of frequency due to bridged taps and binder topology, it improbable that crosstalk is *completely* absent between any two given pairs. Therefore, the conditions of Theorem 6 are likely to apply in practice.
such that $\bar{y}^{(n')}_l > 0$. Define
\[
y^{(n)}_l = \begin{cases} 0 & n = n', \ l = l', \\
\bar{y}^{(n')}_l & \text{else.}
\end{cases}
\] (2.22)

Observe that because $\bar{x}_{n'} = 0$, it holds
\[
J(\bar{x}, [\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}]) = J(\bar{x}, [y^{(1)}, \ldots, y^{(N)}]).
\] (2.23)

Because $\bar{x} \succeq 0$ and $\sum_{n=1}^N x_n > 0$ there must exist some $j \neq n'$, $1 \leq j \leq N$ such that $\bar{x}_j > 0$. Next define
\[
y''^{(n)}_l = \begin{cases} y^{(j)}_l + \bar{y}^{(n')}_l & n = j, \ l = l', \\
y^{(n)}_l & \text{else.}
\end{cases}
\] (2.24)

and observe that $[y''^{(1)}, \ldots, y''^{(N)}] \in S_2$. However, recalling $\alpha_j = \Gamma / \tilde{h}_j > 0$ and $h^{(j)} > 0$ it is the case that
\[
\alpha_n h^{(j)} \gamma^{(j)} < \alpha_n h^{(j)} y''^{(j)}.
\] (2.25)

Recalling from above that $\bar{x}_j > 0$, one has
\[
J(\bar{x}, [y''^{(1)}, \ldots, y''^{(N)}]) < J(\bar{x}, [\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}])
\] (2.26)

which contradicts (2.7). Therefore $\alpha_n h^{(n)} \gamma^{(n)} = 0$ for all $n$. By the exact same argument applied to $(\tilde{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}])$, it follows that $\alpha_n h^{(n)} \tilde{y}^{(n)} = 0$ for all $n$. ■

Note that the requirement $h^{(n)} > 0$ excludes Theorem 6 from applying to the parameters of Example 2.

The previous two Theorems have characterized the interference induced by Player 2. By restricting attention to the FDM setting, it is possible to state explicitly the set of all Nash equilibria of $G$. Equivalently, it is possible to state all strategies of Player 2 that cause worst interferences, and optimal responses by Player 1 thereto. These strategies of Player 2 may be used by practitioners in offline system design or
dynamic operation; for example, it may be possible to operate each modem so that
the worst interference is avoided.

**Theorem 7** If the FDM condition is satisfied, then the set \( P \) of all Nash equilibria
of the WCI game \( G \) is a polytope\(^{10}\)

**Proof**: Given in Appendix C. \( \blacksquare \)

### 2.3.6 Numerical Computation of the Saddle Point

In order to apply the WCI bound in practical settings, it is necessary to develop
numerical algorithms to solve for Nash equilibrium strategies and \( R^* \). Two different
methodologies - dual decomposition and central path analysis - are presented for use
in numerically solving the WCI game.

The methodologies of these approaches are quite distinct. Computing saddle
points (maxmin points) of a function over a set can be significantly more difficult
than minimizing (or maximizing) a function over a set, both in theory and in practice.
For example, useful techniques such as gradient descent or Newton-based algorithms
may not be useful to solve maxmin problems. Also, many key results from nonlinear
optimization theory do not apply directly to maxmin problems.

The approach of the dual decomposition technique is illustrated on the right side of
Figure 2.2. First, the maxmin problem is converted into a single maximization prob-
lem of larger dimensionality. This larger maximization problem may then be solved
using standard techniques. The central path algorithm takes a different approach,
as illustrated on the left in Figure 2.2. One instead solves a sequence of perturbed
versions of the maximin problem. These perturbed maximin problems possess special
properties that allow them to be solved directly. Then, by reducing the amount of
the perturbation, one can approach the solution of the original problem to arbitrary
tolerance.

\(^{10}\)Different definitions of polytopes exist in the literature; this thesis defines a polytope as the
bounded intersection of a finite number of half-spaces [8].
Figure 2.2: High-level flowchart illustrating numerical algorithms used for computation of the Nash equilibria of $G$. 
Dual Decomposition Analysis

The first technique considered to compute $R^*$ is based on a “dual decomposition” that reduces the computation of a saddle point to a single maximization problem. This technique has a limited scope and is applicable subject to the condition $\beta_n = 0$ for all $n$. Recall that in the inner minimization in the left-hand side of (2.14), the objective is convex in each of the variables, and observe that a strictly feasible point exists; therefore Slater’s constraint-qualification condition is satisfied [8], and strong duality holds. Define $\lambda \in \mathbb{R}^{2L}_+$ as the dual variable associated with the vector inequality $[y^{(1)}, \ldots, y^{(N)}] 1 \preceq P^y$. The (partial) Lagrangian may be written as

$$L(x, [y^{(1)} \ldots y^{(N)}], \lambda) = \sum_{n=1}^{N} \log \left( 1 + \frac{x_n}{h^{(n)}y^{(n)} + N_n} \right) + \lambda^T [y^{(1)} \ldots y^{(L)}] 1 - P^y.$$  \hspace{1cm} (2.27)

By strong duality, the optimization (2.14) has the same optimal value as its Lagrangian dual. The Lagrangian dual problem is given by the following optimization problem

$$\max_{x, \lambda} \hspace{1cm} -\lambda^T P^y + \sum_{n=1}^{N} \min_{y^{(n)}} \log \left( 1 + \frac{x_n}{\alpha_n h^{(n)}y^{(n)} + \beta_n x_n + N_n} \right) + \lambda^T y^{(n)}$$

subject to $x \in S_1, 0 \preceq y^{(n)} \preceq C^{y,(n)} \forall n, \lambda \succeq 0.$  \hspace{1cm} (2.28)

Although (2.28) does have the same optimal value as (2.14), it is not necessarily the case that all optimal solutions $(x^*, \lambda^*)$ of (2.28) are optimal or even feasible\textsuperscript{11} for (2.14). It is now shown that that the inner minimization of (2.28) may be solved analytically. One proceeds by considering the following optimization problem

\textsuperscript{11}Uniqueness and feasibility could be guaranteed if, e.g., the objective were strictly convex in $\{y^{(n)}\}$. However, the optimization (2.28) will, in general, have an infeasible optimum with respect to (2.14).
\begin{align*}
\min_y \log \left( 1 + \frac{x}{c + a^T y} \right) + b^T y \\
\text{subject to } 0 \preceq y \preceq C(y).
\end{align*}

Where \( y \in \mathbb{R}^{N \times 1} \) are the decision variables, and \( x \in \mathbb{R}, a \in \mathbb{R}^{1 \times N}, b \in \mathbb{R}^{1 \times N}, \) and \( C(y) \in \mathbb{R}^{N \times 1} \) are constants satisfying \( x \geq 0, a \succeq 0, b \succeq 0, c > 0, \) and \( C(y) \succ 0. \)

**Theorem 8** The optimization problem (2.29) has an analytical solution \( y^* \).

**Proof:** Given in Appendix A.

**Definition 4** The optimal value of the optimization problem (2.29) is denoted by the function \( g(x, a, b, C(y), c) \), where \( g : \mathbb{R}_+ \times \mathbb{R}_+^2 \times \mathbb{R}_+^2 \times \mathbb{R}_+ \mapsto \mathbb{R} \) and

\begin{equation}
\begin{aligned}
g(x, a, b, C(y), c) &= \min_{0 \preceq y \preceq C(y)} \log \left( 1 + \frac{x}{c + a^T y} \right) + b^T y \\
&= \min_{0 \preceq y \preceq C(y)} \log \left( 1 + \frac{x}{c + a^T y} \right) + b^T y
\end{aligned}
\end{equation}

The inner minimization of (2.28) therefore consists of a total of \( N \) instances of the optimization (2.29) by taking \( x = x_n, a = \alpha_n (h^{(n)})^T \succeq 0, b = \lambda \succeq 0, c = N_n > 0, \) and \( C(y) = C_{y,(n)} \succ 0. \) Therefore by simplifying (2.28)

\begin{equation}
\begin{aligned}
\max_{x, \lambda} -\lambda^T P y + \sum_{n=1}^{N} g(x_n, \alpha_n (h^{(n)})^T, \lambda, C_{y,(n)}, N_n)
\end{aligned}
\end{equation}

\begin{equation}
\text{subject to } x \in S_1, \lambda \succeq 0.
\end{equation}

Now equivalently by defining

\begin{equation}
\begin{aligned}
d(x, \lambda) &= -\lambda^T P y + \sum_{n=1}^{N} g(x_n, \alpha_n (h^{(n)})^T, \lambda, C_{y,(n)}, N_n)
\end{aligned}
\end{equation}

and

\begin{equation}
\begin{aligned}
\mathcal{F} = \{(x^T, \lambda^T)^T : x \in S_1, \lambda \succeq 0\}.
\end{aligned}
\end{equation}
The optimization (2.31) may be written concisely as

$$\max_{x, \lambda} \ d(x, \lambda)$$

subject to

$$[x \lambda] \in \mathcal{F}. \quad (2.34)$$

A technique is now developed to solve the optimization problem (2.34). This is possible largely because (2.34) is a convex optimization problem. However, because the objective is non-differentiable, standard descent algorithms or even Newton-step algorithms cannot be employed. Instead, a technique known as subgradient descent is used to solve this concave non-differentiable optimization problem. The intuition of the algorithm is that an analogue of the gradient, called the subgradient, can be used with descent-like algorithms.

**Theorem 9** The following algorithm converges to an optimal solution of the optimization (2.34).

1. Initialize \( k = 1, x = 0, \) and \( \lambda = 0. \)

2. Set

$$\left( \begin{array}{c} x \\ \lambda \end{array} \right) := \mathcal{P}_\mathcal{F} \left( \left( \begin{array}{c} x \\ \lambda \end{array} \right) + \frac{1}{k} \left( \begin{array}{c} \alpha_1 h^{(1)} y^{(1)} \star + N_1 \\ \vdots \\ \alpha_N h^{(N)} y^{(N)} \star + N_N \end{array} \right) \right)$$

and \( k = k+1, \) where \( \mathcal{P}_\mathcal{F} \) denotes Euclidean projection on \( \mathcal{F} \) and \([y^{(1)} \star, \ldots, y^{(N)} \star]\) minimizes (2.30) in \( d(x, \lambda) \) (Theorem 8).

3. Return to Step 2.

**Proof:** Observe that the set \( \mathcal{F} \) is closed and therefore the Euclidean projection on \( \mathcal{F} \) is well-defined and unique; moreover, computing the projection is a convex optimization problem [7, Prop. 2.1.3]. By applying the algorithm of [7, Ex. 6.3.13], [19] it is sufficient to show the following technical conditions.
1. \( d(\mathbf{x}, \lambda) > -\infty \) for all \( \mathbf{x} \in \mathcal{F} \). The value of the optimization (A.1) is bounded because the feasible set is bounded and the objective is continuous on the feasible set. Therefore by (2.32), \( d(\mathbf{x}, \lambda) > -\infty \).

2. \( \sum_{k=1}^{\infty} \frac{1}{k} = \infty \). This is immediate.

3. The vector
   \[
   \left( \frac{\alpha_1 h^{(1)}(y^{(1)}, \star, +1)}{x_1 \alpha_1 h^{(1)}(y^{(1)}, \star, +1) + N_1}, \ldots, \frac{\alpha_N h^{(N)}(y^{(N)}, \star, +1)}{x_N \alpha_N h^{(N)}(y^{(N)}, \star, +1) + N_N} \right)
   \]
   is a supergradient of \( d(\mathbf{x}, \lambda) \) at \( (\mathbf{x}, \lambda) \). This is proven in Lemma 6.

4. Let \( \{(\mathbf{x})_n\} \) be any sequence in \( \mathcal{S}_1 \) and \( \{\lambda_n\} \) be any sequence in \( \mathbb{R}_+^{2L} \). Then it holds
   \[
   \frac{1}{k} \left\| \left( \begin{array}{c}
   \frac{\alpha_1 h^{(1)}(y^{(1)}, \star, +1)}{(x_1)_k \alpha_1 h^{(1)}(y^{(1)}, \star, +1) + N_1} \\
   \vdots \\
   \frac{\alpha_N h^{(N)}(y^{(N)}, \star, +1)}{(x_N)_k \alpha_N h^{(N)}(y^{(N)}, \star, +1) + N_N} \\
   \end{array} \right) \right\|_2^2 \to 0 \quad (2.36)
   \]
as \( k \to \infty \) where \( (y^{(n)}, \star)_k \) is as defined in Lemma 6 corresponding to \( (\mathbf{x})_k \) and \( \lambda_k \). It is proven in Lemma 7 that the supergradient is bounded, which immediately implies the desired condition.

5. Let \( \{(\mathbf{x})_n\} \) be any sequence in \( \mathcal{S}_1 \) and \( \{\lambda_n\} \) be any sequence in \( \mathbb{R}_+^{2L} \). Then it holds
   \[
   \sum_{n=1}^{\infty} \frac{1}{k^2} \left\| \left( \begin{array}{c}
   \frac{\alpha_1 h^{(1)}(y^{(1)}, \star, +1)}{(x_1)_k \alpha_1 h^{(1)}(y^{(1)}, \star, +1) + N_1} \\
   \vdots \\
   \frac{\alpha_N h^{(N)}(y^{(N)}, \star, +1)}{(x_N)_k \alpha_N h^{(N)}(y^{(N)}, \star, +1) + N_N} \\
   \end{array} \right) \right\|_2^2 < \infty. \quad (2.37)
   \]
It is proven in Lemma 7 that the supergradient is bounded, which along with
the fact that $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$ implies the desired condition.

Central-Path Analysis

The second technique presented is based upon interior-point optimization techniques such as the “Infeasible Start Newton method” [8, §10.3]. The general approach of interior-point techniques is to replace the (power and positivity) constraints with barrier functions that become large as the (power and positivity) constraints become tight. By making the increase in the barrier functions progressively sharper, one solves a sequence of problems whose solutions converge to a Nash equilibrium of $G$. This sequence of solution is commonly referred to as a “central path” through the feasible region, and converges to an optimum point as $t \to \infty$.

Central path optimization techniques are well-studied in the optimization literature, having been applied to e.g. linear programming (LP), semidefinite programming (SDP), the linear complementary problem (LCP) [50], as well as general convex optimization problems [8]. We now formally cast the problem (2.14) in the interior interior-point setting and prove that it satisfies certain necessary properties needed for convergence. This method will be shown to allow one to compute $R^*$ to arbitrary accuracy.

Logarithmic barrier functions are employed to enforce the positivity and power constraints. Let the central path parameter be denoted by $t \in \mathbb{R}_{++}$ and define
\( \tilde{S}_1 = \text{int}(S_1), \tilde{S}_2 = \text{int}(S_2), \) and \( \tilde{J} : \tilde{S}_1 \times \tilde{S}_2 \mapsto \mathbb{R}_+, \) where

\[
\tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) = t^{-1} \log \left( P^x - \sum_{n=1}^{N} x_n \right) + \sum_{n=1}^{N} t^{-1} \log (C_n^x - x_n)
\]

\[
+ \sum_{n=1}^{N} \left\{ \log \left( 1 + \frac{x_n}{\alpha_n h^{(n)}y^{(n)} + \beta_n x_n + N_n} \right) + t^{-1} \log(x_n) \right\}
\]

\[
- t^{-1} \sum_{l=1}^{2L} \left[ \log(y_l^{(n)}) + \log(C_l^{y,(n)} - y_l^{(n)}) \right]
\]

\[
- t^{-1} \sum_{l=1}^{2L} \log \left( P_l^y - \sum_{n=1}^{N} y_l^{(n)} \right). \tag{2.38}
\]

Instead of solving (2.4) directly, the central path technique proceeds by solving for a saddle point of \( \tilde{J} \) as in the following modified problem

max \( x \) \( \text{min}_{\{y^{(n)}\}} \) \( \tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) \)

subject to

\( x \succ 0, x \prec C^x \)

\( y^{(n)} \succ 0, y^{(n)} \prec C^{y,(n)} \) \( n = 1, \ldots, N \) \( \tag{2.39} \)

\( 1^T x < P^x, \)

\( [y^{(1)} \ldots y^{(N)}]1 \prec P^y. \)

The following two definitions are required in the proof of the following theorem.

**Definition 5** Consider a function \( f : \mathbb{R} \mapsto \mathbb{R} \), where \( R \subset \mathbb{R}^p \). The function \( f \) is said to be Lipschitz continuous on \( R \) (with coefficient \( L \)) if there exists some \( L < \infty \) such that

\[
||f(z) - f(z')||_2 \leq L||z - z'||_2 \tag{2.40}
\]

for all \( z, z' \in R \).

**Definition 6** Consider a function \( f : \mathbb{R} \mapsto \mathbb{R} \), where \( R \subset \mathbb{R}^p \) having Hessian \( \nabla^2 f(r) \). The function \( f \) is said to have a Lipschitz continuous Hessian on \( R \) (with coefficient
if there exists some \( L < \infty \) such that

\[
\|\nabla^2 f(z) - \nabla^2 f(z')\|_2 \leq L \|z - z'\|_2
\]

for all \( z, z' \in \mathbb{R} \).

**Theorem 10** For any fixed \( t > 0 \), the interior-point Newton step algorithm given in [8, §10.3.4] converges to a saddle point of \( \tilde{J} \); furthermore, this saddle point solves (2.39).

**Proof:** The function \( \tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) \) is convex in \( x \) and jointly concave in \([y^{(1)}, \ldots, y^{(N)}]\) on \( \tilde{S}_1 \times \tilde{S}_2 \) because it is the sum of functions convex in \( x \) and jointly concave in \([y^{(1)}, \ldots, y^{(N)}]\) on \( \tilde{S}_1 \times \tilde{S}_2 \). Given that \( \tilde{J} \) has such a convex-concave structure, convergence of the algorithm to a point satisfying

\[
\nabla \tilde{J}(x', [y'^{(1)}, \ldots, y'^{(N)}]) = 0
\]

may be established by showing that \( \tilde{J} \) satisfies the following technical conditions [8, §10.3.4]

1. The sublevel sets of \( \|
abla \tilde{J}\|_2 \) are closed; this is proven in Lemma 8.
2. The function \( \tilde{J} \) has a Lipschitz continuous Hessian (see Definition 6); this is proven in Lemma 12.
3. The function \( \tilde{J} \) is strongly convex-concave; this is proven in Lemma 17.

However, the condition (2.42) is precisely the first-order optimality conditions [8, §3.1.3] showing that \([y'^{(1)}, \ldots, y'^{(N)}]\) is optimal for the optimization problem

\[
\min_{[y^{(1)}, \ldots, y^{(N)}]} \tilde{J}(x', [y^{(1)}, \ldots, y^{(N)}])
\]

subject to \([y^{(1)}, \ldots, y^{(N)}] \in \tilde{S}_2 \).
and $\mathbf{x}'$ is optimal for

$$\max_{\mathbf{x}} \tilde{J}(\mathbf{x}, [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}])$$

subject to $\mathbf{x} \in \tilde{S}_1$. \hspace{1cm} (2.44)

The optimization problems (2.43) and (2.44) are immediately equivalent to the saddle point condition

$$\tilde{J}(\mathbf{x}', [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}]) \geq \tilde{J}(\mathbf{x}', [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}]) \geq \tilde{J}(\mathbf{x}, [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}]) \hspace{1cm} (2.45)$$

for all $\mathbf{x}' \in \tilde{S}_1$ and $[\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}] \in \tilde{S}_2$. It is proven in Appendix B that the saddle point $(\mathbf{x}', [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}])$ solves (2.39).

The introduction of barrier functions causes inaccuracy in the sense that for a given value of $t$, the values of the Worst Case Interference game $R^*$ (2.4) and the value at the saddle point of the modified objective (2.39) may differ. The issue of the error introduced by such an approximation is rigorized in the following Theorem.

As the central path parameter $t \to \infty$, the barrier functions heuristically approach “brick walls” that become pointwise negligible on the interior of the feasible set, yet rapidly approach infinity near the boundary. The following theorem not only proves convergence of the modified objective saddle point to $R^*$, but further gives the order of the convergence.

**Theorem 11** Let $R : \mathbb{R}_{++} \mapsto \mathbb{R}$ denote the optimal value of the problem (2.39) as a function of $t$. Then it holds

$$R(t) - \frac{L(1+4N)}{t} \leq R^* \leq R(t) + \frac{2N+1}{t} \hspace{1cm} (2.46)$$

**Proof:** See Appendix B.
2.4 The Multiuser Worst Case Interference

The analysis of Section 2.3 developed lower bounds for the achievable rates of individual users in a multiuser binder. In this section, these per-user lower bounds are extended to the multiuser setting by considering lower bounds on the weighted rate-sum of all users when all act greedily. Some intuition for the problem is first developed, which is then formalized and interpreted.

2.4.1 Motivation

Heuristically speaking, if \( R'_k, k = 1, \ldots, L + 1 \) denotes a rate that user \( k \) can achieve by greedily maximizing its own rate when all other users act similarly and \( \mu \in \mathbb{R}^{L+1}_+ \), the construction of the Multiuser Worst Case Interference seeks a lower bound \( L \) to the weighted sum rate

\[
L \leq \sum_{k=1}^{L+1} \mu_k R'_k. \tag{2.47}
\]

For example, by taking \( \mu = 1 \), a lower (achievable) bound to the interference channel sum capacity is computed. Other choices of \( \mu \) yield different bounds.

Observe that one may obtain a trivial bound of the form (2.47) as follows: suppose that \( R'_k^* \) is the WCI rate of Section 2.3 when user \( k \) is considered as the victim modem and all other modems are considered as interfering modems. The rectangular region \( \mathcal{A} \) in Figure 2.3 is therefore achievable. Because \( R'_k \geq R'_k^* \), it is immediate that

\[
\sum_{k=1}^{L+1} \mu_k R'_k \geq \sum_{k=1}^{L+1} \mu_k R'_k^* = L, \tag{2.48}
\]

gives a lower bound. This situation is illustrated in Figure 2.3 for the case of \( L+1 = 2 \).

Visually, for any \( \mu \geq 0, \mu \neq 0 \) the inner product of \( \mu \) and the intersection point of the achievable region \( \mathcal{A} \) with the line \( A \) (having normal vector \( \mu \)) provides a lower bound to \( \sum_{k=1}^{L+1} \mu_k R'_k \). The multiuser WCI analysis gives a nontrivial strengthening of this bound, illustrated by line \( B \); in general, \( B \) may be strictly to the right of \( A \). Note
that any point inside the rectangular region can always be achieved, and furthermore at least one point on $B$ must also be achievable by greedy algorithms. Therefore, the interference channel capacity region $\mathcal{C}_{IC}$ (the outer curved region) must contain at least one point on $B$ (that is, $B \cap \mathcal{C}_{IC} \neq \emptyset$). However, the capacity region may extend beyond $B$ as shown.

### 2.4.2 Formalization of the Multiuser WCI Game

For fixed user weighting $\mu \in \mathbb{R}_{+}^{L+1}$, $\mu \neq 0$, sum power constraints $\mathbf{P}^y \in \mathbb{R}_+^{2(L+1)}$, PSD constraints $\mathbf{C}^{(n)} \in \mathbb{R}_+^{2(L+1)}$, $n = 1, \ldots, N$, interference constants $\alpha_k^{(n)} \geq 0$, $\beta_k^{(n)} \geq 0$, $\sigma_k^{(n)} > 0$, $k = 1, \ldots, L + 1$, $n = 1, \ldots, N$, and crosstalk gains $h^{(n),k} \geq 0$, $n = 1, \ldots, N$, $k = 1, \ldots, L$, let

$$D^{(n)} = [\mathbf{C}_1^{(n)}, \ldots, \mathbf{C}_{L+1}^{(n)}]^T, \quad n = 1, \ldots, N,$$  

$$\mathbf{P}^x = [\mathbf{P}_1^y, \ldots, \mathbf{P}_{L+1}^y]^T.$$  

\[2.49\]  

\[2.50\]
CHAPTER 2. THE WORST-CASE INTERFERENCE

40

Table 2.2: Summary of Multiuser Worst Case Interference game $\mathcal{H}$.

<table>
<thead>
<tr>
<th>Player</th>
<th>Controls</th>
<th>Objective</th>
<th>Strategy Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All Modems max $K$</td>
<td>${[x^{(1)}, \ldots, x^{(N)}] : 0 \leq x^{(n)} \leq D^{(n)} \forall n, \ [x^{(1)}, \ldots, x^{(N)}] \mathbf{1} \leq P^x}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Fictitious Interference min $K$</td>
<td>${[y^{(1)}, \ldots, y^{(N)}] : 0 \leq y^{(n)} \leq C^{(n)} \forall n, \ [y^{(1)}, \ldots, y^{(N)}] \mathbf{1} \leq P^y}$</td>
<td></td>
</tr>
</tbody>
</table>

Define the sets

$$Q_1 = \{[x^{(1)}, \ldots, x^{(N)}] \in \mathbb{R}^{L+1 \times N} : x^{(n)} \succeq 0, [x^{(1)}, \ldots, x^{(N)}] \mathbf{1} \preceq P^x, \ x^{(n)} \preceq D^{(n)}, n = 1, \ldots, N\},$$

$$Q_2 = \{[x^{(1)}, \ldots, x^{(N)}] \in \mathbb{R}^{2(L+1) \times N} : x^{(n)} \succeq 0, [x^{(1)}, \ldots, x^{(N)}] \mathbf{1} \preceq P^y, \ x^{(n)} \preceq C^{(n)}, n = 1, \ldots, N\}. \quad (2.51)$$

Define the map $K : Q_1 \times Q_2 \mapsto \mathbb{R}_+$ as

$$K([x^{(1)}, \ldots, x^{(N)}], [y^{(1)}, \ldots, y^{(N)}]) = \sum_{k=1}^{L+1} \mu_k \sum_{n=1}^{N} \log \left(1 + \frac{x_k^{(n)}}{\beta_k^{(n)} x_k^{(n)} + \alpha_k^{(n)} h_{(n),k}^{(n)} y_k^{(n)} + \sigma_k^{(n)}}\right). \quad (2.53)$$

The strictly-competitive game $\mathcal{H} = (Q_1, Q_2, K)$ is defined to be the Multiuser Worst Case Interference game. A summary is given in Table 2.2. As in the WCI game, an FDM condition will be said to hold if $\beta_k^{(n)} = 0, k = 1, \ldots, L + 1, n = 1, \ldots, N$ and $h_{l,1}^{(n),k} = 0, l = L + 1, \ldots, 2L, k = 1, \ldots, L + 1, n = 1, \ldots, N$.

**Theorem 12** The game $\mathcal{H}$ has a Nash equilibrium in pure strategies and a value, denoted $R^\dagger$.

**Proof:** By Lemma 1, each term in the (non-negative) weighted sum of (2.53) is concave in $[x^{(1)}, \ldots, x^{(N)}]$ for fixed $[y^{(1)}, \ldots, y^{(N)}]$ and convex in $[y^{(1)}, \ldots, y^{(N)}]$ for fixed $[x^{(1)}, \ldots, x^{(N)}]$, the weighted sum $K$ is as well. Since $Q_1$ and $Q_2$ are closed and bounded, they are compact, and the function $K$ is continuous on its domain. By [5, Thm. 4.4], the result follows.
Recall that a pure-strategy Nash equilibrium of $\mathcal{K}$ is a saddle point of $K$ satisfying the saddle point conditions (i.e. (2.7) and (2.8)). The computation of such a Nash equilibrium of $\mathcal{K}$ may be expressed as

$$\max_{\{x^{(n)}\}} \min_{\{y^{(n)}\}} \sum_{k=1}^{L+1} \mu_k \sum_{n=1}^{N} \log \left( 1 + \frac{x_k^{(n)}}{\beta_k^{(n)} x_k^{(n)} + \alpha_k^{(n)} y_k^{(n)} + \sigma_k^{(n)}} \right)$$

subject to

$$[x^{(1)}, \ldots, x^{(N)}] \in \mathcal{Q},$$
$$[y^{(1)}, \ldots, y^{(N)}] \in \mathcal{Q}. \tag{2.54}$$

The Multiuser WCI game also provides a lower bound to the weighted sum rate of each user in the $L + 1$-player noncooperative IW game.

**Theorem 13** In the noncooperative IW game [58], a (possibly non-unique) Nash equilibrium is known to always exist in pure strategies. For a given Nash equilibrium of the $(L + 1)$-player IW game, let user $k$ attain rate $\tilde{R}_k$. Then for any $\mu \succeq 0$ under the FDM condition, it holds

$$\sum_{k=1}^{L+1} \mu_k \tilde{R}_k \geq R^1. \tag{2.55}$$

where $R^1$ is as defined in Theorem 12.

**Proof:** The result is an immediate consequence of the Nash equilibrium condition (2.8).

---

2.4.3 Interpretation of the Multiuser WCI Game

The Multiuser Worst Case Interference game is a generalization of the Worst Case Interference game\footnote{The WCI game falls out by taking $\mu = e_k$ and eliminating irrelevant variables.}. However compared to the WCI game, the interpretation of the variables of this game is somewhat less straightforward.

The constants $h^{(n),k} \geq 0$, $\alpha^{(n),k} \geq 0$, $\beta^{(n),k} \geq 0$, and $\sigma^{(n),k} \geq 0$ may be interpreted in precisely the same manner as in the WCI game, except that they are indexed by user $k$. For example, in the Multiuser WCI game, $\sigma^{(n),k}$ denotes additive noise seen by user
$k$ on tone $n$, whereas in the WCI game, $\sigma_k$ denotes additive noise seen by the victim modem on tone $n$. The non-negative vector $\mu$ denotes a weighting on the users’ rates as discussed in Section 2.4.1.

The variables $[x^{(1)}, \ldots, x^{(N)}]$ denote the power allocation of all the modems when they act greedily. Note that the first $L + 1$ elements of $P^y$ constrain power in the direction of transmission, while the final $L + 1$ elements constrain power in the reverse direction. The variables $[y^{(1)}, \ldots, y^{(N)}]$ do not have a physical manifestation, but may be thought of as a fictitious power allocation that would be malignant in the sense of minimizing a weighted sum of the users’ rates. A Nash equilibrium of $\mathcal{H}$ represents a competitively optimal strategy for each user when experiencing an interference profile generated by the fictitious power allocation $[y^{(1)}, \ldots, y^{(N)}]$.

## 2.5 Lemmata and Proofs

This section comprises proofs of the technical results needed to establish convergence of the two techniques presented for numerically solving for the WCI. These results are used explicitly in proving Theorems 9 and 10. Section 2.5.1 considers proofs related to the subgradient descent algorithm, while Section 2.5.2 considers the central path algorithm. Additional proofs may be found in the Appendices as referenced.

### 2.5.1 Dual Decomposition Analysis

**Lemma 2** The function $g(x, a, b, C^{(y)}, c)$ (as given in Definition 4) is jointly concave in $x$ and $b$ for fixed $a$, $C^{(y)}$, and $c$.

**Proof:** Observe that the function $\phi : \mathbb{R}_+ \times \mathbb{R}_+^{2L} \times \mathbb{R}_+^{2L} \times \mathbb{R}_+ \times \mathbb{R}_+^{2L} \mapsto \mathbb{R}_+$, defined by

$$
\phi(x, a, b, c, y) = \log \left( 1 + \frac{x}{c + a^T y} \right) + b^T y,
$$

is jointly concave in $x$ and $b$ for each fixed $a$, $c$, and $y$. This holds because $\log(\cdot)$ is concave on $\mathbb{R}_+$, $b^T y$ is linear (therefore concave), and so the sum of (jointly) concave
functions is again (jointly) concave. Observe further that \( \phi(x, a, b, c, y) \) is continuous in \( x, b, \) and \( y \). Finally, the set \( \{ y : 0 \preceq y \preceq C^y \} \) is a closed and bounded subset of \( \mathbb{R}^N \), and is therefore compact.

The conditions of Danskin’s theorem [7, Prop. B.25] are thereby satisfied, whence

\[
g(x, a, b, c) = \min_{0 \preceq y \preceq C^y} \phi(x, a, b, c, y) \quad (2.57)
\]

is (jointly) concave in \( x \) and \( b \) for fixed \( a, c \).

**Lemma 3** Let \( a \) be a concave map \( a : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R} \). We shall write \( a(\eta, \nu) \) where \( \eta \in \mathbb{R}^n, \nu \in \mathbb{R}^m \). For each \( \eta \in \mathbb{R}^n \), define the map \( b_\eta : \mathbb{R}^m \mapsto \mathbb{R}_+ \), written \( b_\eta(\nu) \), as \( b_\eta(\nu) = a(\eta, \nu) \). Similarly, for each \( \nu \in \mathbb{R}^m \), define the map \( c_\nu : \mathbb{R}^n \mapsto \mathbb{R}_+ \), written \( c_\nu(\eta) \), as \( c_\nu(\eta) = a(\eta, \nu) \).

If \( \bar{g} \) is a supergradient of \( b_\eta \) at \( \nu \) and \( \bar{g} \) is a supergradient of \( c_\nu \) at \( \eta \), then \( \begin{bmatrix} \eta \\ \nu \end{bmatrix} \) is a supergradient of \( a \) at \( \begin{bmatrix} \eta \\ \nu \end{bmatrix} \).

**Proof:** Indeed by the definition of the supergradient,

\[
a(\eta', \nu') = b_{\eta'}(\nu') \\
geq b_{\eta'}(\nu) + (\nu' - \nu)^T \bar{g} \\
= c_\nu(\eta') + (\nu' - \nu)^T \bar{g} \\
geq c_\nu(\eta) + (\eta' - \eta)^T \bar{g} + (\nu' - \nu)^T \bar{g} \\
= a(\eta, \nu) + \left( \begin{bmatrix} \eta' \\ \nu' \end{bmatrix} - \begin{bmatrix} \eta \\ \nu \end{bmatrix} \right)^T \begin{bmatrix} \bar{g} \\ \bar{g} \end{bmatrix}. \quad (2.58)
\]

**Lemma 4** A supergradient \( g_x \) of \( d(x, \lambda) \) (as defined in (2.32)) in \( x \) for fixed \( \lambda \) is
given by

\[
\mathbf{g}_\mathbf{x} = \begin{pmatrix}
\frac{\alpha_1 h(1) y(1),\star + N_1}{x_1 \alpha_1 h(1) y(1),\star + N_1} \\
\vdots \\
\frac{\alpha_N h(N) y(N),\star + N_N}{x_N \alpha_N h(N) y(N),\star + N_N}
\end{pmatrix}
\] (2.59)

where \( y^{(n),\star} \) achieves the minimum in (2.30) corresponding to tone \( n \).

**Proof:** As in the proof of Lemma 2, the conditions of Danskin’s theorem are satisfied for the minimization in (2.57). Observe that \( \phi \) is differentiable with respect to \( x \) for fixed \( a, b, C(y), c, \) and \( y \). Therefore as a consequence of Danskin’s theorem [7, Prop. B.25], a supergradient of \( g \) (2.30) in \( x \) for fixed \( a, b, C(y), \) and \( c \) is given by the partial derivative of \( \phi \) with respect to \( x \) at any \( y^\star \) achieving the minimum in (2.30). In particular,

\[
\frac{\partial}{\partial x} \phi(x, a, b, C(y), c, y) = \frac{\mathbf{a}^T \mathbf{y} + c}{x \mathbf{a}^T \mathbf{y} + c}.
\] (2.60)

By substituting \( x = \mathbf{x}_n, a = \alpha_n (h^{(n)}), b = \lambda, c = N_n, y = y^{(n),\star} \) and \( C_y = C_y^{(n)} \) as per (2.31), the result is obtained.

**Lemma 5** A supergradient \( \mathbf{g}_\lambda \) of \( d(\mathbf{x}, \lambda) \) (as defined in (2.32)) in \( \lambda \) for fixed \( x \) is given by

\[
\mathbf{g}_\lambda = [y^{(1),\star}, \ldots, y^{(N),\star}] \mathbf{1} - \mathbf{P}^y,
\] (2.61)

where \( y^{(n),\star} \) achieves the minimum in (2.30) corresponding to tone \( n \).

**Proof:** This is a standard result in convex optimization theory; a proof following [7] is reproduced here. Recalling our terminology for convex optimization problems
(Section 1.5), we without loss of generality\(^\text{13}\) consider problems without equality constraints, where \(x_\mu\) is any point minimizing the Lagrangian at \(\lambda\)

\[
x_\mu = \arg \min_{x \in X} \mathcal{L}(x, \lambda).
\]

(2.62)

It then follows that the “vector of infeasibilities” \([f_1(x), \ldots, f_m(x)]^T\) is a subgradient of the dual function at \(\lambda\) because

\[
g(\lambda) = \inf_{x \in X} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)
\leq f_0(x_\mu) + \sum_{i=1}^m \lambda_i f_i(x_\mu)
= f_0(x_\mu) + \sum_{i=1}^m \lambda_i f_i(x_\mu) + \sum_{i=1}^m (\lambda - \lambda_i) f_i(x_\mu)
= g(\lambda) + \sum_{i=1}^m (\lambda - \lambda_i) f_i(x_\mu).
\]

(2.63)

Because \(d(x, \lambda)\) is the Lagrangian (2.27) for fixed \(x\), the result is immediate. \(\blacksquare\)

**Lemma 6** A supergradient \(g\) of \(d(x, \lambda)\) is given by

\[
g = \begin{pmatrix} g_x \\ g_\lambda \end{pmatrix} = \begin{pmatrix} \frac{\alpha_1 h^{(1)} \mathbf{y}^{(1)} + N_1}{x_1 \alpha_1 h^{(1)} \mathbf{y}^{(1)} + N_1} \\ \vdots \\ \frac{\alpha_N h^{(N)} \mathbf{y}^{(N)} + N_N}{x_N \alpha_N h^{(N)} \mathbf{y}^{(N)} + N_N} \end{pmatrix} \text{[}\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]\text{1} - \mathbf{P}^y.
\]

(2.64)

where \(\mathbf{y}^{(n),*}\) achieves the minimum in (2.30) corresponding to tone \(n\).

**Proof:** By Lemma 2, \(g(x_n, \alpha_n(h^{(n)})^T, \lambda, C^{(n)}, N_n)\) is jointly concave in \(x_n\) and \(\lambda\). By definition, \(d(x, \lambda)\) is the (finite) sum of functions that are jointly concave in \(x\) and \(\lambda\), and therefore is itself jointly concave in \(x\) and \(\lambda\). Consequently by applying Lemma 3 to the subgradients \(g_x\) and \(g_\lambda\) from Lemmas 4 and 5, respectively, the result is established. \(\blacksquare\)

---

\(^\text{13}\)Any equality constraint may be transformed into two inequality constraints.
Lemma 7 There exists some \( B < \infty \) such that for all \( \begin{bmatrix} x \\ \lambda \end{bmatrix} \in \mathcal{F} \) (where \( \mathcal{F} \) is defined in (2.33)) it holds

\[
\left\| \begin{pmatrix} \frac{a_1 h^{(1)} y^{(1), \ast} + N_1}{x_1 a_1 h^{(1)} y^{(1), \ast} + N_1} \\ \vdots \\ \frac{a_N h^{(N)} y^{(N), \ast} + N_N}{x_N a_N h^{(N)} y^{(N), \ast} + N_N} \\ [y^{(1), \ast}, \ldots, y^{(N), \ast}] [1 - P^y] \end{pmatrix} \right\|_2 \leq B,
\]

where \( y^{(n), \ast} \) is as defined in Lemma 6.

Proof: By assumption, \( x \in S_1 \). For any \( \lambda \), one has \( [y^{(1), \ast}, \ldots, y^{(n), \ast}] \in S_2 \) due to Lemma 6. Observe that the function \( f : S \mapsto \mathbb{R}^{N+2L} \) with

\[
f(x, [y^{(1)}, \ldots, y^{(n)}]) = \begin{pmatrix} \frac{a_1 h^{(1)} y^{(1), \ast} + N_1}{x_1 a_1 h^{(1)} y^{(1), \ast} + N_1} \\ \vdots \\ \frac{a_N h^{(N)} y^{(N), \ast} + N_N}{x_N a_N h^{(N)} y^{(N), \ast} + N_N} \\ [y^{(1)}, \ldots, y^{(N)}] [1 - P^y] \end{pmatrix}
\]

is continuous on \( S \), the norm \( \| \cdot \|_2 \) is continuous, and therefore their composition is continuous. Because \( S \) is compact, it follows

\[
\sup_{(x, [y^{(1)}, \ldots, y^{(n)}]) \in S} \left\| \begin{pmatrix} \frac{a_1 h^{(1)} y^{(1), \ast} + N_1}{x_1 a_1 h^{(1)} y^{(1), \ast} + N_1} \\ \vdots \\ \frac{a_N h^{(N)} y^{(N), \ast} + N_N}{x_N a_N h^{(N)} y^{(N), \ast} + N_N} \\ [y^{(1)}, \ldots, y^{(N)}] [1 - P^y] \end{pmatrix} \right\|_2 < \infty,
\]

which implies the result.

2.5.2 Central Path Algorithm

Lemma 8 The sublevel sets of the function \( \| \nabla \tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) \|_2 \), where \( \tilde{J} \) is as defined in (2.38), are closed.
CHAPTER 2. THE WORST-CASE INTERFERENCE

Proof: Evaluating the partial derivatives of \( \tilde{J} \), one obtains

\[
\frac{\partial \tilde{J}}{\partial x_n}(x, [y^{(1)}], \ldots, y^{(N)}) = \alpha_n h^{(n)} y^{(n)} + N_n \frac{1}{t x_n} \left( \beta_n x_n + \alpha_n h^{(n)} y^{(n)} + N_n \right) + \frac{1}{t(P x - 1)^T x} - \frac{1}{t(C_n^z - x_n)},
\]

\( (2.68) \)

\[
\frac{\partial \tilde{J}}{\partial (y_m^{(n)})}(x, [y^{(1)}], \ldots, y^{(N)}) = \frac{\alpha_n h^{(n)}}{(1 + \beta_n)x_n + \alpha_n h^{(n)}y^{(n)} + N_n} - \frac{\alpha_n h^{(n)}}{(1 + \beta_n)x_n + \alpha_n h^{(n)}y^{(n)} + N_n} \frac{1}{ty_m^{(n)}} + \frac{1}{t(C_m^{y^{(n)}} - y_m^{(n)})}.
\]

\( (2.69) \)

Since each partial derivative is a continuous function, the continuity of the norm \( \| \cdot \|_2 \) implies that the composition \( \| \nabla \tilde{J} \|_2 \) is continuous on \( \tilde{S}_1 \times \tilde{S}_2 \). Consequently, the sublevel sets \( S_\alpha \) for each \( \alpha \in \mathbb{R} \)

\[ S_\alpha = \{(x, [y^{(1)}], \ldots, y^{(N)}) \in \tilde{S}_1 \times \tilde{S}_2 : \| \nabla \tilde{J}((x, [y^{(1)}], \ldots, y^{(N)}))\|_2 \leq \alpha \} \]

\( (2.70) \)

are closed relative to \( \tilde{S}_1 \times \tilde{S}_2 \) because they are the preimage of the closed set \( \{z \in \mathbb{R} : z \leq \alpha \} \) under a continuous map \( [59] \).

To show that \( S_\alpha \) is closed, we proceed from the sequential definition of set closure. Suppose in particular that \( \{(z)_n\} \) is any sequence in \( S_\alpha \) with \( (z)_n \rightharpoonup z \). To show closure of \( S_\alpha \), we must show that \( z \in S_\alpha \). If \( z \in \tilde{S}_1 \times \tilde{S}_2 = \text{int}(\tilde{S}_1 \times \tilde{S}_2) \), then \( z \in S_\alpha \) by the sequential definition of relative closure. Therefore it remains only to prove that there does not exist any sequence \( \{(z)_n\} \) in \( S_\alpha \) such that \( (z)_n \rightharpoonup z \) with
CHAPTER 2. THE WORST-CASE INTERFERENCE

\[ z \in \partial \text{cl}(\tilde{S}_1 \times \tilde{S}_2). \] The proof is by contradiction. The boundary may be written as

\[
\partial \text{cl}(\tilde{S}_1 \times \tilde{S}_2) = \{(x, [y^{(1)}, \ldots, y^{(N)}]) : x_n = 0 \text{ for some } n\}
\]

\[ \cup \{(x, [y^{(1)}, \ldots, y^{(N)}]) : 1^T x = P^x\} \]

\[ \cup \{(x, [y^{(1)}, \ldots, y^{(N)}]) : x_n = C_n^x \text{ for some } n\} \]

\[ \cup \{(x, [y^{(1)}, \ldots, y^{(N)}]) : y^{(n)}_k = 0 \text{ for some } n, k\} \]

\[ \cup \{(x, [y^{(1)}, \ldots, y^{(N)}]) : \sum_{n=1}^{N} y^{(n)}_k = P^y_k \text{ for some } k\} \]

\[ \cup \{(x, [y^{(1)}, \ldots, y^{(N)}]) : y^{(n)}_k = C_k^{y, (n)} \text{ for some } n, k\}. \] (2.71)

Assume that such a sequence \{(z)_n\} exists, and consider the embedded sequences \{(x)_n\} and \{([y^{(1)}], \ldots, [y^{(N)}])_n\} satisfying \((z)_n = ((x)_n, ([y^{(1)}], \ldots, [y^{(N)}])_n)\). Because \((z)_n\) is a sequence in \(\tilde{S}_1 \times \tilde{S}_2 \subset \mathcal{S}\) and \(\mathcal{S}\) is compact, \{\(z_n\)\} must have a convergent subsequence, call it \{\(z'_n\)\}, with embedded sequences \{(x')_n\} and \{([y']_n, \ldots, [y']_n)\} satisfying \((z'_n) = ((x')_n, [y']_n, \ldots, [y']_n)_n\). The sequence \{(z)_n\} must satisfy at least one of the six conditions (2.71). Each of these conditions are considered in order below.

1. Suppose \((x)_m \rightarrow 0\) for some \(m\) and consider the partial derivative \(\frac{\partial j}{\partial x_m}\) evaluated at the \(n\)th term of the subsequence \{((z')_n)\}, written \(\frac{\partial j}{\partial x_m} ((z')_n)\). The first term in (2.68) is bounded for all \(x \in \mathcal{S}_1\). The second term, \(\frac{1}{l((x')_m)_n} \rightarrow \infty\) because \((x)_m \rightarrow 0\) and every subsequence has the same limit as its parent sequence (when the limit exists). Similarly, the last term \(\frac{1}{l(C_m^x-1)((x')_n)} \rightarrow \frac{-1}{rC_m^x}\). Therefore \(\frac{\partial j}{\partial x_m} ((z')_n) \rightarrow \infty\) unless \(\frac{1}{l(P^x - 1)((x')_n)} \rightarrow \infty\).

But in that case, \(1^T (x')_m \rightarrow P^x > 0\) and therefore it cannot be that \((x')_m \rightarrow 0\) for all elements \(m'\). Considering one such element \(m'\), suppose \((x')_m \rightarrow c > 0\). Examining \(\frac{\partial j}{\partial x_m} ((z')_n)\), the first term in (2.68) again is bounded, and the second term \(\frac{-1}{l((x')_m)_n} \rightarrow \frac{-1}{tc} < 0\). The last term is nonpositive and the third term \(\frac{-1}{l(P^x - 1)((x')_n)} \rightarrow -\infty\). Therefore \(\left| \frac{\partial j}{\partial x_m} ((z')_n) \right| \rightarrow \infty\).
2. If $1^T(x)_n \longrightarrow P^x$ then $1^T(x')_n \longrightarrow P^x$ which implies that one cannot have $(x'_m)_n \longrightarrow 0$ for all $m$. Let $m$ be one such element satisfying $(x'_m)_n \longrightarrow c > 0$ where the limit exists because $\{(x')_n\}$ is convergent. Examining $\frac{\partial J}{\partial x_m}((z')_n)$, the first term in (2.68) again is bounded, the second term converges to $\frac{1}{tc} < \infty$, the last term is nonpositive, and the third term converges to $-\infty$. Therefore $\left| \frac{\partial J}{\partial x_m}((z')_n) \right| \longrightarrow \infty$.

3. If $(x_n)_n \longrightarrow C^*_{m'}$ for some $m$, then $(x'_m)_n \longrightarrow C^*_{m}$. Proceeding as before, the first term in (2.68) again is bounded, the second term converges to $\frac{1}{C^*_{m}}$, the third term is nonpositive, and the last term converges to $-\infty$ by assumption. Therefore $\left| \frac{\partial J}{\partial x_m}((z')_n) \right| \longrightarrow \infty$.

4. This Case proceeds analogously to Case 1. Suppose that $(y^{(m')}_k)_n \longrightarrow 0$ for some $k$ and $m$, in which case $(y^{(m')}_k)_n \longrightarrow 0$ for the same $k$ and $m$. Examining $\frac{\partial J}{\partial y^{(m)}_k}((z')_n)$, the first two terms in (2.69) are bounded for all $[y^{(1)}, \ldots, y^{(N)}] \in S^*_2$. The third term $\frac{1}{t(y^{(m')}_k)_n} \longrightarrow -\infty$, and the fourth term converges to $\frac{1}{tC^{(m')}_k}$. Therefore $\frac{\partial J}{\partial y^{(m)}_k}((z')_n) \longrightarrow -\infty$ unless $\frac{1}{tP^{(m)}_k - \sum_{m=1}^{N} y^{(m')}_k} \longrightarrow \infty$.

In this case, $\sum_{m=1}^{N} (y^{(m')}_k)_n \longrightarrow P^{(m)}_k > 0$. Therefore it cannot be that $(y^{(m')}_k)_n \longrightarrow 0$ for all $m'$. Considering one such $m'$, we have $(y^{(m')}_k)_n \longrightarrow c > 0$. Examining $\frac{\partial J}{\partial y^{(m')}_k}((z')_n)$, the first two terms of (2.69) again are bounded, and the second term converges to $\frac{1}{tc} > 0$. The last term is nonnegative and the fourth term $\frac{1}{tP^{(m')}_k - \sum_{m=1}^{N} y^{(m')}_k} \longrightarrow \infty$. Therefore $\frac{\partial J}{\partial y^{(m')}_k}((z')_n) \longrightarrow \infty$.

5. This Case proceeds analogously to Case 2. If $\sum_{m=1}^{N} (y^{(m)}_k)_n \longrightarrow P^{(m)}_k$ then so must the convergent subsequence $\sum_{m=1}^{N} (y^{(m)}_k)_n \longrightarrow P^{(m)}_k$ which implies that one cannot have $(y^{(m)}_k)_n \longrightarrow 0$ for all $m$. Let $m$ be one such element satisfying $(y^{(m)}_k)_n \longrightarrow c > 0$ where the limit exists because $\{(y^{(m)}_k)_n\}$ is convergent. Examining $\frac{\partial J}{\partial y^{(m)}_k}((z')_n)$, the two terms in (2.69) again are bounded, the third term converges to $\frac{1}{tc} > -\infty$, the last term is nonnegative, and the third term converges to $\infty$. Therefore $\frac{\partial J}{\partial y^{(m)}_k}((z')_n) \longrightarrow \infty$. 
6. This Case proceeds analogously to Case 3. If \((y^{(m)}_k)_n \rightarrow C_k^{y_r(m)}\) for some \(m, k\), then \((y^{(m)}_k)_n \rightarrow C_k^{y)(m)}\). The first two terms in (2.68) are bounded, the third term converges to \(\frac{1}{C_k^{y_r(m)}}\), the fourth term is nonnegative, and the last term converges to \(\infty\) by assumption. Therefore \(\left| \frac{\partial \tilde{I}}{\partial x_m} ((z')_n) \right| \rightarrow \infty\).

Therefore in all cases, \(\left| \frac{\partial \tilde{I}}{\partial x_m} ((z)_n) \right| \) or \(\left| \frac{\partial \tilde{I}}{\partial y_k^{(m)}}((z)_n) \right|\) has a subsequence that converges to \(\infty\) for some \(m\) (respectively some \(k\) and \(m\)). By definition, for every \(\mathbb{R} \ni M < \infty\) there exists an \(n_0 \in \mathbb{N}\) such that \(\left| \frac{\partial \tilde{I}}{\partial x_m}((z')_n) \right| \) or \(\left| \frac{\partial \tilde{I}}{\partial y_k^{(m)}}((z')_n) \right| \) for all \(n > n_0\). Because all norms on \(\mathbb{R}^p\) where \(p < \infty\) are equivalent, in particular the norms \(|| \cdot ||_{\infty}\) and \(|| \cdot ||_2\), for all \(M'\) there therefore exists some \(n_0\) such that \(\left| \nabla \tilde{J} ((z')_n) \right|_2 > M'\) for all \(n > n_0\). Because \(\{z'_n\}\) is a subsequence of \(\{z_n\}\), this implies that for all \(M' > 0\) there exists some \(n \in \mathbb{N}\) such that \(\left| \nabla \tilde{J} ((z)_n) \right|_2 > M'\). This contradicts the assumption that \(\{z_n\}\) is a sequence in \(S_{\alpha}\), thereby establishing the Lemma.

**Lemma 9** Consider two twice-differentiable functions \(f_1 : R \mapsto \mathbb{R}\) and \(f_2 : R \mapsto \mathbb{R}\) where \(R \subset \mathbb{R}^p\) and define the function \(f : R \mapsto \mathbb{R}\) as \(f(z) = f_1(z) + f_2(z)\) for all \(z \in R\).

If \(f_1\) and \(f_2\) have Lipschitz continuous Hessians on \(R\), then so does \(f\).

**Proof:** For all \(z, z' \in R\),

\[
\left| \nabla^2 f(z) - \nabla^2 f(z') \right|_2 = \left| \nabla^2 f_1(z) + \nabla^2 f_2(z) - \nabla^2 f_1(z') - \nabla^2 f_2(z') \right|_2 \\
\leq \left| \nabla^2 f_1(z) - \nabla^2 f_1(z') \right|_2 + \left| \nabla^2 f_2(z) - \nabla^2 f_2(z') \right|_2 \\
\leq (L_1 + L_2) \left| z - z' \right|_2,
\]

where the second line follows by the triangle inequality.

The preceding Lemma is particularly useful in conjunction with the following composition result.

**Lemma 10** Consider a general function \(f : R \mapsto \mathbb{R}\) where \(R \subset \mathbb{R}^p\) and a bounded linear map \(A : \mathbb{R}^q \mapsto \mathbb{R}^p\). Let \(R'\) be any subset of the preimage of \(R\) under \(A\), that is \(R' \subset \{y \in \mathbb{R}^q : Ay \in R\}\). Define \(g : R' \mapsto \mathbb{R}\) as \(g(y) = f(Ay)\) for all \(y \in R'\).
If $f$ has a Lipschitz continuous Hessian on $R$, then $g$ has a Lipschitz continuous Hessian on $R'$.

Proof: For any $y, y' \in R'$, it holds

$$||\nabla^2 g(y) - \nabla^2 g(y')||_2 = ||A\nabla^2 f(z)A^T - A\nabla^2 f(z')A^T||_2$$

$$\leq ||A||_2 ||\nabla^2 f(z) - \nabla^2 f(z')||_2 ||A^T||_2$$

$$= ||A||_2 ||\nabla^2 f(z) - \nabla^2 f(z')||_2$$

$$\leq ||A||_2 L||z - z'||_2,$$

(2.73)

where $z = Ay$ and $z' = Ay'$. Note that because $R'$ is a subset of the preimage of $R$, we have $z, z' \in R$. The second line in (2.73) holds due to submultiplicativity of the norm [57, §6.1], and the third because $||A||_2 = ||A^T||_2$ [57, §6.5]. By assumption, $||A||_2$ is finite, thus the result follows.

Lemma 11 Consider a function $f : R \mapsto \mathbb{R}$, where $R \subset \mathbb{R}^p$ and having Hessian $\nabla^2 f(r)$. If there exists some $L \in \mathbb{R}^{p \times p} < \infty$ such that for all $1 \leq j \leq p$, $1 \leq k \leq p$ and $z, z' \in R$ it holds

$$\left|\left(\nabla f^2(z)\right)_{j,k} - \left(\nabla f^2(z')\right)_{j,k}\right| \leq L_{j,k}||z - z'||_2$$

(2.74)

then $f$ has a Lipschitz continuous Hessian on $R$.

Proof: By defining the following norm\(^{14}\) on $A \in \mathbb{R}^{m \times n}$

$$||A||\Delta = \max_{i,j} |A_{i,j}|$$

(2.75)

\(^{14}\)This is a vector norm, not a matrix norm; thus submultiplicativity $||AB||\Delta \leq ||A||\Delta ||B||\Delta$ need not hold. This distinction turns out to have no bearing on the analysis herein.
Observe that the following inequality chain holds

\[ ||A||_2 = \sup_{||x||_2=1} ||Ax||_2 \]
\[ = \sqrt{\sup_{||x||_2=1} \sum_{i=1}^{m} (a_i^T x)^2} \]
\[ \leq \sqrt{\sum_{i=1}^{m} \sup_{||x||_2=1} (a_i^T x)^2} \]
\[ \leq \sum_{i=1}^{m} ||a_i||_2 \]
\[ \leq \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}^2} \]
\[ \leq \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} ||A||_\Delta^2} \]
\[ = \sqrt{nm}||A||_\Delta, \quad (2.76) \]

where \( a_i \) denotes the \( i \)th row of \( A \). The fourth line is due to Cauchy-Schwartz and the fifth is due to concavity of \((\cdot)^{1/2}\) on \( \mathbb{R}_+ \). The assumed condition \((2.74)\) implies

\[ ||\nabla^2 f(z) - \nabla^2 f(z')||_\Delta \leq L' ||z - z'||_2, \quad (2.77) \]

where \( L' = \max_{i,j} L_{i,j} < \infty \). In turn, \((2.76)\) and \((2.77)\) imply

\[ ||\nabla^2 f(z) - \nabla^2 f(z')||_2 \leq p ||\nabla^2 f(z) - \nabla^2 f(z')||_\Delta \]
\[ \leq pL' ||z - z'||_2. \quad (2.78) \]

In particular, the condition \((2.78)\) is precisely the definition of \( f \) having a Lipschitz continuous Hessian on \( R \) with coefficient \( L = p \max_{i,j} L_{i,j} \).

\[ \]

Lemma 12 The function \( \tilde{J} \) has a Lipschitz continuous Hessian on \( S_\alpha \) for every fixed \( \alpha \).
\( \alpha < \infty, t > 0. \)

**Proof:** The proceeds by showing that each term of \( \tilde{J} \) (2.38) has a Lipschitz-continuous Hessian on \( S_\alpha \). It is then possible to show that the result follows easily from this property.

- For each term \( t^{-1} \log(C^x_n - x_n) \), consider the bounded linear map \( A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R} \)

\[
A(x, [y^{(1)}, \ldots, y^{(N)}]) = x_n. \tag{2.79}
\]

And let \( R \) be the image of \( S_\alpha \) under \( A \)

\[
R = \{ p \in \mathbb{R} : \exists (x', [y'^{(1)}, \ldots, y'^{(N)}]) \in S_\alpha \text{ such that } A((x', [y'^{(1)}, \ldots, y'^{(N)}])) = p \}, \tag{2.80}
\]

which is compact because \( S_\alpha \) is closed (Lemma 8) and bounded (therefore compact) and \( A \) is bounded (therefore continuous).

Consider \( a = 0, b = C^x_n, a' = \min_{x \in R} x, \) and \( b' = \max_{x \in R} x \). Note that the definition of \( a' \) and \( b' \) is well-posed because \( R \) is compact and therefore the minimum and maximum exist. Also, \( a < a' < b' < b \) because \( S_\alpha \subset \text{int}(S) \) hence \( 0 \notin R, b \notin R \) (refer to (2.71)). Thus we have \( t^{-1} \log(b - p) = h(p) \) for all \( p \in R \), and therefore by Lemma 15, \( t^{-1} \log(b - p) \) has a Lipschitz continuous Hessian on \( R \). Observe that by construction, \( S_\alpha \) is a subset of the preimage of \( R \) under the bounded linear map \( A \). Therefore by Lemma 10, \( t^{-1} \log(C^x_n - x_n) \) has a Lipschitz continuous Hessian on \( S_\alpha \).

- For each term \( t^{-1} \log(C^y_i^{(n)} - y_i^{(n)}) \), consider the bounded linear map \( A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R} \)

\[
A(x, [y^{(1)}, \ldots, y^{(N)}]) = y_i^{(n)}. \tag{2.81}
\]
And let $R$ be the image of $S_\alpha$ under $A$

$$R = \{ p \in \mathbb{R} : \exists (x', [y^{(1)}, \ldots, y^{(N)}]) \in S_\alpha \text{ such that } A((x', [y^{(1)}, \ldots, y^{(N)}])) = p \}$$

(2.82)

which is compact because $S_\alpha$ is closed (Lemma 8) and bounded (therefore compact) and $A$ is bounded (therefore continuous).

Consider $a = 0$, $b = C_y^{(n)}$, $a' = \min_{x \in R} x$, and $b' = \max_{x \in R} x$. Note that the definition of $a'$ and $b'$ is well-posed because $R$ is compact and therefore the minimum and maximum exist. Also, $a < a' < b' < b$ because $S_\alpha \subset \text{int}(S)$ hence $0 \notin R, b \notin R$ (refer to (2.71)). Thus we have $t^{-1} \log(b - p) = h(p)$ for all $p \in R$, and therefore by Lemma 15, $t^{-1} \log(b - p)$ has a Lipschitz continuous Hessian on $R$. Observe that by construction, $S_\alpha$ is a subset of the preimage of $R$ under the bounded linear map $A$. Therefore by Lemma 10, $t^{-1} \log(C_y^{(n)} - y^{(n)}_l)$ has a Lipschitz continuous Hessian on $S_\alpha$.

- For the term $t^{-1} \log(P^x - \sum_{n=1}^{N} x_n)$, consider the bounded linear map $A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R}$

$$A(x, [y^{(1)}, \ldots, y^{(N)}]) = \sum_{n=1}^{N} x_n.$$  

(2.83)

And let $R$ be the image of $S_\alpha$ under $A$

$$R = \{ p \in \mathbb{R} : \exists (x', [y^{(1)}, \ldots, y^{(N)}]) \in S_\alpha \text{ such that } A((x', [y^{(1)}, \ldots, y^{(N)}])) = p \}$$

(2.84)

which is compact because $S_\alpha$ is closed (Lemma 8) and bounded (therefore compact) and $A$ is bounded (therefore continuous).

Consider $a = 0$, $b = P^x$, $a' = \min_{x \in R} x$, and $b' = \max_{x \in R} x$. Note that the definition of $a'$ and $b'$ is well-posed because $R$ is compact and therefore the minimum and maximum exist. Also, $a < a' < b' < b$ because $S_\alpha \subset \text{int}(S)$ hence $0 \notin R, b \notin R$ (refer to (2.71)). Thus $t^{-1} \log(b - p) = h(p)$ for all $p \in$
and therefore by Lemma 15 it holds that $t^{-1} \log(b - p)$ has a Lipschitz continuous Hessian on $R$. Observe that by construction, $S_\alpha$ is a subset of the preimage of $R$ under the bounded linear map $A$. Therefore by Lemma 10, $t^{-1} \log(Px - \sum_{n=1}^{N} x_n)$ has a Lipschitz continuous Hessian on $S_\alpha$.

- For the term $t^{-1} \log(Py_l - \sum_{n=1}^{N} y_l^{(n)})$, consider the bounded linear map $A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R}$

$$A(\mathbf{x}, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]) = \sum_{n=1}^{N} y_l^{(n)}.$$ (2.85)

And let $R$ be the image of $S_\alpha$ under $A$

$$R = \{ p \in \mathbb{R} : \exists (\mathbf{x}', [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}]) \in S_\alpha \text{ such that } A((\mathbf{x}', [\mathbf{y}'^{(1)}, \ldots, \mathbf{y}'^{(N)}])) = p \}$$ (2.86)

which is compact because $S_\alpha$ is closed (Lemma 8) and bounded (therefore compact) and $A$ is bounded (therefore continuous).

Consider $a = 0$, $b = P_y$, $a' = \min_{x \in R} x$, and $b' = \max_{x \in R} x$. Note that the definition of $a'$ and $b'$ is well-posed because $R$ is compact and therefore the minimum and maximum exist. Also, $a < a' < b' < b$ because $S_\alpha \subset \text{int}(S)$ hence $0 \notin R, b \notin R$ (refer to (2.71)). Thus we have $t^{-1} \log(b - p) = h(p)$ for all $p \in R$, and therefore by Lemma 15 it holds that $t^{-1} \log(b - p)$ has a Lipschitz continuous Hessian on $R$. Observe that by construction, $S_\alpha$ is a subset of the preimage of $R$ under the bounded linear map $A$. Therefore by Lemma 10, $t^{-1} \log(P_i - \sum_{n=1}^{N} y_l^{(n)})$ has a Lipschitz continuous Hessian on $S_\alpha$.

- For each term $t^{-1} \log(y_l^{(n)})$, consider the bounded linear map $A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R}$

$$A(\mathbf{x}, [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]) = y_l^{(n)}.$$ (2.87)
And let \( R \) be the image of \( S_\alpha \) under \( A \)

\[
R = \{ p \in \mathbb{R} : \exists (x', [y^{(1)}, \ldots, y^{(N)}]) \in S_\alpha \text{ such that } A((x', [y^{(1)}, \ldots, y^{(N)}])) = p \}
\]

(2.88)

which is compact because \( S_\alpha \) is closed (Lemma 8) and bounded (therefore compact) and \( A \) is bounded (therefore continuous). Also, for \( p > 0 \) for all \( p \in R \) because \( S_\alpha \subset \text{int}(S) \) (refer to (2.71)). Thus we have \( t^{-1} \log(p) = g(p) \) for all \( p \in R \), and therefore by Lemma 14 it holds that \( t^{-1} \log(p) \) has a Lipschitz continuous Hessian on \( R \). Observe that by construction, \( S_\alpha \) is a subset of the preimage of \( R \) under the bounded linear map \( A \). Therefore by Lemma 10, \( t^{-1} \log(y^{(n)}_i) \) has a Lipschitz continuous Hessian on \( S_\alpha \).

- For each term \( t^{-1} \log(x_n) \), consider the bounded linear map \( A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \rightarrow \mathbb{R} \)

\[
A(x, [y^{(1)}, \ldots, y^{(N)}]) = x_n.
\]

(2.89)

And let \( R \) be the image of \( S_\alpha \) under \( A \)

\[
R = \{ p \in \mathbb{R} : \exists (x', [y^{(1)}, \ldots, y^{(N)}]) \in S_\alpha \text{ such that } A((x', [y^{(1)}, \ldots, y^{(N)}])) = p \}
\]

(2.90)

which is compact because \( S_\alpha \) is closed (Lemma 8) and bounded (therefore compact) and \( A \) is bounded (therefore continuous). Also, for \( p > 0 \) for all \( p \in R \) because \( S_\alpha \subset \text{int}(S) \) (refer to (2.71)). Thus we have \( t^{-1} \log(p) = g(p) \) for all \( p \in R \), and therefore by Lemma 14 it holds that \( t^{-1} \log(p) \) has a Lipschitz continuous Hessian on \( R \). Observe that by construction, \( S_\alpha \) is a subset of the preimage of \( R \) under the bounded linear map \( A \). Therefore by Lemma 10, \( t^{-1} \log(x_n) \) has a Lipschitz continuous Hessian on \( S_\alpha \).

- Finally, for each term \( \log \left( 1 + \frac{x_n}{\beta_n x_n + \alpha_n h^{(n)}_i y^{(n)} + N_n} \right) \), consider the bounded linear
map $A : \mathbb{R}^N \times \mathbb{R}^{2L \times N} \mapsto \mathbb{R} \times \mathbb{R}$

$$A(x, [y^{(1)}, \ldots, y^{(N)}]) = (x_n, h^{(n)}y^{(n)}). \quad (2.91)$$

And let $R$ be the image of $S_\alpha$ under $A$

$$R = \{(x, \eta) \in \mathbb{R} \times \mathbb{R} : \exists (x', [y'^{(1)}, \ldots, y'^{(N)}]) \in S_\alpha \text{ such that } A((x', [y'^{(1)}, \ldots, y'^{(N)}])) = (x, \eta)\} \quad (2.92)$$

which is compact because $S_\alpha$ is closed (Lemma 8) and bounded (therefore compact) and $A$ is bounded (therefore continuous). Also, $x \geq 0$, $\eta \geq 0$ for all $(x, \eta) \in R$ because $h^{(n)} \geq 0$ and $S_\alpha \subset S$. Let $\gamma = N_n > 0$, $\alpha = \alpha_n \geq 0$, $\beta = \beta_n \geq 0$.

Thus we have $t^{-1} \log \left(1 + \frac{x}{\beta x + \alpha \eta + \gamma}\right) = f(x, \eta)$ for all $(x, \eta) \in R$, and therefore by Lemma 13 it holds that $\log \left(1 + \frac{x}{\beta x + \alpha \eta + \gamma}\right)$ has a Lipschitz continuous Hessian on $R$. Observe that by construction, $S_\alpha$ is a subset of the preimage of $R$ under the bounded linear map $A$. Therefore by Lemma 10, $t^{-1} \log \left(1 + \frac{x_n}{\beta x_n + \alpha_n h^{(n)}y^{(n)} + N_n}\right)$ has a Lipschitz continuous Hessian on $S_\alpha$.

Because each term of $\tilde{J}$ (2.38) has a Lipschitz-continuous Hessian on $S_\alpha$, by Lemma 9 their sum does.

\[\square\]

**Lemma 13** Consider any $\beta > 0$, $\gamma > 0$, $\alpha \geq 0$, and $R$ that is a compact subset of $\mathbb{R}_+ \times \mathbb{R}_+$. Define

$$\epsilon_0 = \begin{cases} 
\min \left(\frac{\gamma}{4\beta(\beta+1)}, \frac{\gamma}{2\alpha}\right) & \alpha > 0, \\
\min \left(\frac{\gamma}{4\beta(\beta+1)}, 1\right) & \text{else},
\end{cases} \quad (2.93)$$

$$R_0 = \{(x, \eta) \in \mathbb{R}^2 : x > -\epsilon_0, \ y > -\epsilon_0\}. \quad (2.94)$$
Further, define the function \( f : R_0 \rightarrow \mathbb{R}_+ \) as
\[
f(x, \eta) = \log \left( 1 + \frac{x}{\beta x + \alpha \eta + \gamma} \right).
\] (2.95)

The function \( f \) has a Lipschitz continuous Hessian on \( R \).

Proof: It is clear that \( R \subset R_0 \) and furthermore that \( R_0 \) is open and convex. To prove that \( f \) has a Lipschitz continuous Hessian on \( R \), it is sufficient (by Lemma 11) to show that each of the four terms of the Hessian of \( f \) satisfy (2.74). The latter condition holds by Lemma 16 if each term of the Hessian is a smooth function on \( R_0 \). Therefore to prove that \( f \) has a Lipschitz continuous Hessian on \( R \), it suffices to compute the respective partial derivatives and show that they are continuous functions on \( R_0 \).

First consider the Hessian term
\[
\frac{\partial^2 f}{\partial x^2}(x, \eta) = \frac{(\alpha \eta + \gamma)(2 \beta(\beta + 1)x + (2 \beta + 1)(\alpha \eta + \gamma))}{-(\alpha \eta + \beta x + \gamma)^2(\alpha \eta + (\beta + 1)x + \gamma)^2}.
\] (2.96)

and observe that the function \( \frac{\partial^2 f}{\partial x^2}(x, \eta) : R_0 \mapsto \mathbb{R} \) has continuous partial derivatives with respect to \( x \) and \( \eta \) on \( R_0 \), that is,
\[
\frac{\partial}{\partial x} \frac{\partial^2 f}{\partial x^2}(x, \eta) =
\frac{2(\alpha \eta + \gamma)(3 \beta^2(\beta + 1)^2x^2 + 3(\alpha \eta + \gamma)\beta(\beta + 1)(2 \beta + 1)x + (\alpha \eta + \gamma)^2(3 \beta^2 + 3 \beta + 1))}{(\beta x + \alpha \eta + \gamma)^3((\beta + 1)x + \alpha \eta + \gamma)^3}
\]
\[
\frac{\partial}{\partial \eta} \frac{\partial^2 f}{\partial x^2}(x, \eta) =
\frac{2(\eta^3 \alpha^3(2 \beta + 1) + 3 \eta^2(\beta(\beta + 1)x + (2 \beta + 1)\gamma)\alpha^2 + 3 \eta(2 \beta(\beta + 1)x + (2 \beta + 1)\gamma)\alpha \gamma)}{(\beta x + \alpha \eta + \gamma)^3((\beta + 1)x + \alpha \eta + \gamma)^3}
\]
\[
- \frac{2(\beta^2(\beta + 1)^2x^3 + 3 \beta(\beta + 1)\gamma^2x + (2 \beta + 1)\gamma^2)}{(\beta x + \alpha \eta + \gamma)^3((\beta + 1)x + \alpha \eta + \gamma)^3}
\] (2.97)

are continuous functions\(^{15}\) of \( x \) and \( \eta \) on \( R_0 \).

\(^{15}\)It is easy to see that all the numerators and denominators are continuous, and that the denominators are strictly positive on \( R_0 \).
The argument for the other three terms of the Hessian proceeds similarly. The function \( \frac{\partial^2 f}{\partial \eta^2}(x, \eta) : R_0 \mapsto \mathbb{R} \) where
\[
\frac{\partial^2 f}{\partial \eta^2}(x, \eta) = \frac{\alpha^2 (2\alpha\eta + (2\beta + 1)x + 2\gamma)x}{(2\alpha\eta + (2\beta + 1)x + 2\gamma)^2 (2\beta + 1)x + \alpha\eta + \gamma)^2}
\] (2.98)
has continuous partial derivatives on \( R_0 \), given by
\[
\frac{\partial}{\partial x} \frac{\partial^2 f}{\partial \eta^2}(x, \eta) = -\frac{2\alpha(2(\beta + 1)(2\beta + 1)x^3 + 3(\alpha\eta + \gamma)\beta(\beta + 1)x^2 - (\alpha\eta + \gamma)^3)}{(\beta x + \alpha\eta + \gamma)^3((\beta + 1)x + \alpha\eta + \gamma)^3}
\]
(2.99)
Respecting the remaining two terms, one has \( \frac{\partial}{\partial \eta} \frac{\partial f}{\partial x}(x, \eta), \frac{\partial}{\partial x} \frac{\partial f}{\partial \eta}(x, \eta) : R_0 \mapsto \mathbb{R} \) where
\[
\frac{\partial}{\partial \eta} \frac{\partial f}{\partial x}(x, \eta) = \frac{\partial}{\partial x} \frac{\partial f}{\partial \eta}(x, \eta) = \frac{\alpha(\beta(\beta + 1)x^2 - (\alpha\eta + \gamma)^2)}{(\beta x + \alpha\eta + \gamma)^3((\beta + 1)x + \alpha\eta + \gamma)^3}.
\] (2.100)
The partial derivatives of this term of the Hessian are given by
\[
\frac{\partial}{\partial x} \frac{\partial}{\partial x} \frac{\partial f}{\partial \eta}(x, \eta) = -\frac{2\alpha(2(\beta + 1)^2x^3 + 3(\alpha\eta + \gamma)^2\beta(\beta + 1)x - (\alpha\eta + \gamma)^3(2\beta + 1))}{(\beta x + \alpha\eta + \gamma)^3((\beta + 1)x + \alpha\eta + \gamma)^3}
\]
(2.101)
which are again continuous on \( R_0 \), and therefore the last two terms on the Hessian satisfy the Lipschitz property (2.74). Applying Lemma 11, we conclude that \( f \) has a
Lipschitz continuous Hessian on $R$. \hfill $\blacksquare$

**Lemma 14** Consider any compact set $R$, $R \subset \mathbb{R}^{++}$, and define

$$R_0 = \{ x \in \mathbb{R} : 2 \sup_{r \in R} r > x > \frac{1}{2} \inf_{r \in R} r \}. \tag{2.102}$$

Further, define the function $g : R_0 \mapsto \mathbb{R}$

$$g(x) = t^{-1} \log(x). \tag{2.103}$$

The function $g$ has a Lipschitz continuous Hessian on $R$.

**Proof:** Proceeding as in the proof of Lemma 13, observe that $R_0$ is convex, open, and $R \subset R_0$. Compute

$$\frac{\partial g}{\partial x} = \frac{1}{tx} \tag{2.104}$$

$$\frac{\partial^2 g}{\partial x^2} = -\frac{1}{tx^2} \tag{2.105}$$

$$\frac{\partial^3 g}{\partial x^3} = \frac{1}{tx^3}. \tag{2.106}$$

Evidently $\frac{\partial^3 g}{\partial x^3}(x)$ is continuous on $R_0$ whence by Lemma 16, $\frac{\partial^3 g}{\partial x^3}(x)$ is Lipschitz continuous on $R$. \hfill $\blacksquare$

**Lemma 15** Consider any open interval $R_0 = (a, b)$ satisfying $\infty > b > a > 0$ and any closed interval $R = [a', b']$ where $a < a' < b' < b$. Define $h : R_0 \mapsto \mathbb{R}$ such that

$$h(x) = t^{-1} \log(b - x). \tag{2.107}$$

The function $h$ has a Lipschitz continuous Hessian on $R$. 
Proof: Proceeding as in the proof of Lemma 13, compute:

\[
\frac{\partial h}{\partial x} = \frac{1}{t(b - x)}
\]  \hspace{1cm} (2.108)

\[
\frac{\partial^2 h}{\partial x^2} = \frac{-1}{t(b - x)^2}
\]  \hspace{1cm} (2.109)

\[
\frac{\partial^3 h}{\partial x^3} = \frac{1}{t(b - x)^3}.
\]  \hspace{1cm} (2.110)

It is immediate the \( R_0 \) is open and convex and that \( R \) is compact. Also, \( \frac{\partial^3 h}{\partial x^3} (x) \) is continuous on \( R_0 \) and therefore by Lemma 16, \( h \) has a Lipschitz continuous Hessian on \( R \).

\[\text{Lemma 16}\] Let \( R \) be a compact subset of \( \mathbb{R}^n \), \( n < \infty \) and let \( f \) be a smooth function on an open convex set \( R_0 \) such that \( R \subset R_0 \). Then \( f \) is Lipschitz continuous on \( R \).

Proof: Following [64, Thm. 1.1.7], write the first-order Taylor expansion of \( f \) at an arbitrary \( x_0 \in R \):

\[
f(x) = f(x_0) + g(x')^T(x - x_0), \quad x \in R
\]  \hspace{1cm} (2.111)

for some \( x' = \theta x + (1 - \theta)x_0 \) where \( 0 \leq \theta \leq 1 \). Because \( R \) is convex, \( x' \in R \). By the Cauchy-Schwartz inequality, it follows:

\[
|f(x) - f(x_0)| = |g(x')^T(x - x_0)| \leq ||g(x')||_2 ||x - x_0||_2.
\]  \hspace{1cm} (2.112)

Because \( R \) is compact and \( ||g(x)||_2 : R \mapsto \mathbb{R} \) is continuous on \( R \), it follows that \( \sup_{x \in R} ||g(x)|| = L < \infty \). Therefore for arbitrary \( x \in R, x_0 \in R \):

\[
|f(x) - f(x_0)| \leq L||x - x_0||_2,
\]  \hspace{1cm} (2.113)

which is precisely the definition of Lipschitz continuity.

\[\text{Lemma 17}\] For any fixed \( t > 0 \), the modified payoff function \( \tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) \) is strongly convex in \( x \) for fixed \( [y^{(1)}, \ldots, y^{(N)}] \) and strongly concave in \( [y^{(1)}, \ldots, y^{(N)}] \)
for fixed $x$. Formally, if one fixes $t > 0$ and defines $\tilde{J}_{[y^{(1)}, \ldots, y^{(N)}]} : \bar{S}_1 \mapsto \mathbb{R}$ and $J_x : \bar{S}_2 \mapsto \mathbb{R}$ such that:

$$\tilde{J}_{[y^{(1)}, \ldots, y^{(N)}]}(x) = J_x([y^{(1)}, \ldots, y^{(N)}]) = J(x, [y^{(1)}, \ldots, y^{(N)}]).$$  \hspace{1cm} (2.114)

Then there exists some $m > 0$ such that for all $(x, [y^{(1)}, \ldots, y^{(N)}]) \in \bar{S}_1 \times \bar{S}_2$:

$$\nabla^2 J_x([y^{(1)}, \ldots, y^{(N)}]) \succeq mI$$ \hspace{1cm} (2.115)

$$\nabla^2 \tilde{J}_{[y^{(1)}, \ldots, y^{(N)}]}(x) \preceq -mI.$$ \hspace{1cm} (2.116)

**Proof:** The function $\tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}])$ is written as a sum of terms (2.38), each of which is concave in $x$ for fixed $[y^{(1)}, \ldots, y^{(N)}]$ and convex in $[y^{(1)}, \ldots, y^{(N)}]$ for fixed $x$. Therefore in order to show the result, it is sufficient to show that any subset of these terms satisfy (2.116) and (2.115).

To establish (2.115), consider the terms corresponding to the positivity constraints on the interfering modems:

$$-t^{-1} \sum_{n=1}^{N} \sum_{l=1}^{2L} \log(y_i^{(n)}).$$ \hspace{1cm} (2.117)

The term corresponding to $y_i^{(n)}$ depends *only* on $y_i^{(n)}$ and moreover it was proven in Theorem 12, (2.105) that the respective term is strictly convex in $y_i^{(n)}$. Note that because for all $[y^{(1)}, \ldots, y^{(N)}] \in S_2$ it holds

$$y_i^{(n)} \leq P_i^y \leq \max_i P_i^y$$ \hspace{1cm} (2.118)
it is the case that (2.105) implies

\[
\frac{\partial}{\partial y^{(n)}_l} \frac{\partial (-t^{-1} \log(y^{(n)}_l))}{\partial y^{(n)}_l} (x, [y^{(1)}, \ldots, y^{(N)}]) \geq \frac{t^{-1}}{(\max_i P^y_i)^2}
\]

\[
\frac{\partial}{\partial y^{(n')}_{l'}} \frac{\partial (-t^{-1} \log(y^{(n)}_l))}{\partial y^{(n)}_l} (x, [y^{(1)}, \ldots, y^{(N)}]) = 0, \ (m, l) \neq (m', l'). \quad (2.119)
\]

Recall that the function \( \tilde{J}(x, [y^{(1)}, \ldots, y^{(N)}]) \) is written as a sum of terms convex in \([y^{(1)}, \ldots, y^{(N)}]\) for fixed \(x\), and by (2.119) there is at least one strictly convex term corresponding to each \(y^{(n)}_l\). Thus for any \(v \in \mathbb{R}^{2LN}\), \((x, [y^{(1)}, \ldots, y^{(N)}]) \in \tilde{S}_1 \times \tilde{S}_2:\)

\[
v^T \left( \nabla^2_{x} \tilde{J}_x ([y^{(1)}, \ldots, y^{(N)}]) \right) v \geq v^T v \frac{t^{-1}}{(\max_i P^y_i)^2}, \quad (2.120)
\]

which means precisely that (2.115) holds for \(m'' = \frac{t^{-1}}{(\max_i P^y_i)^2}\).

To establish (2.116), consider the terms corresponding to the rate of the victim user. Recall from Theorem 2 that in the sum

\[
\sum_{n=1}^{N} \log \left( 1 + \frac{x_n}{\alpha_n h^{(n)} y^{(n)} + \beta_n x_n + N_n} \right), \quad (2.121)
\]

the term corresponding to the \(n\)th tone is strictly concave in \(x_n\) for fixed \([y^{(1)}, \ldots, y^{(N)}]\). Therefore their sum is strictly concave in \(x\) for fixed \([y^{(1)}, \ldots, y^{(N)}]\), implying (2.116) for some \(m' > 0\). Finally, taking \(m = \min(m', m'') > 0\) establishes the result.
Chapter 3

DSL Applications of the WCI Analysis

This chapter considers an application of the Worst Case Interference analysis to spectrum management in DSL.

The scope of the WCI analysis extends generally to DMT-based DSL systems. This section examines two particular cases that are deployed prevalently: VDSL and ADSL. In VDSL, a prominent interference issue is the upstream near-far effect, which is caused by crosstalk from short ("near") lines FEXT-coupling into longer ("far") lines. In ADSL, the issue of RT FEXT injection into longer CO lines is similarly of concern. Numerical results for these sample deployments demonstrate the practicality of the WCI analysis and show surprising commonalities between the different scenarios. In all simulations, the interior-point technique is used with an error tolerance of $< 0.1\%$.

3.1 VDSL

The WCI rate bound is first applied to two different upstream VDSL scenarios exhibiting the near-far effect. The binder configuration is illustrated in Figure 3.1. For all simulations, $19 \times 300$ m lines, $10 \times 1200$ m lines, and one line of varying length occupy the binder of 24 AWG twisted-pairs. Thus, the binder contains a total of 30
twisted pairs. The FTTE Ex M2 (998 FDM) bandplan is employed with HAM bands notched, and the usual PSD constraints removed. Tones below 138 kHz are disabled for ADSL compatibility, and the normal PSD masks are not applied. The FDM condition is satisfied for this configuration, hence $\beta_n = 0$. For $10^{-7}$ BER, assume coding gain of 3 dB, with 6 dB margin, thus $\Gamma = 12.5$ dB. Each line is limited to 14.5 dBm power ($P^x = 14.5$ dBm, $P^y = 1 \cdot 14.5$ dBm).

**WCI rate as a function of line length**

First, consider the WCI rate bound when the variable-length line is the victim line (Player 1). Numerical results are shown in Figure 3.3, where a rate lower bound as well as the rate obtained when all lines execute full-power rate-adaptive (RA) IW are plotted as a function of victim line length. Note that full-power RA IW is quite different from fixed-margin (FM) IW, where power is minimized while achieving a fixed rate and margin [88]. To investigate practical bitloading constraints numerically, RA IW with discrete bit constraints [72] is executed on the victim modem assuming the WCI (2.14). Player 1’s achieved rate with discrete bit loading is plotted as $R^*_d$. Evidently, $R^*_d \leq R^*$, and therefore $R^*_d$ is also a lower bound to the achievable rate under the WCI.
CHAPTER 3. DSL APPLICATIONS OF THE WCI ANALYSIS

Figure 3.2: Binder configuration for downstream RT ADSL simulations (not to scale). A common RT is used for each line.

Observe that for most line lengths, the rate achieved by RA IW is fairly close to the WCI bound, particularly near 200 m and 900 m. For intermediate lengths (≈ 650 m), rate-adaptive IW can perform up to ≈ 75% better than the WCI bound, though the absolute difference is small. As a corollary, the interference generated by IW in this configuration is deleterious in the sense that it is close in rate to the WCI saddle-point. This finding is consistent with results [10] showing that other centralized DSM strategies can significantly outperform IW in such cases. Furthermore, fixed-margin (FM) IW can also be seen to perform significantly better than the WCI bound when rates are adjudicated reasonably [88].

WCI rate as a function of PBO

Motivated by the results of the previous section showing that the full-power WCI rate bound can decrease precipitously as loop length increases, the efficacy of upstream power backoff (UPBO) at mitigating this effect is considered. This section examines a simple power backoff strategy in the form of power-constrained RA IW for Level 0-1 DSM. Though the use of RA IW is retained, an effect similar to fixed-margin (rate-constrained) IW [88] is induced by imposing various tighter sum-power constraints. In particular, the variable-length line is set to length 300 m, and (sum) power back-off
Figure 3.3: Achievable rates in upstream VDSL as a function of victim line length (200-1000 m).
Figure 3.4: Achievable rates in upstream VDSL as a function of short line (300 m) power back-off.
is imposed on all (20) 300 m lines with full power retained on the (10) 1200 m lines. By taking the victim line to be one of the 300 m lines, the 300 m WCI curve in Figure 3.4 is generated, yielding a lower bound to the achievable data rate for all 300 m lines in the binder. The 1200 m WCI curve represents the case where the victim modem is instead taken to be one of the 1200 m lines. To compare standardized SSM techniques to DSM, the rates achieved using the SSM VDSL UBPO masking technique defined for the Noise A environment [2] are illustrated by dashed horizontal lines.

The results illustrate that a tradeoff exists between the rates of the short and long lines. Examining the 1200 m lines, the proposed technique improves both the RA IW-achieved and WCI bounds significantly up to approximately $-30 \text{ dBm}$, with diminishing returns for further PBO as the 300 m line FEXT no longer dominates the interference profile. However, further PBO further decreases the achievable rates of the 300 m lines, as expected. The WCI bound is again fairly tight. Thus by employing such a simple PBO scheme with Level 1 DSM, one can dynamically control the tradeoff between short and long lines to best match desired operating conditions, i.e. operating with guaranteed $\approx 4 \text{ MBPS}$ on the 1200 m lines and $\approx 7.75 \text{ MBPS}$ on the 300 m lines. In this example, the SSM technique achieves approximately the same performance as this simple DSM technique at one tradeoff point ($\approx -22 \text{ dB PBO}$).

### 3.2 ADSL

The WCI rate bound is also applicable to ADSL. This section considers an RT ADSL configuration as illustrated in Figure 3.2. For all simulations, 25 ADSL lines are located 2000 m from a fiber-fed RT 4000 m from the CO. Additionally, $5 \times 5000$ m lines are present in the binder. The FDM ADSL standard [33] parameters are assumed. As in the VDSL simulations, $\Gamma = 12.5 \text{ dB}$. Each line is limited to $20.4 \text{ dBm}$ downstream power ($P^x = 20.4 \text{ dBm}$, $P^y = 1 \cdot 20.4 \text{ dBm}$), and the standard PSD masks are neglected.

A common problem of such configurations is that the signal from the CO to the non-RT (7000 m) modems will be saturated by FEXT from the RT lines. As in
the VDSL example, the efficacy of (sum) power backoff for the RT lines as a means of improving the rate of the CO lines is studied. Figure 3.5 shows the dependence of rates on the level of power backoff (relative to 20.4 dBm) for the RT lines. The horizontal lines represent the performance obtained by SSM with the standardized PSD masks.

The WCI bound is reasonably close to actual power-controlled RA IW performance on both RT and CO lines. Figure 3.6 shows the spectrum adopted at the (approximate) Nash equilibrium, as well as the power allocation chosen by discrete IW against the noise induced by Player 2, yielding $R^*_d$ (in discrete IW, tones above 47 are not used because they correspond to fractional bit loadings). The simulation shows that Player 1’s interference is dominated by interference from the RT modems; these modems induce a “kindred-like” noise while the CO lines to concentrate their power at low frequencies. Also illustrated by example is that the Player 2’s optimal strategy may be highly frequency-selective, and therefore the existing interference analysis technique of setting tight PSD masks for each modem cannot capture the WCI unless the masks are set very high\textsuperscript{1}. As in VDSL, a wide range of useful operating points may be attained; for example, it is possible (through proper power control) to guarantee 3 MBPS service on all lines, whereas this rate point was far from feasible with SSM or with full-power rate-adaptive IW. However without any power backoff, the performance of RA IW and the WCI bound are near that of SSM, showing the key role of power control in obtaining DSM gains in this setting.

\textsuperscript{1}Doing so would consistently overestimate interference power, and underestimate achievable DSM performance.
Figure 3.5: Achievable rates in downstream ADSL as a function of RT line power backoff (relative to 20.4 dBm nominal TX power).
Figure 3.6: Spectral allocations \( (x, [y^{(1)}, \ldots, y^{(N)}]) \) of Players 1 and 2 for the rightmost lower (0 dB PBO) operating point in Figure 3.5, where Player 1 is a CO line. Note Player 2 has different Nash equilibrium spectra for its CO and RT loops, and that Player 2’s RT line spectrum overlaps Player 1’s Nash equilibrium strategy on most tones.
Chapter 4

The Vector Worst Case Interference

The use of vectored transmission (also termed Level 3 DSM) has been proposed [38] for next-generation DSL to achieve near Shannon multiuser capacity in twisted-pair channels. This chapter examines the interference properties of such vectored DSL systems. In particular, a framework is developed to calculate the interference effects of vectoring systems into colocated vectoring systems.

The interference analysis of vectored systems differs from that of non-vectored systems (Chapter 2) in two critical aspects:

1. Coordination enables new transmission and reception schemes that offer fundamentally superior performance than is achievable under independent (per-line) encoding and decoding schemes.

2. Coordination across multiple copper twisted pairs admits the possibility of spatially-correlated transmission. The interference characteristics of spatially-correlated signals differ fundamentally; this is because, generally speaking, signal powers do not add linearly (as they do for uncorrelated transmission).
CHAPTER 4. THE VECTOR WORST CASE INTERFERENCE

4.1 Background

4.1.1 Gaussian Broadcast and Multiple Access Channels

A number of relevant surveys [23] [78] [36] and texts [22] [35] [24] have been written on topics in multiuser information theory, and in particular multiple-access and broadcast channels. The broadcast channel was initiated by Cover [21] and has been developed extensively in the period since. We shall have particular interest in the non-scalar Gaussian broadcast channel; the sum capacity\(^1\) of this channel is known [90] [81] [82], and recent results [85] assert that the entire capacity region corresponds to a “dirty-paper” region. Also of interest is the capacity of multiple-access channels, which admit a concise single-letter characterization in the finite-alphabet setting [4] [56], and also in the scalar Gaussian case, where the capacity region is given by the Cover-Wyner “pentagon” region [22]. It was shown in [76] that this region is in fact polymatroidal. The scalar Gaussian capacity generalizes in a straightforward fashion to the multiple-input multiple-output (MIMO) Gaussian multiple access channel (MAC) (e.g. [39]).

The following standard definitions [22] are stated for completeness. The 2-user Gaussian broadcast channel (BC) is illustrated in Figure 4.1. A sender wishes to send individual messages to remote users 1 and 2 subject to a maximum power constraint of \(P^x\). The rate pair \((R_1, R_2)\) is said to be achievable for the BC in Figure 4.1 if there exists a sequence of \((2^{nR_1}, 2^{nR_2}, n)\) codes defining an encoding function

\[
X : \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} \mapsto \prod_n \mathbb{R}^{N_T}
\]

\(^1\)This text exclusively considers capacity with asymptotically vanishing average probability of error [77].
and decoding functions\(^2\)

\[
\hat{W}_1 : \prod_n \mathbb{R}^{N_1} \mapsto \{1, \ldots, 2^{nR_1}\}, \tag{4.2}
\]

\[
\hat{W}_2 : \prod_n \mathbb{R}^{N_2} \mapsto \{1, \ldots, 2^{nR_2}\}, \tag{4.3}
\]

such that for a uniform distribution of \(W_1\) and \(W_2\) on \(\{1, \ldots, 2^{nR_1}\}\) and \(\{1, \ldots, 2^{nR_2}\}\), respectively, it holds

\[
P_e^n = P(\hat{W}_1(y_1, \ldots, y_n) \neq W_1 \text{ or } \hat{W}_2(y_1, \ldots, y_2) \neq W_2) \longrightarrow 0 \tag{4.4}
\]
as \(n \longrightarrow \infty\) and

\[
\frac{1}{n} \sum_{i=1}^{N_1} \sum_{j=1}^{n} |[X^n(w)]_{ij}|^2 \leq P_x, \tag{4.5}
\]

for all \(w \in \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\}\).

The rate pair \((R_1, R_2)\) is said to be achievable for the MAC in Figure 4.2 if there exists a sequence of \((2^{nR_1}, 2^{nR_2}, n)\) codes defining encoding functions

\[
X_1 : \{1, \ldots, 2^{nR_1}\} \mapsto \prod_n \mathbb{R}^{N_1}, \tag{4.6}
\]

\[
X_2 : \{1, \ldots, 2^{nR_2}\} \mapsto \prod_n \mathbb{R}^{N_2}, \tag{4.7}
\]

as well as a decoding function

\[
\hat{X} : \prod_n \mathbb{R}^{N_R} \mapsto \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} \tag{4.8}
\]

having probability of error satisfying (for uniform distribution on \(W_1, W_2\))

\[
P_e^{(n)} = P(\hat{X}(y^1, \ldots, y^n) \neq (W_1, W_2)) \longrightarrow 0 \tag{4.9}
\]

\(^2\)Note that the decoding functions (as well as the encoding function) vary with \(n\); this dependence is suppressed in the notation \(e.g. \hat{W}_1\).
as \( n \to \infty \) and

\[
\frac{1}{n} \sum_{i=1}^{N_1} \sum_{j=1}^{n} |[X_1(w_1)]_{i,j}|^2 \leq P_1, \quad (4.10)
\]

\[
\frac{1}{n} \sum_{i=1}^{N_2} \sum_{j=1}^{n} |[X_2(w_2)]_{i,j}|^2 \leq P_2, \quad (4.11)
\]

for all \( w_1 \in \{1, \ldots, 2^{nR_1}\} \), \( w_2 \in \{1, \ldots, 2^{nR_2}\} \), and \( n \in \mathbb{Z}_{++} \).
4.1.2 Practical Schemes for BC and MAC

Implementable schemes for transmission over Gaussian BCs and MACs have been considered in the literature. For the BC, much work has been undertaken from the viewpoint of coding theory, e.g. [26] [27] [43] study classes of codes for the degraded BC in terms of distance and error probability. Li studied the capacities of BCs for time-division multiple access (TDMA) and superposition schemes under fading [55].

Numerous investigations have examined trellis coding techniques for unequal error protection (UEP) of multiple data streams [12] [49]. Wei [84] comparatively studied a particular trellis-coded scheme with a TDMA scheme. For the ad hoc techniques considered therein, it was observed that TDMA outperformed superposition-coded modulation (SCM) when there was relatively more high-priority information. In other cases, SCM offered higher coding gain and better impulsive noise protection. Gadkari and Rose [34] further investigated the conditions under which TDMA outperforms SCM, and found that for the studied suboptimal schemes, the relative merit was related to the ratio of high-priority to low-priority information. Wang and Orchard [83], recently developed a low-complexity superposition coding scheme based on trellis-coded modulation, showing both UEP and significant coding gain. However, a material gap between the performance and the information-theoretic bound remained.

The use of iterative ‘turbo’ decoding has been studied in connection with UEP. Parallel concatenated codes having nonuniform puncturing and interleaving were studied [18] and shown to achieve UEP. However, information-theoretic aspects were not directly explored. Recently, Sun and Wesel [74] implemented a turbo-coding structure [32] on top a superposition trellis-coded modulation scheme (Turbo Trellis-Coded Modulation (TCM)). The results show performance near the Shannon bound for many different configurations [75]. Another high-performance scheme based on trellis and convolutional precoding has recently been introduced [92].

Even in near-Shannon-limit schemes such as turbo superposition coded modulation (SCM) [74] and trellis and convolutional precoding [92], there is a “gap” between
rates achievable in Shannon theory\(^3\) and those rates achievable with acceptable BER, delay, and implementation properties.

The Generalized Decision-Feedback Equalization (GDFE) architecture has been proposed for multiuser decoding in the Gaussian MAC. It has been shown to achieve the sum capacity [79], and can also be used to implement transmission systems that achieve any extreme point in the capacity region [89]. Other architectures have been proposed for scalar Gaussian MACs [65] and code-division multiple access (CDMA) systems (e.g. [80]).

### 4.1.3 Jamming Interference Analysis

In many communication applications, e.g. military communications, the transmission channel may be corrupted by a hostile adversary who seeks to impair transmission. Such hostile transmission is referred to as jamming, which has been studied in a great variety of applications in the literature [30] [41] [54].

A class of jamming analyses where the performance criterion is channel mutual information is of particular interest. In this context, the game-theoretic analysis of a single user channel with a single jammer has been performed under various conditions [22] [60] [68] [6] [25]. The terminology of “correlated jamming” has been used to denote correlation between a legitimate user’s signal and the jammer’s signal. Recent results extend this scope to multiuser channels; scalar multiple-access channels have been studied under various conditions [69] [68] [51].

### 4.1.4 Level 3 DSM

Level 3 Dynamic Spectrum Management implies that signal-level coordination is possible on the central office (CO) side of the DSL binder and hence “vectored” (MIMO) transmission is possible. The terminology of “vectoring” is selected to contrast with the information-theoretic [22] terminology of scalar MAC and BC channels. Unlike the classical formulation of the Gaussian vector broadcast channel and vector multiple

\(^3\)In general, achieving points in the BC capacity region at a given BER requires arbitrarily large delay, memory, and complexity; practical schemes with constrained parameters suffer a performance loss with respect to the asymptotic Shannon limit.
access channel (e.g. [22]), a DSL system may experience interference that varies as a function of time and the multiuser interference from external interference sources. These effects are not explicitly considered in the sequel.

Neglecting alien interference effects, vectored upstream transmission in DSL may be modeled as a frequency-selective, non-scalar Gaussian multiple access channel (MAC) and downstream transmission may be modeled as a frequency-selective, non-scalar, non-degraded\(^4\) Gaussian broadcast channel.

### 4.2 System Model

This section describes the system and channel models used for vectored upstream and downstream transmission in DSL. A discrete multitone (DMT) modulation scheme is employed, so that transmission over the frequency-selective MAC and BC channels is decoupled into \(N\) independent subcarriers or tones\(^5\). It is assumed that upstream and downstream transmission occur on distinct frequencies, so that NEXT is eliminated. Frequency bleeding that may occur between adjacent bands due to finite windowing is not explicitly considered; the “zipper” digital duplexing scheme [61] may be used in practice to eliminate this effect.

If bonding\(^6\) is implemented, then more than one twisted-pair is allocated for a given user. For clarity of exposition, it is assumed that each user is equipped with \(B\) twisted pairs; the common configuration of \(B = 1\) then falls out as a special case. For example, in upstream transmission with \(B\) twisted pairs deployed to each of \(K\) households, the vector Gaussian MAC channel would consist of \(B\) inputs per user and a total of \(BK\) outputs at the central office if differential excitation is employed\(^7\). The

\(^4\)In general, the user channels are not statistically degraded; definitions and a discussion of degradedness as it relates to the BC may be found in [22].

\(^5\)The error induced by this approximation for a finite number of tones or carriers is not explicitly considered.

\(^6\)Bonding refers to combined transmission on more than one twisted pair to each DSL modem (consumer premises equipment (CPE)); it is common practice in many localities to by default deploy more than one twisted pair per household in case additional services are desired in the future, rendering bonded DSL service a possibility.

\(^7\)There exist techniques such as non-differential “phantom” mode excitation [52] that can improve performance relative to a differential excitation implementation. Although beyond the present scope,
generalization to channels where there a different number of twisted-pairs deployed to each user is straightforward.

\section*{4.2.1 Upstream Channel}

A total of $K$ independent users implement an $N$-tone digital multitone (DMT) transmission scheme to combat channel frequency-selectivity in the binder channel. Each user is equipped with $B$ twisted pairs and transmits to a common central office with $BK$ physical channel outputs. A malicious jammer operates a total of $NJ$ twisted pairs in the same binder. The linear discrete-time channel\footnote{The independent subcarriers of the DMT transmission are incorporated into this model by taking $H_k$ to be a block-diagonal matrix where each block diagonal term corresponds to a particular independent subcarrier.} is governed by

\begin{equation}
\mathbf{y} = \sum_{k=1}^{K} H_k \mathbf{x}_k + Gz + n, \tag{4.12}
\end{equation}

where $H_k \in \mathbb{R}^{NBK \times NB}$ is the gain between the $k$th user and the receiver, $G \in \mathbb{R}^{NBK \times NN_J}$ is the gain between the jammer and the receiver, $\mathbf{y} \in \mathbb{R}^{NBK}$ is the channel output, and $\mathbf{x}_k \in \mathbb{R}^{NB}$ is $k$th user’s signal\footnote{Note that this notation differs from that of Chapter 2 where $\mathbf{x}, \mathbf{y}$ denoted user powers.}. Because transmission on each tone is independent, $H_k$ is block-diagonal such that

\begin{equation}
H_k = \begin{bmatrix}
H_k^{(1)} \\
\vdots \\
H_k^{(N)}
\end{bmatrix}, \tag{4.13}
\end{equation}

where $H_k^{(n)} \in \mathbb{R}^{BK \times B}$ is the gain between the $k$th user and the receiver on tone $n$. An identical decomposition applies to $G$.

A Gaussian jammer emits signal $z$, where $z \sim \mathcal{N}(0, Z)$. $\mathbf{n} \in \mathbb{R}^{NBK}$ denotes AWGN independent of the users and jammer, distributed as $\mathbf{n} \sim \mathcal{N}(0, \Lambda)$ where $\Lambda \succ 0$. Each legitimate transmitter has a power constraint: $\mathbf{E}[\mathbf{x}_k^T \mathbf{x}_k] \leq P_k \gamma^2$, for all $k = 1, \ldots, K$, as does the jammer: $\mathbf{E}[z^T z] \leq P_j$. The following analysis may be generalized to consider non-differential excitation as well.
All the transmitters and the jammer are assumed to have perfect channel state information (CSI). Although the jammer may correlate its interference in space ($Z$), eavesdropping by the jammer is not allowed, i.e. the jammer has no knowledge of the users’ signals ($\{x_k\}$) and thus may not perform “correlated jamming”.

### 4.2.2 Downstream Channel

A single transmitter, equipped with $BK$ physical channel inputs, transmits to $K$ independent users each with $B$ physical channel outputs. A jammer operates a total of $N_J$ twisted pairs. An $N$-tone DMT technique is again employed. The discrete-time channel is governed by the following linear model:

$$y_k = H_k x + G_k z + n_k,$$

where $H_k \in \mathbb{R}^{NB \times NBK}$ is the gain between the transmitter and the $k$th user, $G_k \in \mathbb{R}^{NB \times NN_J}$ is the gain between the jammer and user $k$, $y_k \in \mathbb{R}^{NB}$ is the channel output observed by user $k$, $x \in \mathbb{R}^{NBK}$ is the transmitter’s signal, and $z_k \in \mathbb{R}^{NB}$ is AWGN observed by user $k$. The distribution of the AWGN, which is independent of the jammer, is $n_k \sim \mathcal{N}(0, \Lambda_k)$ where $\Lambda_k \succ 0$ for all $k = 1, \ldots, K$. Because transmission on each tone is independent, $H_k$ is block-diagonal such that

$$H_k = \begin{bmatrix} H_k^{(1)} & & \\ & \ddots & \\ & & H_k^{(N)} \end{bmatrix},$$

where $H_k^{(n)} \in \mathbb{R}^{B \times BK}$ is the gain between the transmitter and the $k$th receiver on tone $n$. An identical decomposition applies to $G_k$.

The transmitter and jammer’s power are upper-bounded: $E[x^T x] \leq P_x$, $E[z^T z] \leq P_J$. Both the users and the jammer are assumed to have perfect knowledge of all parameters (CSI), but the jammer again has no knowledge of $x$. 
CHAPTER 4. THE VECTOR WORST CASE INTERFERENCE

4.3 Upstream Interference Analysis

This section analyzes the interference characteristics of a hostile Gaussian jammer to transmission in MIMO multiple access channels. Basic properties of multiple-access channel capacity are first reviewed, which subsequently allow one to consider the jammer’s behavior in a game-theoretic framework.

4.3.1 Gaussian MAC Capacity

Begin by considering the case when the jammer is removed \( N_J = 0 \). In this case, the channel (4.12) corresponds to a Gaussian Multiple Access Channel. The (Shannon) capacity of this channel is well-known [39], and is given by the following expression

\[
C_{\text{MAC}}(\Lambda) = \bigcup_{\text{Tr}(X_k) \leq P_k, \ X_k \succeq 0, \ k = 1, \ldots, K} \left\{ R : 2 \sum_{s \in S} R_s \leq \log \left| \frac{\sum_{m \in S} H_m X_m H_m^T + \Lambda}{|\Lambda|} \right| \forall S \subset E \right\},
\]

(4.16)

where \( E = \{1, \ldots, K\} \) is an index set, \( R \in \mathbb{R}_+^K \) is a vector of user rates wherein \( R_k \) is the rate of user \( k \), and \( X_k \in S_+^{N_T} \) denotes the transmit covariance of user \( k \). The notation \( C_{\text{MAC}}(\Lambda) \) is employed to denote explicitly the dependence of the capacity region on \( \Lambda \).

Turning to the more general case with the jammer present \( (N_J > 0) \), denote the net (Gaussian) interference covariance at the receiver by \( \Psi \)

\[
\text{Cov}[n + Gz] = \Lambda + GZG^T = \Psi(Z).
\]

(4.17) \hspace{1cm} (4.18)

We consider the problem of computing

\[
\max \mu^T R \hspace{1cm} \text{subject to} \ R \in C_{\text{MAC}}(\Psi(Z)),
\]

(4.19)
where $\mu$ is a fixed non-negative weighting $\mu \in \mathbb{R}_+^K$ and $Z \succ 0$ is fixed. Note that the maximum is attained because $C_{\text{MAC}}(\Psi(Z))$ is closed and bounded. This optimization is illustrated in Figure 4.3 as the intersection of a tangent plane having normal vector $\mu$ with $C_{\text{MAC}}(\Lambda)$. It can be shown that (4.19) is a convex optimization problem in the sense of [8]; without loss of generality\(^{10}\), by assuming $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_K$ one has the equivalent problem [91]

$$
\begin{align*}
\max & \quad \frac{1}{2} \sum_{k=1}^{K} (\mu_k - \mu_{k-1}) \log \left| \frac{\Psi(Z) + \sum_{m=k}^{K} H_m X_m H_m^T}{|\Psi(Z)|} \right| \\
\text{subject to} & \quad \text{Tr}(X_k) \leq P_k, \\
& \quad X_k \succeq 0,
\end{align*}
$$

where in a slight abuse of notation, we define $\mu_0 = 0$. Note that the formulation (4.20) has only $K$ total log $|\cdot|$ expressions, which is a substantial simplification of the naive formulation of (4.19) based on the $2^K - 1$ convex inequalities defining $C_{\text{MAC}}(\Lambda)$ (4.16).

### 4.3.2 Game-Theoretic Formulation

Suppose now that the jammer wishes to choose a fixed covariance $Z$ so as to minimize the weighted rate achievable by the transmitters. The jammer’s optimization is

$$
\begin{align*}
\inf_Z \max_{\{X_k\}} & \quad \frac{1}{2} \sum_{k=1}^{K} (\mu_k - \mu_{k-1}) \log \left| \frac{\Psi(Z) + \sum_{m=k}^{K} H_m X_m H_m^T}{|\Psi(Z)|} \right| \\
\text{subject to} & \quad (X_1, X_2, \ldots, X_K) \in S_1, \\
& \quad Z \in S_2,
\end{align*}
$$

where we define $S_1 = \{(X_1, \ldots, X_K) : X_k \succeq 0, \text{Tr}(X_k) \leq P_k^x, k = 1, \ldots, K\}$ and $S_2 = \{Z : Z \succeq 0, \text{Tr}(Z) \leq P_j\}$.

It is possible to obtain additional insight to the problem (4.21) by appealing to game-theoretic results. Consider the following two-player strictly-competitive game,

\(^{10}\)Formally, consider a permutation $\pi$ on the user identifications such that $\mu_{\pi(1)} \leq \ldots \leq \mu_{\pi(K)}$ and define $\mu_{\pi_0} = 0$.\]
as summarized in Table 4.1. The objective function of the game is

\[ J(X_1, \ldots, X_K, Z) = \frac{1}{2} \sum_{k=1}^{K} (\mu_k - \mu_{k-1}) \log \left( \frac{|\Psi(Z) + \sum_{m=k}^{K} H_m X_m H_m^T|}{|\Psi(Z)|} \right), \tag{4.22} \]

where

\[ \Lambda + GZG^T \triangleq \Psi(Z). \tag{4.23} \]

The non-negative weighting \( \mu \in \mathbb{R}_+^K \) is fixed. Player 1 (corresponding to the legitimate users) chooses transmit covariances \( (X_1, \ldots, X_K) \) from the set \( S_1 \) to maximize the objective function (corresponding to their weighted rate). Player 2 (corresponding to the jammer) chooses a covariance \( (Z) \) to minimize the objective function. The game \( \mathcal{M} = (J, S_1, S_2) \) is defined as the Multiple Access Channel Worst-Case Weighted Rate game.

This game-theoretic formulation of (4.21) admits the application of Nash equilibrium results from game theory. A Nash equilibrium (in pure strategies) of the game \( \mathcal{M} \), denoted \( (X_1^*, \ldots, X_K^*, Z^*) \), has the physical interpretation of a worst interference covariance \( (Z^*) \) by the jammer, and the optimal transmit covariance for the users \( (X_1^*, \ldots, X_K^*) \) under such interference.

The following two lemmata are useful in the subsequent analysis:

**Lemma 18 ([25])** The function \( f : \mathbb{S}_+^n \to \mathbb{R} \) defined as

\[ f(K_x) = \log (|K_x + K_z|/|K_z|), \tag{4.24} \]

is convex in \( K_z \), where \( 0 \preceq K_x \in \mathbb{S}_+^n \). Furthermore, the convexity is strict if \( K_x > 0 \).
CHAPTER 4. THE VECTOR WORST CASE INTERFERENCE

Lemma 19 ([25] [20]) The function \( g : S_+^n \rightarrow \mathbb{R} \) defined as

\[
g(K_x) = \log \left( \frac{|K_x + K_z|}{|K_z|} \right),
\]

(4.25)
is strictly concave in \( K_x \), where \( 0 < K_z \in S^n \).

Theorem 14 The game \( \mathcal{M} \) always has at least one pure strategy Nash equilibrium, denoted \((X_1^*, \ldots, X_K^*, Z^*)\). Also, the game has a value.

Proof: The sets \( S_1 \) and \( S_2 \) are both compact and convex. The function \( J \) is continuous, and each \( \log |\cdot| \) term in the summation (4.22) is convex in \( Z \in S_2 \) for any fixed \((X_1, \ldots, X_K) \in S_1 \) due to Lemma 18 and the affine composition property. The differences \( \mu_k - \mu_{k-1} \) are all nonnegative, and therefore \( J \) is the nonnegative weighted sum of terms convex in \( Z \); hence \( J \) is convex in \( Z \) (for any fixed \((X_1, \ldots, X_K) \in S_1 \)). Also, for any fixed \( Z \in S_2 \), each \( \log |\cdot| \) term in the sum (4.22) is a concave function of \((X_1, \ldots, X_K) \) on \( S_1 \) due to Lemma 19 and the affine composition property. Therefore \( J \) is the nonnegative weighted sum of terms concave in \((X_1, \ldots, X_K) \) (for any fixed \( Z \in S_2 \)). The necessary conditions of [5, Thm. 4.4] are thereby satisfied and the
result follows.

As a corollary, the infimum in (4.21) is achieved.

It can be shown that, in general, the Nash equilibrium of the game $\mathcal{M}$ is not unique. However, the following result gives conditions under which a certain “partial” uniqueness holds.

**Theorem 15** Let $(X^*_1, \ldots, X^*_K, Z^*)$ be any Nash equilibrium of $\mathcal{M}$. If the matrices $H_1, \ldots, H_K$ are each full column rank and $0 < \mu_1 < \ldots < \mu_K$, then for all $(\hat{X}_1, \ldots, \hat{X}_K, \hat{Z})$ that are Nash equilibria of $\mathcal{M}$ it holds $\hat{X}_k = X^*_k$ for every $k = 1, \ldots, K$.

**Proof:** By the interchangeability property, $(\hat{X}_1, \ldots, \hat{X}_K, Z^*)$ is also a Nash equilibrium. Define $(\tilde{X}_1, \ldots, \tilde{X}_K) = \frac{1}{2}(\hat{X}_1, \ldots, \hat{X}_K) + \frac{1}{2}(X^*_1, \ldots, X^*_K)$. Because $J$ is concave in $(X_1, \ldots, X_K)$ with $Z^*$ fixed and the game has a value, it holds

$$J(\tilde{X}_1, \ldots, \tilde{X}_K, Z^*) = \frac{1}{2}J(X^*_1, \ldots, X^*_K, Z^*) + \frac{1}{2}J(\hat{X}_1, \ldots, \hat{X}_K, Z^*).$$

(4.26)

And because each term in the summation (4.22) similarly is concave in $(X_1, \ldots, X_K)$, it holds

$$(\mu_k - \mu_{k-1}) \log \frac{|\Psi(Z^*) + \sum_{m=k}^{K} H_m \tilde{X}_m H_m^T|}{|\Psi(Z^*)|}$$

$$\geq \frac{1}{2}(\mu_k - \mu_{k-1}) \log \frac{|\Psi(Z^*) + \sum_{m=k}^{K} H_m X^*_m H_m^T|}{|\Psi(Z^*)|}$$

$$+ \frac{1}{2}(\mu_k - \mu_{k-1}) \log \frac{|\Psi(Z^*) + \sum_{m=k}^{K} H_m \hat{X}_m H_m^T|}{|\Psi(Z^*)|},$$

(4.27)

for each $k = 1, \ldots, K$. Together (4.26) and (4.27) imply that the inequality in (4.27) holds with equality. Because $\Psi(Z^*) \succ 0$ and the differences $\mu_k - \mu_{k-1}$ are positive, Lemma 19 implies that

$$\sum_{m=k}^{K} H_m \tilde{X}_m H_m^T = \sum_{m=k}^{K} H_m X^*_m H_m^T,$$

(4.28)
for each $k = 1, \ldots, K$. Consider now the case of $k = K$. Because $H_K$ has full column rank, it has a left inverse $H_K^\dagger$ such that $H_K^\dagger H_K = I$. Then by (4.28), $H_K^\dagger H_K \tilde{X}_K H_K^T (H_K^T) = H_K^\dagger H_K X_K H_K^T (H_K^T)$ and hence $\tilde{X}_K = X_K$.

Now consider $k = K - 1$. By (4.28) it holds

$$
\sum_{m=K-1}^{K} H_m \tilde{X}_m H_m^T = \sum_{m=K-1}^{K} H_m X_m^* H_m^T.
$$

Because $\tilde{X}_K = X_K^*$ this implies that $H_{K-1} \tilde{X}_{K-1} H_{K-1}^T = H_{K-1} X_{K-1}^* H_{K-1}^T$. $H_{K-1}$ also has full column rank and a left inverse, whence $\tilde{X}_{K-1} = X_{K-1}^*$. By an identical induction argument, $\tilde{X}_k = X_k^*$ for each $k = 1, \ldots, K$, implying the result.

In practical DSL systems, the full column rank condition can be shown to hold in deployed loop channels [38] [9]. Therefore, in such DSL systems there exists a unique “robust” transmit covariance for each of the $K$ modems that maximizes the worst-case weighted rate\textsuperscript{11}.

### 4.4 Downstream Interference Analysis

This section considers a “dual” configuration whereby a single transmitter (i.e. central office) wishes to communicate with several independent receivers in the presence of hostile Gaussian jamming.

#### 4.4.1 Sum Capacity

It has been shown [90] (see also [81] [82]) that without the jammer ($N_J = 0$) in the channel (4.14), the Gaussian broadcast channel sum capacity $C_{\text{sum}}$ is given by

\textsuperscript{11}Under the mild non-degeneracy condition $0 < \mu_1 < \ldots \mu_K$. 
\[
\begin{align*}
\max_X \min_\Upsilon & \quad \frac{1}{2} \log |H X H^T + \Upsilon| - \frac{1}{2} \log |\Upsilon| \\
\text{subject to} & \quad \Upsilon^{[k]} = \Lambda_k \quad k = 1, \ldots, K, \\
& \quad \text{Tr}(X) \leq P_x, \\
& \quad \Upsilon \succeq 0, \\
& \quad X \succeq 0.
\end{align*}
\] (4.30)

where the notation
\[
H \triangleq \begin{bmatrix} H_1 \\ \vdots \\ H_K \end{bmatrix},
\] (4.31)
is employed and \(\Upsilon^{[k]}\) denotes the \(k\)th \(NB \times NB\) block-diagonal of the matrix \(\Upsilon \in \mathbb{R}^{N_B K \times N_B K}\). An interpretation [90, Thm. 3] of the formulation (4.30) is a strictly-competitive two-player game between the transmitter \((X)\) choosing an optimal transmit covariance and a “malicious nature” \((\Upsilon)\) choosing a worst joint distribution for the broadcast channel; furthermore, solutions of (4.30) correspond to Nash equilibria of the game. The sum capacity may be achieved by \textit{e.g.} dirty-paper coding [39], or trellis and convolutional precoding [92].

Considering now the interference experienced from the Gaussian jammer, the noise covariance seen by user \(k\) is given by
\[
\text{Cov}[n_k + G_k z] = \Lambda_k + G_k Z G_k^T.
\] (4.32)

For notational convenience, define
\[
\Psi(Z) \triangleq \begin{bmatrix} \Lambda_1 + G_1 Z G_1^T \\ \vdots \\ \Lambda_K + G_K Z G_K^T \end{bmatrix},
\] (4.33)

where all the off block-diagonal terms are 0. For any \textit{fixed} \(Z \succeq 0\), the sum capacity
CHAPTER 4. THE VECTOR WORST CASE INTERFERENCE

89

Player Controls Objective Strategy Set
1 Opr. A max \( J \) \( \{X : X \succeq 0, \ \text{Tr}(X) \leq P^x\} \)
2 Opr. B, “Nature” min \( J \) \( \{S, Z : Z \succeq 0, \ \text{Tr}(Z) \leq P^j, \ \Psi(Z) + S \succeq 0, S^{[k]} = 0\} \)

\[ \begin{array}{|c|c|c|}
\hline
\text{Player} & \text{Controls} & \text{Objective} & \text{Strategy Set} \\
\hline
1 & \text{Opr. A} & \max J & \{X : X \succeq 0, \ \text{Tr}(X) \leq P^x\} \\
2 & \text{Opr. B, “Nature”} & \min J & \{S, Z : Z \succeq 0, \ \text{Tr}(Z) \leq P^j, \ \Psi(Z) + S \succeq 0, S^{[k]} = 0\} \\
\hline
\end{array} \]

Table 4.2: Summary of Broadcast Channel game \( \mathcal{B} \).

The channel sum capacity \( C_{\text{sum}}(Z) \) may be written

\[
\min_S \max_X \frac{1}{2} \log |H X H^T + S + \Psi(Z)| - \frac{1}{2} \log |S + \Psi(Z)| \\
\text{subject to} \quad S^{[k]} = 0 \quad k = 1, \ldots, K, \\
\text{Tr}(X) \leq P^x, \\
\text{Tr}(Z) \leq P^j, \\
S + \Psi(Z) \succeq 0, \\
X \succeq 0, \\
Z \succeq 0.
\]

(4.34)

4.4.2 Game-Theoretic Formulation

Suppose that the jammer wishes to choose its covariance \( Z \) so as to minimize the channel sum capacity \( C_{\text{sum}}(Z) \). The jammer’s optimization is

\[
\inf_Z \min_S \max_X \frac{1}{2} \log |H X H^T + S + \Psi(Z)| - \frac{1}{2} \log |S + \Psi(Z)| \\
\text{subject to} \quad X \in \mathcal{S}_1, \\
(S, Z) \in \mathcal{S}_2,
\]

(4.35)

where \( \mathcal{S}_1 = \{X : X \succeq 0, \text{Tr}(X) \leq P^x\} \) and \( \mathcal{S}_2 = \{(S, Z) : S^{[k]} = 0, k = 1, \ldots, K, Z \succeq 0, \text{Tr}(Z) \leq P^j, \Psi(Z) \succeq -S\} \).

The formulation (4.35) may be interpreted as a strictly-competitive game, and is summarized in Table 4.2. In this two-player game, the objective function \( K : \mathcal{S}_1 \times \mathcal{S}_2 \mapsto \mathbb{R} \) is given by

\[
K(X, (S, Z)) = \frac{1}{2} \log |H X H^T + S + \Psi(Z)| - \frac{1}{2} \log |S + \Psi(Z)|.
\]

(4.36)
Player 1 chooses $X$ from the set $S_1$ to maximize the objective, while Player 2 chooses $(S, Z)$ from the set $S_2$ to minimize the objective. The game $\mathcal{B} = (J, S_1, S_2)$ is defined as the Broadcast Worst Throughput game. Observe that by setting $P^j = 0$, the Broadcast Worst Throughput game reduces to the game (4.30) defined for broadcast channel sum capacity; thus the former game generalizes the latter.

A pure-strategy Nash equilibrium represents a “worst” choice transmit covariance $(Z)$ by the jammer and channel joint distribution $(S)$ by “nature” [90], as well as the optimal response $(X)$ to this interference by the transmitter such that neither player has a unilateral incentive to deviate its strategy. It turns out that such a Nash equilibrium always exists in $\mathcal{B}$.

**Theorem 16** The game $\mathcal{B}$ has a Nash equilibrium in pure strategies and a value.

**Proof:** The crux of the proof is a theorem due to Diggavi and Cover [25] that has been utilized in a similar fashion to prove the Gaussian broadcast channel sum capacity [90]. Define $H = [H_1^T \ldots H_K^T]^T$ and $G = [G_1^T \ldots G_K^T]^T$, and let

$$F = \begin{bmatrix} I \\ G \end{bmatrix}.$$  \hfill (4.37)

Define $\mathcal{K}_j$ as the image of $S_2$ under the map $F(\cdot)F^T$

$$\mathcal{K}_j = \left\{ R : F \begin{bmatrix} S & 0 \\ 0 & Z \end{bmatrix} F^T = R \text{ for some } (S, Z) \in S_2 \right\}.$$  \hfill (4.38)

The set $\mathcal{K}_j$ is convex and compact because it is the image of a convex, compact set under a bounded linear (hence continuous) map. Define $\mathcal{K}_v$ similarly

$$\mathcal{K}_v = \{ R : HXH^T = R \text{ for some } X \in S_1 \}.$$  \hfill (4.39)

Define the sets of probability measures

$$\mathcal{V} = \text{cl}\{p_v : E[v] = 0, \ K_v \in \mathcal{K}_v \},$$  \hfill (4.40)

$$\mathcal{J} = \text{cl}\{p_j : E[j] = 0, \ K_j \in \mathcal{K}_j \}.$$  \hfill (4.41)
where \( K_v, K_j \) denote the covariance matrices of random variables distributed according to the measures \( p_v, p_j \) respectively.

By Lemma 20, a Gaussian saddle point exists in the strictly competitive game \((\mathcal{V}, \mathcal{J}, I(v; v+j))\). Let \((\tilde{V}, \tilde{J})\) denote the covariance matrices at such a saddle point. Thus \( \tilde{V} \in \mathcal{K}_v \) and \( \tilde{J} \in \mathcal{K}_j \). Let \( X^* \in \mathcal{S}_1 \) be such that \( H X^* H^T = \tilde{V} \); such an \( X^* \) exists by the definition (4.39). Similarly, let \((S^*, Z^*) \in \mathcal{S}_2 \) be such that \( F \begin{bmatrix} S^* & 0 \\ 0 & Z^* \end{bmatrix} F^T = \tilde{Z} \).

It may be readily verified that \( K(X^*, (S^*, Z^*)) = I(\tilde{v}; \tilde{v}+\tilde{j}) \) where \( \tilde{v}, \tilde{j} \) are zero-mean Gaussian with covariances \( \tilde{V}, \tilde{J} \) respectively.

It is claimed that \((X^*, (S^*, Z^*))\) is a Nash equilibrium of \( B \). This can be seen from the definition: suppose that there exists some \( X' \in \mathcal{S}_1 \) such that \( K(X', (S^*, Z^*)) > K(X^*, (S^*, Z^*)) \). Then one would have \( H X'H \in \mathcal{K}_v \) and for a Gaussian \( p_v' \) having this covariance it would hold \( I(v'; v' + j) > I(v; v + j) \), which contradicts the fact that \((\tilde{V}, \tilde{J})\) is a saddle point. A similar contradiction ensues if there exists some \((S', Z') \in \mathcal{S}_2 \) such that \( K(X^*, (S', Z')) < K(X^*, (S^*, Z^*)) \).

**Lemma 20** Consider the channel \( y = x + j \) where \( x \in \mathbb{R}^n, j \in \mathbb{R}^n \), and impose the constraints \( p_x \in \mathcal{X}, p_j \in \mathcal{J} \) where

\[
\mathcal{X} = \text{cl}\{p_x : E[x] = 0, \ K_x \in \mathcal{K}_x\}, \quad (4.42)
\]
\[
\mathcal{J} = \text{cl}\{p_j : E[j] = 0, \ K_j \in \mathcal{K}_j\}, \quad (4.43)
\]

\( p_x, p_j \) are probability measures defined on the Borel \( \sigma \)-algebra of \( \mathbb{R}^n \), the closure is in terms of the weak topology of probability measures [28, §2.2], \( K_x, K_j \) are the covariances of \( x, j \) and \( \mathcal{K}_x, \mathcal{K}_j \) are closed, bounded, and convex sets. Then there exists a pair \((p_{xG}^*, p_{jG}^*)\) that is a saddle point of \( I(x; x+j) \). Moreover, the pair \((p_{xG}^*, p_{jG}^*)\) where \( p_{xG}^*, p_{jG}^* \) are Gaussian distributions with the same covariance as \( p_{x}^*, p_{j}^* \), is also a saddle point.

**Proof:** The proof can be found in [25]. Note that the formulation of [25] contains a slightly less general formulation whereby

\[
\mathcal{X} = \text{cl}\{p_x : E[x] = 0, \ \text{Tr}(K_x) \leq P\}, \quad (4.44)
\]
however, as noted in [25], an identical argument may be applied to establish the result for the condition given in this Lemma.

In general, there need not exist a unique Nash equilibrium of $B$. However, sufficient conditions are given in the following Theorem for the uniqueness of $X^*$.

**Theorem 17** **In the game** $B$, **if** $H$ **has full column rank and if there exists a Nash equilibrium** $(X^*, S^*, Z^*)$ **such that** $\Psi(Z^*) \succ -S^*$, **then for every Nash equilibrium** $(\tilde{X}, \tilde{S}, \tilde{Z})$, **it holds** $\tilde{X} = X^*$.

**Proof:** The convex optimization problem

$$
\begin{align*}
\max & \quad \frac{1}{2} \log |Q + S + \Psi(Z)| - \frac{1}{2} \log |S + \Psi(Z)| \\
\text{subject to} & \quad Q \in S_1
\end{align*}
$$

(4.45)

has a unique solution [8], denoted $Q^*$ because by Lemma 19, the objective is strictly convex in $Q$. Therefore $HX^*H^T = Q^*$. By the exchangeability property of Nash equilibria, $(\tilde{X}, S^*, Z^*)$ is also a Nash equilibrium. Therefore by the identical argument above, $H\tilde{X}H^T = Q^*$ and hence $H\tilde{X}H^T = HX^*H^T$. Because $H$ is full column rank, it has a left inverse $H^\dagger$. Thus $H^\dagger H\tilde{X}H^T(H^\dagger)^T = H^\dagger HX^*H^T(H^\dagger)^T$, which simplifies to $\tilde{X} = X^*$.

Note that in contrast to Theorem 17, no conditions on the Nash equilibria are required by Theorem 15.

### 4.5 Application to Vectored DSL

The analysis of Sections 4.3 and 4.4 may be readily applied to the interference analysis of next-generation vectored DSL systems. Practical DSL loops are quite unlikely to experience interference from a *bona fide* jammer, but due to loop unbundling regulations that apply in certain legal jurisdictions, competing operators may need to share the same physical binder channel. As competitors, the different operators may have an incentive to act greedily or even to impair transmission of its competitors to the extent possible. Therefore, the interference properties of vectored DSL transmission...
are of central concern. The canonical binder configuration considered in this Section is illustrated in Figure 4.4.

The interference experienced by a “victim” local exchange carrier (LEC) A from a competing LEC B may be considered through the jamming framework. The “victim” LEC suffers upstream and downstream interference from (in general) multiple lines operated by competing LEC B (shown at bottom), who acts as a jammer. Crosstalk from LEC A into LEC B is not illustrated in the figure. For any fixed $\mu \geq 0$, a guarantee on the weighted rates of LEC A may be computed based on the foregoing Nash equilibrium analysis. For example in the upstream direction, by taking $\mu = [1 \ 0 \ 0 \ldots 0]^T$, the rate of User 1 may be lower-bounded, and by taking $\mu = 1$, the sum rate is lower bounded. These lower bounds constitute guarantees on the performance of a vectored DSL system.

By relabeling the parties such that LEC A is the jammer and LEC B is the “victim”, an interference analysis may be performed from the perspective of LEC B. It is straightforward to generalize this configuration to settings where multiple (> 2)
LEC’s share a binder by selecting a given LEC as the victim and considering all the others as jammers.

4.6 Numerical Examples

This section presents simple numerical examples that illustrate the numerical solution of the games $\mathcal{M}$ and $\mathcal{B}$. A DSL deployment has sufficient numerical parameters\textsuperscript{12} that printing them for a typical deployment would dwarf the present volume; therefore, simplified examples are presented to clarify the exposition.

4.6.1 Upstream Transmission

This section considers two numerical examples in upstream transmission.

Example 3 Let $N = 1$, $K = 2$, $B = 2$, $P_1^x = P_2^x = 1$, $P^j = 0$, $\mu = 1$ and

$$H_1 = \begin{bmatrix} 0.2 & 0.1 \\ -0.2 & 1.0 \end{bmatrix},$$

(4.46)

$$H_2 = \begin{bmatrix} -0.3 & 0.0 \\ 0.5 & 0.6 \end{bmatrix},$$

(4.47)

$$G = \begin{bmatrix} 0.2 & 0.1 \\ -0.3 & 0.15 \end{bmatrix},$$

(4.48)

$$\Lambda = I.$$  

(4.49)

Note that because the jammer is restricted to use zero total power and $\mu = 1$, this problem reduces to the computation of the sum capacity of the Gaussian vector multiple access channel. It may be verified (e.g. using the modified waterfilling algorithm of [91]) that for these parameters, $\mathcal{M}$ has the following Nash equilibrium.

\textsuperscript{12}Although channel modelling is well-understood in systems where interference is treated as noise, there is no standardized model for vectored DSL transmission. Therefore, any numerical results in DSL (in analogy with Chapter 3) would depend on the channel and noise models adopted.
Example 4 This example considers the same parameters as that of Example 3, except that $P_j = 1$. The following is a Nash equilibrium of $\mathcal{M}$ with these parameters:

$$X_1^* = \begin{bmatrix} 0.020 & -0.141 \\ -0.141 & 0.980 \end{bmatrix},$$

(4.50)

$$X_2^* = \begin{bmatrix} 0.686 & 0.464 \\ 0.464 & 0.314 \end{bmatrix},$$

(4.51)

$$Z^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

(4.52)

$$J(X_1^*, X_2^*, Z^*) = 0.508.$$

(4.53)

As expected, throughput is reduced by the jammer. Because the jammer’s interference covariance

$$GZ^*G^T = \begin{bmatrix} 0.019 & -0.047 \\ -0.047 & 0.112 \end{bmatrix}$$

(4.58)

is not large compared to the AWGN ($\Lambda = I$), the throughput loss is not substantial. Note further that in this example, the jammer uses full power ($\text{Tr}(Z^*) = P_j = 1$) as does Player 1 ($\text{Tr}(X_1^*) = P_1^x$, $\text{Tr}(X_2^*) = P_2^x$).
4.6.2 Downstream Transmission

Example 5 Let $N = 1$, $B = 1$, $K = 3$, $N_J = 2$, $P_x = 5$, $P_j = 0$, and

$$H = \begin{bmatrix} 1.0 & -0.3 & 0.2 \\ -0.4 & 2.0 & 0.5 \\ -0.1 & 0.2 & 3.0 \end{bmatrix}$$ (4.59)

$$G = \begin{bmatrix} 0.1 & -0.2 \\ 2.0 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$$ (4.60)

$$\Lambda_1 = \Lambda_2 = \Lambda_3 = 1.$$ (4.61)

This example corresponds to a multiple-input single-output\(^{13}\) (MISO) broadcast channel. The game $\mathcal{B}$ with these parameters has the following Nash equilibrium

$$X^\star = \begin{bmatrix} 1.076 & -0.233 & -0.0074 \\ -0.233 & 1.864 & 0.039 \\ -0.0074 & 0.039 & 2.060 \end{bmatrix},$$ (4.62)

$$Z^\star = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$ (4.63)

$$S^\star = \begin{bmatrix} 0 & -0.129 & 0.049 \\ -0.129 & 0 & 0.031 \\ 0.049 & 0.031 & 0 \end{bmatrix},$$ (4.64)

$$K(X^\star, (S^\star, Z^\star)) = 2.895.$$ (4.65)

Example 5 verifies the claim that the game $\mathcal{B}$ with $P_j = 0$ reduces to the vector Gaussian broadcast channel sum computation: the above results match exactly the numerical results of the sum capacity of this channel found by Yu [90].

Example 6 Now consider the same parameters as Example 5, except that $P_j = 5$.\(^{13}\)The channel has a single output for each decoder because $B = 1$.\(^{13}\)
This game has the following Nash equilibrium

\[
X^* = \begin{bmatrix}
1.734 & -0.480 & 0.032 \\
-0.480 & 0.671 & 0.126 \\
0.032 & 0.126 & 2.596
\end{bmatrix}, \quad (4.66)
\]

\[
Z^* = \begin{bmatrix}
0.873 & 1.898 \\
1.898 & 4.127
\end{bmatrix}, \quad (4.67)
\]

\[
S^* = \begin{bmatrix}
0 & 0.034 & 0.469 \\
0.034 & 0 & -2.979 \\
0.469 & -2.979 & 0
\end{bmatrix}, \quad (4.68)
\]

\[
K(X^*, (S^*, Z^*)) = 1.765. \quad (4.69)
\]

As expected, the value of \( B \) is reduced in Example 6 as \( P^j \) is increased from 0. Example 6 also shows that even when \( \Psi(Z^*) \gg -S^* \), the jammer’s Nash equilibrium strategy may be rank-deficient \( (Z^* \not\succ 0) \) as

\[
Z^* = \begin{bmatrix}
0.873 & 1.898 \\
1.898 & 4.127
\end{bmatrix} = 5 \begin{bmatrix}
0.418 \\
0.906
\end{bmatrix} \begin{bmatrix}
0.418 & 0.906
\end{bmatrix}. \quad (4.70)
\]
Chapter 5

Conclusions

5.1 Summary and Inferences

In the broadest terms, this thesis has studied the interference properties of a number of emerging DSL technologies. Motivated by the inability of existing techniques to characterize the range of performance of dynamic spectrum management DSL, a worst-case framework was developed and applied to different types of dynamic spectrum management (DSM) problems.

For Level 0-2 DSM where interference is treated as noise, a game-theoretic analysis has shown that a certain interference game has a pure-strategy Nash equilibrium that can be computed using known optimization techniques. The Nash equilibrium, which in practice is unique, provides a useful lower bound to the achievable rate for a DSL modem employing DSM under any power-constrained interference profile. One is therefore well-founded to speak of “the worst-case interference.”

The bound also establishes certain properties of the solutions of the noncooperative iterative waterfilling game. Furthermore, the structure of the Nash equilibrium reveals that for FDM systems, the well-known iterative waterfilling algorithm is optimal.

When applied to upstream near-far VDSL scenario, the bound was found to be numerically close to IW performance. The utility of a simple DSM UPBO strategy employing RA IW was compared to SSM UPBO, were it was found that control of
rate trade-offs is possible with DSM, which may allow significantly preferable operating rates. A similar tradeoff was observed in RT ADSL systems, where CO line performance benefits significantly from proper power control. These results suggest that the parameter of transmit power is important to DSM performance, in the sense that proper power control can beget large performance gains in this setting.

For Level 3 or vectored DSL, the properties of Gaussian jamming in MIMO multiple access and broadcast channels have been examined from the standpoint of game theory. Worst-case interferences and optimal responses thereto were obtained from Nash equilibria of suitably-defined strictly competitive games: in the multiple access channel, arbitrary weightings of users’ rates were adopted as the performance criterion. In the broadcast channel, throughput is considered, in a generalization of a game previously formulated for Gaussian sum capacity.

Partial uniqueness properties of these Nash equilibria have been developed. Under appropriate conditions, the Nash equilibrium strategies of the legitimate users were shown to be unique; such strategies therefore may be interpreted as “robust” transmit covariances that afford protection against a hostile Gaussian jammer.

In each game constructed in this thesis, a worst-case interference approach yielded tractable, strictly competitive games. In each case, the Nash equilibria of these games could be interpreted as representing a worst interference and an optimal response thereto. In each case, benign conditions - often satisfied in DSL practice - were shown to imply that the worst interference and/or the optimal response thereto were unique.

5.2 Future Research

The nascent field of spectrum management for DSM systems offers many further opportunities for study. While this thesis has adopted a worst-case framework for interference analysis, other comparatively benign interference models may be useful in some settings. This is particularly true if DSM algorithms for DSL modems become standardized. In such a case, the algorithm may prohibit many choices of power allocation, which may improve interference characteristics.
Even within the worst-case framework, many additional applications can be envisioned. An application of the multiuser worst-case interference analysis may be possible by optimizing the bound as a performance criterion. For example, the proposed band preference spectrum management scheme [17] simplifies spectrum management in Level 2 DSM by manipulating spectrum masks over large bands, or subdivisions of spectrum. In such a scheme, one seeks algorithms to judiciously select the spectrum mask parameters. One possible criterion for such algorithms is a maximization of the multiuser worst-case interference bound for some given $\mu$. By maximizing this bound, one thereby maximizes a (weighted) rate guarantee. By adjusting $\mu$, the priority of the users may be varied.

In its present form, the formulation of the vectored game for upstream (multiple access) is more general than that for downstream (broadcast). It may be possible to extend the downstream analysis to consider arbitrary weighting of users’ rates by assuming an achievable scheme (such as “dirty-paper coding”) and making assumptions on the users’ transmit covariances, encoding ordering, etc.

Further empirical studies may be conducted to estimate the difference between the achieved performance of “vectored” DSL systems and theoretical (Shannon) bounds. One candidate for such study is that of a receiver\footnote{The generalized decision feedback equalizer (GDFE) [89] employs such a structure.} employing successive interference cancellation (SIC). In practice, SIC receivers may suffer from channel estimation error, error propagation effects, and restrictions on their complexity. Different interferences may affect each of these constraints in a varied fashion. Pragmatic details such as discrete bit loading constraints, maximum bit loading constraints, and non-Gaussian (e.g. impulsive) noise may also affect numerical results. Such studies may thus lead to interference analyses - other than the worst case - that are relevant to vectored DSL.
Appendix A

Dual Decomposition Subproblem

This section considers the following convex optimization problem

\[
\min_y \log \left(1 + \frac{x}{\hat{c} + \hat{a}^T y}\right) + b^T y \quad (A.1)
\]

subject to \(0 \preceq y \preceq C(y)\),

where \(y \in \mathbb{R}^N\) are the decision variables, and \(x \in \mathbb{R}, a \in \mathbb{R}^N, b \in \mathbb{R}^N,\) and \(C(y) \in \mathbb{R}^N\) are constants satisfying \(x \geq 0, a \succeq 0, b \succeq 0, c > 0,\) and \(C(y) \succ 0\). Subsequently, the feasible set of \(y\) in the minimization shall be denoted \(\mathcal{Y} = \{y : 0 \preceq y \preceq C(y)\}\). Observe that \(\mathcal{Y}\) is closed and bounded, therefore compact. Because the objective is a continuous function on \(\mathcal{Y}\), the minimum over \(\mathcal{Y}\) is achieved.

In particular, if \(x = 0\), any feasible \(y\) achieves the minimum, and the optimization problem has a value of 0. The remainder of this Appendix considers the nontrivial case where \(x > 0\). By dividing through by \(x\) in the denominator, one obtains the following equivalent optimization

\[
\min_y \log \left(1 + \frac{1}{c + a^T y}\right) + b^T y \quad (A.2)
\]

subject to \(0 \preceq y \preceq C(y)\).

where \(c = \hat{c}/x > 0\) and \(a = \hat{a}/x \succeq 0\).
Observe that if \( a_n = 0 \) for some \( n \), then the objective is increasing in \( y_n \), and hence \( y_n = 0 \) is optimal and the variable may be eliminated. By eliminating all such variables, the following considers \( a \succ 0 \). Further, it is assumed without loss of generality that \( b_1/a_1 \leq b_2/a_2 \leq \ldots \leq b_N/a_N \) by reordering variables as necessary.\(^1\)

It is the case that this optimization problem (A.2) has an analytical solution, which is developed in the remainder of this Appendix. While closed-form, the computation of the exact solution requires up to \( N \) iterations of a simple algorithm. In this narrow computational sense, it is similar to the “water-filling” solution to parallel independent Gaussian channels. This result is stated formally and proved in Section A.1, and subsequently a heuristic explanation is developed in Section A.2.

### A.1 Analytical Solution

This section utilizes the suffix notation \( v[n] \) to denote the value of a set or variable \( v \) on the \( n \)th step of algorithm iteration. Recall that the notation \( e_n \) denotes a vector whose \( k \)th element is equal to 1 and all other elements are equal to 0.

**Theorem 18** The optimization problem (A.2) is solved by the following algorithm in at most \( N \) iterations

1. Initialize \( k = 1 \), \( y[k] = 0 \), \( \nu[k] = 0 \).

2. Set \( S[k] = \{k, \ldots, N\} \), and \( T[k] = \{1, \ldots, k\} \).

3. If \( b_n/a_n \geq (a^T y[k] + c)^{-1}(1 + a^T y[k] + c)^{-1} \) then set \( y[k+1] = y[k] \), \( \nu[k+1] = \nu[k] \) and stop. Otherwise, let

\[
\alpha[k] = \frac{\sqrt{4a_n + b_n} - (1 + 2c)\sqrt{b_n}}{2a_n\sqrt{b_n}}, \quad (A.3)
\]

and let \( \beta[k] = \min(\alpha[k], C_k^\eta) \).

\(^1\)Formally, consider a permutation \( \pi \) on the vectors \( a \) and \( b \) such that \( b_{\pi(1)}/a_{\pi(1)} \leq b_{\pi(2)}/a_{\pi(2)} \leq \ldots \leq b_{\pi(N)}/a_{\pi(N)} \).
4. Set $y[k+1] = y[k] + \beta[k]e_k$. Set

$$
\nu[k+1] = 0 - \sum_{m=1}^{k} e_m \left( b_m - \frac{a_m}{(a^T y[k+1] + c)(1 + a^T y[k+1] + c)} \right). \quad (A.4)
$$

If $k = N$, then stop. Otherwise, and increment $k$ and return to Step 2.

Proof: The KKT conditions associated with (A.2) are given by

$$
b_n + \frac{-a_n}{(a^T y + c)(1 + a^T y + c)} \geq 0, \quad \nu_n = 0 \text{ if } y_n = 0, \quad (A.5)
$$

$$
b_n + \frac{-a_n}{(a^T y + c)(1 + a^T y + c)} = 0, \quad \nu_n = 0 \text{ if } 0 < y_n < C^y_n \quad (A.6)
$$

$$
b_n + \frac{-a_n}{(a^T y + c)(1 + a^T y + c)} + \nu_n = 0, \quad \text{if } y_n = C^y_n \quad (A.7)
$$

$$
n = 1, \ldots, N, \quad y \succeq 0, \quad \nu \succeq 0. \quad (A.8)
$$

The claim that the algorithm terminates in at most $N$ iterations follows because $k$ is incremented once per iteration, and Step 4 terminates when $k = N$.

Let $\overline{k}$ be the value of $k$ at algorithm termination. To show optimality, it is sufficient to show that $y[\overline{k}+1]$ and $\nu[\overline{k}+1]$ satisfy the KKT conditions (A.5)-(A.8). This proof proceeds in two steps.

1. First, it is argued that for any $0 < m \leq \overline{k}$, the KKT conditions are satisfied for all $n \in T[m]$ by feasible $y[m+1]$ and $\nu[m+1]$.

2. Second, it is shown that for all $n \in S[\overline{k}+1]$, the condition (A.5) is satisfied by $y[\overline{k}+1]$ and $\nu[\overline{k}+1]$.

These two conditions suffice because $T[\overline{k}] \cup S[\overline{k}+1] = \{1, \ldots, N\}$ implies that the KKT conditions are satisfied for all $n$ at termination, and hence by the KKT theorem [8] $y[\overline{k}+1]$ is optimal.

The first claim is proven inductively. Consider the basis step of $k = 1$ so that $T[k] = \{1\}$ and $y[k] = 0$. If $b_k/a_k \geq (a^T y[k] + c)^{-1}(1 + a^T y[k] + c)^{-1}$, then the algorithm terminates at Step 3, and the KKT condition (A.5) is satisfied for $n = k$. 

APPENDIX A. DUAL DECOMPOSITION SUBPROBLEM

by \( y[k+1] = y[k] \) and \( \nu[k+1] = \nu[k] \). If the algorithm proceeds to Step 4, then observe that
\[
\alpha[k] \text{ is the unique positive solution of the equation } b_k/a_k = (a_k \alpha + a^T y[k] + c)^{-1} (1 + a_k \alpha + a^T y[k] + c). 
\]
If \( \beta[k] = \alpha[k] \), then (A.6) is satisfied for \( n = k \) by \( y[k+1] \) and \( \nu[k+1] \) (where \( \nu[k+1]_k = 0 \) by A.4). If \( \beta[k] = C_n y \), then because \( (a_k \alpha + a^T y[k] + c)^{-1} (1 + a_k \alpha + a^T y[k] + c) \) is monotone decreasing in \( \alpha \) for \( \alpha \geq 0 \), and \( \beta[k] \leq \alpha[k] \), it follows \( \nu[k+1] \geq 0 \) by (A.4) and (A.7) consequently is satisfied for \( n = k \) by \( y[k+1] \) and \( \nu[k+1] \).

Now consider the inductive step, so that \( k > 1 \). Recall that \( b_k/a_k \geq b_{k-1}/a_{k-1} \). If \( \beta[k-1] < C_{k-1} y \), then \( \beta[k-1] = \alpha[k-1] \) and thus
\[
\frac{b_k}{a_k} \geq \frac{b_{k-1}}{a_{k-1}} = (a^T y[k-1] + c)^{-1} (1 + a^T y[k-1] + c)^{-1}. \tag{A.9}
\]
Therefore the algorithm will terminate at Step 3 of iteration \( k \), whereupon (A.5) is satisfied for \( n = k \) by \( y[k+1] = y[k] \) and \( \nu[k+1] = \nu[k] \). Because the KKT conditions are satisfied for all \( n \in T[k-1] \) by the inductive hypothesis and \( y[k+1] = y[k] \), \( \nu[k+1] = \nu[k] \) together this implies that they are satisfied for all \( n \in T[k] \).

If instead the algorithm proceeds to Step 4, then \( y[n] = C_n y \) for each \( n \in T[k-1] \) (since otherwise termination would have occurred previously). By the exact same argument as in the basis step, (A.6) or (A.7) is satisfied for \( n = k \) by \( y[k+1] \) and \( \nu[k+1] \) as
\[
\frac{b_k}{a_k} + \frac{1}{(a^T y[k+1] + c)(1 + a^T y[k+1] + c)} = 0, \tag{A.10}
\]
if \( \alpha[k] = \beta[k] \), or
\[
\frac{b_k}{a_k} + \frac{1}{(a^T y[k+1] + c)(1 + a^T y[k+1] + c)} \leq 0, \tag{A.11}
\]
otherwise.

It remains to be shown that the KKT conditions are also satisfied for all \( n \in T[k-1] \) by \( y[k+1] \) and \( \nu[k+1] \). Because \( b_k/a_k \geq a_n/b_n \) for all \( n \in T[k-1] \), it
follows that
\[
\frac{b_n}{a_n} + \frac{-1}{(a^T y[k+1] + c)(1 + a^T y[k+1] + c)} \leq 0, \tag{A.12}
\]
and thus (A.8) is satisfied for all \( n \in T[k] \) by \( y[k+1] \) and \( \nu[k+1] \geq 0 \).

To show the second claim, observe that if the algorithm terminates at Step 4 or at Step 3 with \( \overline{k} = N \), then \( S[\overline{k} + 1] = \emptyset \) and the result is immediate. Suppose instead that termination occurs at Step 3 and that \( \overline{k} < N \). Recall that \( b_N/a_N \geq \ldots \geq b_1/a_1 \). Termination at Step 3 implies
\[
\frac{b_k}{a_k} \geq (a^T y[\overline{k} + 1] + c)^{-1}(1 + a^T y[\overline{k} + 1] + c)^{-1}, \tag{A.13}
\]
and therefore \( b_m/a_m \geq (a^T y[\overline{k} + 1] + c)^{-1}(1 + a^T y[\overline{k} + 1] + c)^{-1}, y[\overline{k} + 1]_m = 0, \) and \( \nu[\overline{k} + 1]_m = 0 \) for each \( m \in S[\overline{k} + 1] \). Thus (A.5) is satisfied for each \( m \in S[\overline{k} + 1] \) as desired.

\section*{A.2 Analytical Interpretation}

This section considers a simple and heuristic interpretation of the solution of (A.1). No rigorous arguments are given in this section; consult Section A.1 for these details. Although (A.1) has been presented strictly a mathematical problem, using an example similar to that of Chapter 2 lends itself to ready interpretation.

The problem (A.2), which is equivalent to (A.1), may be interpreted as an optimization performed by a jammer against a transmitting user. The transmitting user has fixed power equal to 1 and achieves the rate given in the logarithm term of objective in (A.2). The jammer’s choice \( y \geq 0 \) denotes the power it allocates to a number of different jamming antennas, and the fixed vector \( h \geq 0 \) denotes the gain of the power coupling from those antennas into the transmitting user. The net amount (power) of interference induced by the jammer is given by \( a^T y \).

Ideally, the jammer wishes to make the transmitting user’s rate as small as possible. However, the jammer must pay a cost for the power it uses, and this cost counts
against the reduction in the transmitting user’s rate, as reflected in the objective of (A.2). The per-antenna price of the power used by the jammer is given by the vector \( b \succ 0 \), so that the net cost experienced by the jammer is \( b^T y \).

For the moment, temporarily disregard the constraint \( y \preceq C(y) \) to simplify the problem. Consider the regime where the jammer asymptotically increases its power on any one of the antennas (say, the \( k \)th). The net jamming power \( a^T y \) increases linearly in \( y_k \), and the transmitting user’s rate decreases to 0 sub-linearly. However, the cost \( b^T y \) continues to increase linearly. Thus as \( y_k \) becomes very large, the cost term dominates and the objective value diverges to \( \infty \). Thus, for the jammer, using too much power is undesirable.

Consider the regime where the jammer asymptotically decreases its power to 0 on all antennas. In this case, the cost term \( b^T y \) tends to 0 and the objective tends to \( \log(1 + 1/c) < \infty \).

It may be the case that for small amounts of jammer power, the objective decreases more quickly than the cost term increases. In such a case, the jammer could do better by using some power instead of none. Furthermore, the jammer should use only the antenna which induces the most interference\(^2\) for a given cost, that is, has the largest ratio of \( a_k/b_k \). This is because a \( y \) with positive power on any other antenna can be strictly improved upon by transferring power to the antenna with largest \( a_k/b_k \). It is thus intuitive that the jammer should increase its power up to the point where the increase in price and decrease in transmitting user’s rate balance each other.

Now consider the power constraints \( y \preceq C(y) \), which limit the maximum power that may be allocated to a given antenna. In such a case, it is intuitive that the jammer would (starting from \( y = 0 \)) begin with the antenna having largest \( a_k/b_k \) and increase its power on that antenna up to the point where the increase in price and decrease in transmitting user’s rate balance each other. This might not be possible due to the peak power constraints. If the peak power constraints are active, then the jammer would allocate up to the maximum allowed for that antenna, and then choose the antenna with the next-highest \( a_k/b_k \). For this antenna, the jammer then increases

\(^2\)If the ratios \( a_k/b_k = a_l/b_l \) for \( k \neq l \), then without loss of generality all power may be allocated to only one antenna. Note that this does not hold when the power constraints \( y \preceq C(y) \) are included.
power up to the point where the increase in price and decrease in transmitting user’s rate balance each other. If it is not possible to allocate this much power, then the jammer proceeds to subsequent antennas in the same fashion. This algorithm is precisely that defined in Section A.1, the optimality of which is proven in Theorem 18.
Appendix B

Saddle Point Error Bounds

This Appendix adapted from [37] [8].

B.1 Formulation

This section considers the problem of computing

\[
\min_u \max_v f_0(u, v) \quad \text{subject to} \quad \hat{f}_i(u) \leq 0, \quad i = 1, \ldots, m \\
\tilde{f}_i(v) \leq 0, \quad i = 1, \ldots, \tilde{m}. \tag{B.1}
\]

where

\[
U = \{ u \in \mathbb{R}^p : \hat{f}_i(u) \leq 0, \quad i = 1, \ldots, m \}, \tag{B.2}
\]

and

\[
V = \{ v \in \mathbb{R}^q : \tilde{f}_i(v) \leq 0, \quad i = 1, \ldots, \tilde{m} \}. \tag{B.3}
\]
Assume that the functions $\hat{f}_i : \mathbb{R}^p \rightarrow \mathbb{R}$ and $\tilde{f}_i : \mathbb{R}^q \rightarrow \mathbb{R}$ are convex and differentiable (and therefore continuous [59, 6.3.1]) on $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and that the objective function $f_0 : U \times V \rightarrow \mathbb{R}$ is differentiable and continuous at all points in $U \times V$, and convex-concave. By convex-concave, it is meant that for any fixed $v' \in V$, the function $f_0(\cdot, v') : U \rightarrow \mathbb{R}$ is convex in $u$ and for any fixed $u' \in U$, the function $f_0(\cdot, u') : V \rightarrow \mathbb{R}$ is concave in $v$. It can be shown that $U$ and $V$ are closed convex sets; furthermore, assume that $U$ and $V$ are compact and nonempty, so that the minimum and maximum are well-defined and achieved (Theorem 3).

B.2 Error Bound

For any fixed $t > 0$, define $f_t : \text{int}(U) \times \text{int}(V) \rightarrow \mathbb{R}$ as

$$
f_t(u, v) = t f_0(u, v) - \sum_{i=1}^{m} \log(-\hat{f}_i(u)) + \sum_{i=1}^{\tilde{m}} \log(-\tilde{f}_i(v)), \tag{B.4}
$$

and consider solving instead

$$
\inf_u \sup_v f_t(u, v)
$$

subject to $\ (u, v) \in \text{int}(U) \times \text{int}(V), \tag{B.5}
$$

where a solution is denoted $(u^*(t), v^*(t))$. These definitions are motivated by the observation that the function $\frac{1}{t} f_t$ converges pointwise to $f_0$ on $\text{int}(U) \times \text{int}(V)$ in (B.1) as $t \rightarrow \infty$; the barrier term is heuristically an approximation to a “brick-wall” indicator function corresponding to the inequality constraints. To solve the convex-concave game (B.1), it will be shown that under certain technical conditions, one may solve a sequence of problems of the form (B.5) as $t \rightarrow \infty$. In particular, two simple bounds depending only on the problem dimensions and decreasing to zero with $t$ shall be derived.

Observe that $f_t$ is convex-concave from the convex-concave property of $f_0$, convexity of $-\log(-\hat{f}_i)$ on $\text{int}(U)$, and concavity of $\log(-\tilde{f}_i)$ on $\text{int}(V)$. Also, $f_t$ is evidently
differentiable at all points in \( \text{int}(U) \times \text{int}(V) \).

Finally, assume that \( f_t(u, v) \) has an interior saddle point, \( i.e., \) there exists some \((u^*(t), v^*(t)) \in \text{int}(U) \times \text{int}(V)\) such that:

\[
f_t(u, v^*(t)) \leq f_t(u^*(t), v^*(t)) \leq f_t(u^*(t), v)
\]

for all \( u \in \text{int}(U) \) and \( v \in \text{int}(V) \), or equivalently that \( u^*(t) \) is optimal for

\[
\min_u f_t(u, v^*(t)) \quad \text{subject to} \quad u \in \text{int}(U),
\]

and \( v^*(y) \) is optimal for

\[
\max_v f_t(u^*(t), v) \quad \text{subject to} \quad v \in \text{int}(V).
\]

Note that this assumption strengthens the earlier assumption that \( U \) and \( V \) are nonempty. Because \((u^*(t), v^*(t))\) is an interior saddle point \(1\) (in \( \text{int}(U) \times \text{int}(V) \)), the gradient of \( f_t \) with respect to \( u \), and also with respect to \( v \), vanishes in (B.7) and (B.8), respectively [62, 4.1.3]

\[
t \nabla_u f_0(u^*(t), v^*(t)) + \sum_{i=1}^{m} \frac{1}{-\hat{f}_i(u^*(t))} \nabla \hat{f}_i(u^*(t)) = 0,
\]

\[
t \nabla_v f_0(u^*(t), v^*(t)) + \sum_{i=1}^{\tilde{m}} \frac{-1}{-\tilde{f}_i(v^*(t))} \nabla \tilde{f}_i(v^*(t)) = 0.
\]

\(^1\)The saddle-point condition also (B.6) implies that \((u^*(t), v^*(t))\) achieves the infimum and supremum in both (B.5) and (B.9) below; this condition is termed the strong max-min property [8] in analogy with strong duality

\[
\sup_u \inf_v f_t(u, v) \quad \text{subject to} \quad (u, v) \in \text{int}(U) \times \text{int}(V).
\]
Note that in the following optimization problem
\[
\min_{u \in \text{int}(U)} f_0(u, v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u),
\]  
(B.12)
where \( \lambda_i = 1/(-t\hat{f}_i(u^*(t))) \), the objective is convex in \( u \) and the first-order optimality conditions [8, §3.1.3] are precisely (B.10). Because \( u^*(t) \) satisfies (B.10), \( u^*(t) \) is also optimal for (B.12).

It turns out that \( u^*(t) \) also minimizes (B.12) over the set \( U \). To see this, suppose to the contrary that there exists some \( u' \in U \) satisfying
\[
f_0(u^*(t), v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u^*(t)) - f_0(u', v^*(t)) - \sum_{i=1}^{m} \lambda_i \hat{f}_i(u') = \epsilon > 0,
\]  
(B.13)
and consider the sequence \( \{(u)_n\} \)
\[
(u)_n = \frac{1}{n} u^*(t) + \frac{n-1}{n} u',
\]  
(B.14)
which is a sequence in \( \text{int}(U) \) because (by construction of \( U \)) \( \hat{f}_i(u^*(t)) < 0, \hat{f}_i(u') \leq 0 \) implies \( \hat{f}_i((u)_n) < 0 \) for each \( i \). By continuity of (B.12) in \( u \), it must be that
\[
f_0(u', v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u') - f_0((u)_n, v^*(t)) - \sum_{i=1}^{m} \lambda_i \hat{f}_i((u)_n) \rightarrow 0.
\]  
(B.15)
And hence for some \( n < \infty \) it holds
\[
\left| f_0(u', v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u') - f_0((u)_n, v^*(t)) - \sum_{i=1}^{m} \lambda_i \hat{f}_i((u)_n) \right| < \epsilon/2,
\]  
(B.16)
whence
\[
f_0(u^*(t), v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u^*(t)) - f_0((u)_n, v^*(t)) - \sum_{i=1}^{m} \lambda_i \hat{f}_i((u)_n) > \epsilon/2,
\]  
(B.17)
which is a contradiction of the fact that \( u^*(t) \) is a minimizer of (B.12) over \( \text{int}(U) \).
Therefore for all \( u \in U \), one has
\[
f_0(u^*(t), v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u^*(t)) \leq f_0(u, v^*(t)) + \sum_{i=1}^{m} \lambda_i \hat{f}_i(u). \tag{B.18}
\]

The left hand side is equal to \( f_0(u^*(t), v^*(t)) - m/t \), and for all \( u \in U \) the second term on the right hand side is nonpositive, yielding:
\[
f_0(u^*(t), v^*(t)) \leq \min_{u \in U} f_0(u, v^*(t)) + m/t. \tag{B.19}
\]

This easily implies
\[
f_0(u^*(t), v^*(t)) - m/t \leq \min_{u \in U} f_0(u, v^*(t)) \leq \max_{v \in V} \min_{u \in U} f_0(u, v). \tag{B.20}
\]

An identical argument yields
\[
f_0(u^*(t), v^*(t)) \geq \max_{v \in V} f_0(u^*(t), v) - \tilde{m}/t, \tag{B.21}
\]

and hence
\[
f_0(u^*(t), v^*(t)) + \tilde{m}/t \geq \max_{v \in V} f_0(u^*(t), v) \geq \min_{u \in U} \max_{v \in V} f_0(u, v). \tag{B.22}
\]

Combining the bounds from (B.19) and (B.21), one obtains the following bound on the “suboptimality” of \((u^*(t), v^*(t))\) in \(f_0\)
\[
\max_{v \in V} f_0(u^*(t), v) - \min_{u \in U} f_0(u, v^*(t)) \leq \frac{m + \tilde{m}}{t}. \tag{B.23}
\]

And by combining the bounds from (B.20) and (B.22), one obtains a bound on the error of using \(f_0(u^*(t), v^*(t))\) as an estimate of (B.1)
\[
f_0(u^*(t), v^*(t)) - m/t \leq \min_{u \in U} \max_{v \in V} f_0(u, v)
\]
\[
= \max_{v \in V} \min_{u \in U} f_0(u, v) \leq f_0(u^*(t), v^*(t)) + \tilde{m}/t. \tag{B.24}
\]
B.3 Applications

B.3.1 Worst Case Interference

Applying the bound (B.24) to the WCI game (2.4) establishes Theorem 10. In particular, one makes the straightforward identifications $U = S_2$, $V = S_1$, $f_0 = J$, and the inequality constraints of (B.1) corresponding to the power, positivity, and PSD constraints of (2.4).

There are a total of $\tilde{m} = 2N + 1$ constraint functions on the victim modem ($N$ positivity constraints, $N$ tone power constraints, and 1 sum power constraint) and $m = L(1+4N)$ constraints on the interfering modems ($2LN$ positivity constraints, $2LN$ tone power constraints, and $L$ user sum power constraints).

The algorithm of Theorem 10 was proven to converge to a saddle point of (2.39), the limit point is in $\text{int}(U) \times \text{int}(V)$. Hence the assumption of Section B.2 that an interior saddle point exists is valid; the remaining assumptions of Section B.2 follow immediately.

B.3.2 Multiuser Worst Case Interference

In much the same manner as the Worst Case Interference, the bound (B.24) may be applied to the Multiuser Worst Case Interference Game $\mathcal{H}$. There are a total of $\tilde{m} = (2N + 1)(L + 1)$ constraint functions on the modems ($N$ positivity constraints, $N$ tone power constraints, and 1 sum power constraint for each of $L + 1$ lines) and similarly $m = (L + 1)(1 + 4N)$ constraints on the fictitious interferers.
Appendix C

Nash Equilibria Under an FDM Condition

This appendix proves that under the condition $\beta_n = 0$, $n = 1, \ldots, N$, the set of all Nash equilibria of the WCI game $G$ is a polytope. This condition is implied by the “FDM Condition” of Chapter 2.

Proof: Let $P$ denote the set of all Nash equilibria of $G$. The result is proven by constructing a polytope $Q$ and subsequently showing that $P = Q$. To construct $Q$, take any $(\hat{x}, [\hat{y}^{(1)}, \ldots, \hat{y}^{(N)}]) \in P$ (such a point must exist by Theorem 3). Define $D = \{n : \hat{x}_n = 0\}$, $E = \{n : 0 < \hat{x}_n < C_n^x\}$, $F = \{n : \hat{x}_n = C_n^x\}$ and $I = E \cup F$. (2.7) holds that $\hat{x}$ must be an optimum solution of the convex optimization problem

$$\max_{\mathbf{x}} \sum_{n=1}^{N} \log \left( 1 + \frac{x_n}{\alpha_n h^{(n)} \bar{y}^{(n)} + N_n} \right)$$

subject to

$$\begin{align*}
\mathbf{x} &\succeq 0, \ n = 1, \ldots, N \\
\sum_{n} x_n &\leq P^x, \\
C_x^x &\succeq \mathbf{x}.
\end{align*}$$

Associate Lagrangian dual variables $\lambda \in \mathbb{R}$ and $\nu \in \mathbb{R}^N$ with constraints (C.3) and

\footnote{The reader is referred to Chapter 2 for relevant definitions.}
(C.4) respectively. Because the objective is concave in \( x \) and Slater’s constraint qualification condition is satisfied [8] the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality (for fixed \( [\hat{y}^{(1)}, \ldots, \hat{y}^{(N)}] = [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \))

\[
\frac{1}{\alpha_n h^{(n)}(n) y^{(n)} + x_n + N_n} - \lambda \leq 0, \ \nu_n = 0 \text{ if } x_n = 0, \tag{C.5}
\]

\[
\frac{1}{\alpha_n h^{(n)}(n) y^{(n)} + x_n + N_n} - \lambda = 0, \ \nu_n = 0 \text{ if } 0 < x_n < C_n^x \tag{C.6}
\]

\[
\frac{1}{\alpha_n h^{(n)}(n) y^{(n)} + x_n + N_n} - \lambda - \nu_n = 0, \text{ if } x_n = C_n^x \tag{C.7}
\]

\[
\lambda \left( \sum_n x_n - P^x \right) = 0, x \in S_1, \ \lambda \geq 0, \ \nu \geq 0. \tag{C.8}
\]

Suppose that the KKT conditions are satisfied by the triplet \((\hat{x}, \hat{\lambda}, \hat{\nu})\). The triplet \((\hat{x}, \hat{\lambda}, \hat{\nu})\) need not be unique, in general. However the first element is unique (by Theorem 5) and thus it remains to be seen whether the ordered pair \((\hat{\lambda}, \hat{\nu})\) is unique. If \( E \neq \emptyset \), then the pair is unique. To see this, consider \( n_0 \in E \) which by (C.6) uniquely determines \( \hat{\lambda} \) and along with (C.5) and (C.7) uniquely determines \( \hat{\nu} \). Because \( 1/(\alpha_{n_0} h^{(n_0)}(n_0) \hat{y}^{(n_0)} + x_{n_0} + N_{n_0}) > 0 \) for all \( x \in S_1 \), in account of (C.6) it must be that \( \hat{\lambda} > 0 \). In this case, we define \( \hat{\lambda} = \hat{\lambda} \) and \( \hat{\nu} = \hat{\nu} \).

In the event that \( E = \emptyset \), observe that because the objective (C.1) is strictly increasing in \( x \), it must be that \( I \neq \emptyset \). Also, because \( E \subset E \cup F = I \neq \emptyset \), one has \( F \neq \emptyset \). Define

\[
\hat{\lambda} = \hat{\lambda} + \min_{m \in F} \hat{\nu}_m \tag{C.9}
\]

\[
\hat{\nu}_n = \begin{cases} 
\hat{\nu}_n - \min_{m \in F} \hat{\nu}_m & n \in F, \\
0 & \text{else.}
\end{cases} \tag{C.10}
\]

It is easily verified that \((\hat{x}, \hat{\lambda}, \hat{\nu})\) also satisfies the KKT conditions. Observe that by (C.10), \( \hat{\nu}_{n'} = 0 \) for at least one \( n' \in I \). Because \( N_n > 0 \) for all \( n \), it holds \( 1/(\alpha_{n'} h^{(n')} \hat{y}^{(n')} + x_{n'} + N_{n'}) > 0 \) for all \( x \in S_1 \). By (C.6) or (C.7), one has that \( \hat{\lambda} > 0 \). It is therefore the case that triplet \((\hat{x}, \hat{\lambda}, \hat{\nu})\) satisfies the KKT conditions and \( \hat{\lambda} > 0 \) whether \( E = \emptyset \) or \( E \neq \emptyset \).
APPENDIX C. NASH EQUILIBRIA UNDER AN FDM CONDITION

For each \( n \in D \), define \( \varphi_n \) as the solution of the equation \( 1/(\tilde{x}_n + \varphi) = \tilde{\lambda} + \tilde{\nu}_n \), namely \( \varphi_n = \frac{1}{\lambda} - \tilde{x}_n \). Define the polytope \( Q \)

\[
Q = \{(x, [y^{(1)}, \ldots, y^{(N)}]) \in S : x = \tilde{x}, \quad \alpha_n h^{(n)} y^{(n)} = \alpha_n h^{(n)} \tilde{y}^{(n)} \forall n \in I, \quad \\
\alpha_n h^{(n)} y^{(n)} + N_n \geq \varphi_n \forall n \in D \}.
\] (C.11)

It remains to be shown that \( P = Q \); it is first argued that \( Q \subset P \). Recall that \((\tilde{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) was used to construct \( Q \), and consider any other Nash equilibrium \((\bar{x}, [\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}]) \in Q \). Note that \( \bar{x} = \tilde{x} \) by construction of \( Q \). The inequality (2.8) requires that \([\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}]\) be an optimum solution of the convex optimization problem

\[
\min_{[y^{(1)}, \ldots, y^{(N)}]} \quad \sum_{n=1}^{N} \log \left( 1 + \frac{\bar{x}_n}{\alpha_n h^{(n)} y^{(n)} + \beta_n \bar{x}_n + N_n} \right)
\]
subject to \([y^{(1)}, \ldots, y^{(N)}] \in S_2 \). (C.12)

However since by Theorem 5, \( \alpha_n h^{(n)} \tilde{y}^{(n)} = \alpha_n h^{(n)} \bar{y}^{(n)} \) for all \( n \in I \), the objective value is equal and hence (2.8) is satisfied. (2.7) is equivalent to requiring the KKT conditions (C.5)-(C.8) to be satisfied for some ordered pair \((\lambda, \nu)\) where \( x = \tilde{x} \) and \([y^{(1)}, \ldots, y^{(N)}] = [\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}]\) are fixed. It is now argued that the choice of \((\lambda, \nu) = (\tilde{\lambda}, \tilde{\nu})\) satisfies the conditions. For each \( n \in \{1, \ldots, N\} \), if \( n \in D \) then \( \alpha_n h^{(n)} \tilde{y}^{(n)} + N_n \geq \varphi_n \) implies that \((\tilde{x}_n + \alpha_n h^{(n)} \tilde{y}^{(n)} + N_n)^{-1} = \tilde{\lambda} \leq 0 \) by monotonicity of \( 1/(x+a) \) in \( x \geq 0 \) for \( a > 0 \). If \( n \in I \), then \( \alpha_n h^{(n)} \tilde{y}^{(n)} + N_n = \alpha_n h^{(n)} \bar{y}^{(n)} + N_n \) by construction of \( Q \), and accordingly (C.6) or (C.7) is satisfied. Because both (2.7) and (2.8) are satisfied, it follows by definition that \((\tilde{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) and hence \( Q \subset P \).

It is now argued that \( P \subset Q \). Recall that \((\tilde{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \) was used to construct \( Q \) and consider any \((\bar{x}, [\bar{y}^{(1)}, \ldots, \bar{y}^{(N)}]) \in P \). By Theorem 5, \( \bar{x} = \tilde{x} \). Also by Theorem 5, one has \( \alpha_n h^{(n)} \tilde{y}^{(n)} = \alpha_n h^{(n)} \bar{y}^{(n)} \) for all \( n \in I \), and therefore it remains only to prove that \( \alpha_n h^{(n)} \tilde{y}^{(n)} + N_n \geq \varphi_n \) for all \( n \in D \).

Because \((\tilde{x}, [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}]) \in P \), there must exist a pair \((\tilde{\lambda}, \tilde{\nu})\) such that the triplet \((\tilde{x}, \tilde{\lambda}, \tilde{\nu})\) satisfies the KKT conditions for \([y^{(1)}, \ldots, y^{(N)}] = [\tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)}] \).
In the event that \( E \neq \emptyset \), define \( \tilde{\lambda} = \tilde{\lambda}^0 \) and \( \tilde{\nu} = \tilde{\nu}^0 \). Clearly the triplet \((\tilde{x}, \tilde{\lambda}, \tilde{\nu})\) also satisfies the same KKT conditions. Observe by Theorem 5 that because for \( n' \in E \) one has \( \alpha_n h^{(n')} \tilde{y}^{(n')} = \alpha_n h^{(n')} \tilde{y}^{(n')} \), it follows by (C.6) that \( \tilde{\lambda} = \hat{\lambda} \).

In the event that \( E = \emptyset \), observe that because \( \emptyset = E \subset I \neq \emptyset \), we have \( F = I - E \neq \emptyset \). Define

\[
\tilde{\lambda} = \lambda^0_n + \min_{m \in F} \tilde{v}^0_m, \tag{C.13}
\]
\[
\tilde{v}_n = \begin{cases} 
\nu^0_n - \min_{m \in F} \nu^0_m & n \in F \\
0 & \text{else}
\end{cases} \tag{C.14}
\]

It may be readily verified that the triplet \((\tilde{x}, \tilde{\lambda}, \tilde{\nu})\) satisfies the KKT conditions for \([y_1^{(1)}, \ldots, y^{(N)}] = [\tilde{y}_1^{(1)}, \ldots, \tilde{y}^{(N)}]\). By (C.14), there must exist some \( n' \in F \) such that \( \tilde{v}_{n'} = 0 \). Similarly, recall that there must exist some \( m' \in F \) such that \( \tilde{v}_{m'} = 0 \). It is now argued that there exists some \( m \in F \) such that both \( \tilde{v}_m = 0 \) and \( \tilde{v}_{m} = 0 \). In particular, let \( m = m' \). Then by (C.7) and the fact that the triplet \((\tilde{x}, \tilde{\lambda}, \tilde{\nu})\) satisfies the KKT conditions for \([y_1^{(1)}, \ldots, y^{(N)}] = [\tilde{y}_1^{(1)}, \ldots, \tilde{y}^{(N)}]\), one has

\[
\frac{1}{\alpha_n h^{(n')} \tilde{y}^{(n')} + x_n + N_n} \leq \frac{1}{\alpha_m h^{(m)} \tilde{y}^{(m)} + x_m + N_m} \leq \frac{1}{\alpha_n h^{(n)} \tilde{y}^{(n)} + x_n + N_n} \text{ for all } n \in F.
\]

However, \( \alpha_n h^{(n')} \tilde{y}^{(n')} = \alpha_n h^{(n')} \tilde{y}^{(n')} \) for all \( n \in F \), and therefore

\[
\frac{1}{\alpha_n h^{(n')} \tilde{y}^{(n')} + x_n + N_n} \leq \frac{1}{\alpha_n h^{(n')} \tilde{y}^{(n')} + x_n + N_n} \text{ for all } n \in F.
\]

This (along with the fact that \( \tilde{v}_{n'} = 0 \) for some \( n' \in F \)) implies that \( \tilde{v}_m = 0 \). Then (C.7) for this choice of \( m \) implies that \( \tilde{\lambda} = \hat{\lambda} \).

Because it is always the case that \( \hat{\lambda} = \tilde{\lambda} \), the triplet \((\tilde{x}, \hat{\lambda}, \tilde{\nu})\) satisfies the KKT conditions (for \([y_1^{(1)}, \ldots, y^{(N)}] = [\tilde{y}_1^{(1)}, \ldots, \tilde{y}^{(N)}]\)). Therefore

\[
\frac{1}{\alpha_n h^{(n')} \tilde{y}^{(n')} + x_n + N_n} - \hat{\lambda} \leq 0 \text{ for all } n \in D \text{ implies that } \alpha_n h^{(n)} \tilde{y}^{(n)} + N_n \geq \varphi_n \text{ for all } n \in D.
\]

Thus \((\tilde{x}, [\tilde{y}_1^{(1)}, \ldots, \tilde{y}^{(N)}]) \in Q.\)
Bibliography


