THÈSE

en vue de l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ DE LORRAINE

(arrêté ministériel du 7 Août 2006)

Spécialité INFORMATIQUE

présentée par

TRAN THI THUY

Titre de la thèse :

LA PROGRAMMATION DC ET DCA POUR CERTAINES CLASSES DE PROBLÈMES DANS LES SYSTÈMES DE COMMUNICATION SANS FIL

—

DC PROGRAMMING AND DCA FOR SOME CLASSES OF PROBLEMS IN WIRELESS COMMUNICATION SYSTEMS

soutenue le 24 Avril 2017

Composition du Jury :

Rapporteurs	Christian PRINS	*Professeur, Université de Technologie de Troyes*
	Yaroslav D. Sergeyev	*Professeur, Université de Calabria*
Examinateurs	Tao PHAM DINH	*Professeur, INSA de Rouen*
	Nathalie SAUER	*Professeur, Université de Lorraine*
	Viet Hung NGUYEN	*MCF, Université de Paris 6*
Directrice de thèse	Hoai An LE THI	*Professeur, Université de Lorraine*
Co-encadrant	Alain GÉLY	*MCF, Université de Lorraine*

THÈSE PRÉPARÉE AU SEIN DE LABORATOIRE
D'INFORMATIQUE THÉORIQUE ET APPLIQUÉE (LITA)
UNIVERSITÉ DE LORRAINE, METZ, FRANCE

Remerciements

En premier lieu, je tiens à exprimer ma profonde gratitude à ma directrice de thèse, Madame Hoai An LE THI, Professeur des Universités à l'Université de Lorraine, pour son soutien permanent et ses précieux conseils tout au long de mes études. Sans son encadrement, il m'aurait été impossible de finaliser cette thèse. Grâce à ses connaissances approfondies dans le domaine de l'optimisation, à la fois théoriques et appliquées, ses compétences pédagogiques et son expérience, Madame LE THI m'a aidée à surmonter les difficultés initiales rencontrées dans mes travaux de recherche, puis guidée, étape par étape, pour acquérir les connaissances et les compétences nécessaires pour enrichir ma thèse. Ses commentaires critiques m'ont permis d'acuérir progressivement une maturité dans le domaine de la recherche scientifique. Elle a su faire preuve d'un grand enthousiasme et d'une patience remarquable lors de la lecture-relecture de ma rédaction et l'encadrement de mon travail. Je lui suis très reconnaissante pour toute l'attention qu'elle m'a accordée au cours plus des trois années de doctorat.

En second lieu, je souhaite exprimer ma sincère gratitude à mon second encadrant de thèse, Monsieur Alain Gély, Maître de Conférences à l'Université de Lorraine, pour m'avoir permis de bénéficier de ses connaissances scientifiques et pour son soutien. Je voudrais le remercier également pour ses remarques qui m'ont permis d'améliorer la rédaction de certaines parties de la thèse.

Je souhaite aussi exprimer toute ma gratitude à Monsieur Tao Pham Dinh, Professeur des Universités à l'INSA de Rouen, pour avoir créé la programmation DC et DCA qui sont des outils puissants d'optimisation non convexe et connaissent un grand succès dans la résolution de nombreux problèmes d'application dans divers domaines de sciences appliquées. Je voudrais le remercier pour les connaissances précieuses et les documents très intéressants qu'il a partagés avec moi. Sa passion pour la recherche est un excellent exemple pour nous tous. Sa connaissance profonde et riche et ses exposés toujours sources d'inspiration nous motivent pour étudier plus et ouvrir notre esprit à la culture scientifique. Je voudrais le remercier pour m'avoir fait l'honneur de siéger au jury de ma thèse

Je souhaite remercier vivement Professeur Christian PRINS à l'Université de Technologie de Troyes et Professeur Yaroslav D. Sergeyev à l'Université de Calabria, pour m' avoir fait l'honneur d'être rapporteurs de ma thèse et pour leurs temps précieux consacrés.

Je souhaite remercier également Madame Nathalie Sauer, Professeur des Universités à

l'Université de Lorraine, et Monsieur Viet Hung NGUYEN, Maître de conférences au Laboratoire d'Informatique de Paris 6, pour leur participation au jury de soutenance de thèse.

Je remercie tous mes amis du LITA et de Metz : Manh Cuong, Bich Thuy, Minh Thuy, Anh Vu, Minh Tam, Hoai Minh, Duy Nhat, Xuan Thanh, Vinh Thanh, Tran Bach, Sara,...pour leur aide et leurs encouragements, ainsi que pour les agréables moments passés ensemble lors de mon séjour en France. Je remercie particulièrement Dr. Xuan Thanh Vo pour sa disponibilité et son aide constante. J'ai acquis en le côtoyant un grand nombre de connaissances et compétences nécessaires à mes recherches. Je voudrais remercier également Dr. Manh Cuong Nguyen, et M. Vinh Thanh Ho pour leurs partages de connaissances en matière de programmation. Je voudrais remercier M. Nhu Tuan Nguyen pour les discussions très intéressantes que nous avons eues lors de notre collaboration.

Je remercie ma famille pour leur amour et soutien illimité. Je suis particulièrement reconnaissante envers mes parents, mon mari et mes plus jeunes soeurs pour avoir pris soin de mes enfants pendant que j'étudiais loin de chez nous.

Enfin, je souhaite remercier le Gouvernement Vietnamien pour son soutien financier dans la réalisation de mes études en France. Je n'oublie pas de remercier l'Université de FPT qui a toujours soutenu ma démarche de poursuite d'études dans l'enseignement supérieur français en prenant toutes les dispositions nécessaires internes pour permettre et pallier mon absence de l'établissement.

TRAN Thi Thuy

Né le 03 Juin, 1979 (Viet Nam)

Tél: 06 95 90 98 19

E-mail: thi-thuy.tran@univ-lorraine.fr

Adresse personnelle: P6201, Res Univ Saulcy, Ile du Saulcy, 57010 Metz

Adresse professionnelle: Bureau E408, LITA – Université de Lorraine, Ile du Saulcy, 57045 Metz

Situation Actuelle

Depuis Décembre 2013	Doctorant au Laboratoire d'Informatique Théorique et Appliquée (LITA EA 3097) de l'Université de Lorraine. Encadré par Prof. Hoai An Le Thi et MCF. Alain Gély. Sujet de thèse : **"La Programmation DC et DCA pour certaines classes de problèmes dans les systèmes de communication sans fil"**

Experience Professionnelle

2001–2008	Enseignant, Université des ressources en eau - Hanoi Ville, Vietnam
2009–2013	Enseignant, Université de FPT - Hanoi Ville, Vietnam.

Diplôme et Formation

2013 au present	Doctorant en Informatique. LITA–Université de Lorraine - Metz Ville, France.
2001–2003	Master en Mathématiques, Université de Hanoi de l'Éducation - Hanoi Ville, Vietnam.
1997–2001	Diplôme universitaire en Mathématiques, Université de Hanoi de l'Éducation - Hanoi Ville, Vietnam.

Publications

Refereed international journal papers

[1] Thi Thuy Tran, Hoai An Le Thi, Tao Pham Dinh. DC programming and DCA for Enhancing Physical Layer Security via Cooperative Jamming. *Publish online in 18 November 2016* in Journal of Computers and Operations Research (http://dx.doi.org/10.1016/j.cor.2016.11.003).

[2] Hoai An Le Thi, Thi Thuy Tran, Tao Pham Dinh, Nhu Tuan Nguyen. DC programming and DCA for Enhancing Physical Layer Security via Relay Beamforming Strategies. *Submitted* to Applied Mathematics and Computation.

Refereed papers in books / Refereed international conference papers

[1] Hoai An Le Thi, Thi Thuy Tran, Tao Pham Dinh, Alain Gély. DC Programming and DCA for Transmit Beamforming and Power Allocation in Multicasting Relay Network. in H.A. Le Thi et al. (Eds.), ICCSAMA 2016: Proceedings of the 4th International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2016, AISC 453, pp 29-41, Springer 2016.

[2] Tran Thi Thuy, Nguyen Nhu Tuan, Le Thi Hoai An and Alain Gély. DC programming and DCA for Enhancing Physical Layer Security via Relay Beamforming Strategies. in N.T. Nguyen et al. (Eds.), ACIIDS 2016: Proceedings of the 8th Asian Conference on Intelligent Information and Database Systems, LNCS 9622, pp 640-650, Springer 2016.

[3] Thi Thuy Tran, Hoai An Le Thi, Tao Pham Dinh. DC Programming and DCA for a Novel Resource Allocation Problem in Emerging Area of Cooperative Physical Layer Security. in H.A. Le Thi et al. (Eds.), ICCSAMA 2015: Proceedings of 3rd International Conference on Computer Science, Applied Mathematics and Applications, AISC 358, pp. 57-68, Springer 2015.

[4] Hoai An Le Thi, Thi Thuy Tran. DC programming and DCA for secrecy rate maximization in a wireless network with multiple eavesdroppers. *accepted by* ACACOS'17.

Communications in national / International conferences

[1] Thi Thuy Tran, Hoai An Le Thi, Tao Pham Dinh. DC Programming and DCA for transmit Beamforming and Power Allocation in Multicasting Relay Network. Presentation in The 27th European Conference on Operational Research, Glasgow, UK, July 12 - 15, 2015.

[2] Thi Thuy Tran, Hoai An Le Thi, Tao Pham Dinh. DC programming and DCA for DEFT (Distributed DC programming and DCA for DEFT). Presentation in The 20th Conference of the International Federation of Operational Research Societies (IFORS 2014) Barcelona, Spain, July 13 - 18, 2014.

Contents

List of Figures

List of Tables

Abbreviations and Notations

Throughout the dissertation, we use bold letters to denote matrices and vectors, and normal letters for scalars. Vectors are also regarded as matrices with one column. The table below summarizes some of the abbreviations and notations used in the dissertation.

DC	Difference of convex functions
DCA	DC Algorithms
AF	Amplify and Forward
DF	Decode and Forward
CJ	Cooperative Jamming
BF	Beamforming
SNR	Signal to Noise Ratio
SRM	Secrecy Rate Maximization
QoS	Quality of Service
SDP	Semidefinite Program
SDR	Semidefinite Relaxation
CCCP	Convex-Concave Procedure
QCQP	Quadratically Constrained Quadratic optimization Problem
MIMO	Multiple-Input Multiple-Output
SIMO	Single-Input Multiple-Output
MISO	Multiple-Input Single-Output
SCA	Successive Convex Approximation

$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbb{R}^n$	set of real column vectors of size n
$\mathbb{C}^n$	set of complex column vectors of size n
$\mathbb{R}^{m\times n}$	set of real matrices of size m - by - n
$\|\cdot\|$	Euclidean norm, $\|\mathbf{x}\| = (\sum_{i=1}^{n} \|x_i\|^2)^{1/2}$, $\mathbf{x} \in \mathbb{R}^n$
	matrix ℓ_2-norm/spectral norm,
	$\|\mathbf{X}\| = \max_{\mathbf{u}\in\mathbb{R}^n, \|\mathbf{u}\|=1} \|\mathbf{X}\mathbf{u}\|$, $\mathbf{X} \in \mathbb{R}^{m\times n}$
$\langle\ ,\ \rangle$	scalar product, $\langle \mathbf{X}, \mathbf{Y}\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}Y_{ij}$,
	$\mathbf{X}$,$\mathbf{Y} \in \mathbb{R}^{m\times n}$
$\mathbf{X}(i,:)$	i^{th} row of $\mathbf{X}$
$\mathbf{X}(:,j)$	j^{th} column of $\mathbf{X}$
X_{ij}	element located at the position (i,j) of $\mathbf{X}$
$\mathbf{X}^T$	transpose of a matrix $\mathbf{X}$, $(\mathbf{X}^T)_{ij} = \mathbf{X}_{ji}$
$\mathbf{X}^*$	conjugate of a matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	conjugate transpose of a matrix $\mathbf{X}$
$\mathrm{diag}(\mathbf{X})$	vector of diagonal-elements of X, $(\mathrm{diag}(\mathbf{X}))_i = X_{ii}$
$\mathrm{diag}(\mathbf{x})$	diagonal matrix whose the main diagonal is the
	vector $\mathbf{x}$
$\mathrm{blkdiag}([\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_n])$	denotes the block diagonal matrix formed from
	the matrices $\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_n$.
$\mathrm{vec}(\mathbf{X})$	vector formed by stacking the columns of
	$\mathbf{X} \in \mathbb{R}^{m\times n}$ into one vector of size mn
$\mathrm{tr}(\mathbf{X})$	the trace of matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(\mathbf{X}) = \sum_i^n X_{ii}$
$E\{\mathbf{V}\}$	expectation of the variable $\mathbf{V}$
$\mathrm{Re}(x)$ and $\mathrm{Im}(x)$	the real part and the imaginary part of the
	complex number x
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker product between matrices $\mathbf{X}$ and $\mathbf{Y}$
$\mathbf{X} \preceq \mathbf{Y}$	$\mathbf{Y} - \mathbf{X}$ is positive semi-definite matrix
	(all eigenvalues are nonnegative)
$\mathbf{I}_n$	identity matrix of size n
$\mathrm{Proj}_\Omega(\mathbf{X})$	projection of $\mathbf{X} \in \mathbb{R}^{m\times n}$ onto $\Omega \subset \mathbb{R}^{m\times n}$
$[\mathbf{X}]_+$	projection of $\mathbf{X} \in \mathbb{R}^{m\times n}$ onto $\mathbb{R}_+^{m\times n}$,
	$[\mathbf{X}]_+ = \max(\mathbf{X}, 0)$
$\chi_C(\cdot)$	the indicator function of C, $\chi_C(x) = 0$ if $x \in C$
	and $+\infty$ otherwise
$\nabla f(x)$	the gradient of f at x
$\nabla^2 f(x)$	the Hessian of f at x
$\partial f(x)$	the subdifferential of f at x
$\mathcal{CN}(0, \Gamma)$	the circularly symmetric complex Gaussian distribution

Résumé

La communication sans fil joue un rôle de plus en plus important dans de nombreux domaines. Un grand nombre d'applications sont exploitées tels que l'e-banking, l'e-commerce, les services médicaux, ... Ainsi, la qualité de service (QoS), et la confidentialité d'information sur le réseau sans fil sont primordiales dans la conception du réseau sans fil. Dans le cadre de cette thèse, nous nous concentrons sur le développement des approches d'optimisation pour résoudre certains problèmes concernant les deux sujets suivants : la qualité de service et la sécurité de la couche physique. Nos méthodes sont basées sur la programmation DC (Difference of convex functions) et DCA (DC Algorithms) qui sont reconnues comme de puissants outils d'optimisation non convexes et non différentiables. Ces outils ont connu de grands succès au cours des deux dernières décennies dans la modélisation et la résolution de nombreux problèmes d'applications dans divers domaines de sciences appliquées.

Outre les chapitres d'introduction et de conclusion, le contenu principal de cette thèse est divisé en quatre chapitres: Le chapitre 2 concerne la QoS dans les réseaux sans fil tandis que les trois chapitres suivants étudient la sécurité de la couche physique. Le chapitre 2 considère un critère de QoS qui consiste à assurer un service équitable entre les utilisateurs dans un réseau sans fil. Plus précisement, on doit s'assurer qu'aucun utilisateur ne souffre d'un mauvais rapport signal sur bruit ("signal to noise ratio (SNR)" en anglais). Le problème revient à maximiser le plus petit SNR. Il s'agit donc un problème d'optimisation DC général (minimisation d'une fonction DC sur un ensemble défini par des contraintes convexes et des contraintes DC). La programmation DC et DCA ont été développés pour résoudre ce problème. Tenant compte de la structure spécifique du problème, nous avons proposé une nouvelle décomposition DC qui était plus efficace que la précédente décomposition. Une méthode de résolution basée sur la programmation DC et DCA a été développée. De plus, nous avons prouvé la convergence de notre algorithme.

L'objectif commun des trois chapitres suivants (Chapitre 3, 4, 5) est de garantir la sécurité de la couche physique d'un système de communication sans fil. Nous nous concentrons sur l'approche qui consiste à maximiser le taux de secret ("secrecy rate" en anglais). Trois diverses architectures du réseau sans fil utilisant différentes techniques coopératives pour la transmission sont considérées dans ces trois chapitres. Dans le chapitre 3, nous considérons un réseau point-à-point utilisant une technique coopérative de brouillage. Le chapitre 4 étudie un réseau de relais utilisant une combi-

naison de technique de formation de faisceau (“beamforming technique” en anglais) et de technique de relais coopératifs. Deux protocoles de technique de relais coopératifs, Amplifier-et-Transmettre (“Amplify-and-Forward (AF)”) et Décoder-et-Transmettre (“Decode-and-Forward (DF)” en anglais), sont considérés. Dans le chapitre 3 et le chapitre 4, nous considérons qu'il y a seulement un espion (“eavesdropper” en anglais) dans le réseau tandis que le chapitre 5 est une extension du chapitre 4 où on peut avoir plusieurs espions. Tous ces problèmes sont des problèmes d'optimisation non-convexes qui peuvent être ensuite reformulés sous forme d'une programmation DC pour lesquels nous développons les méthodes efficaces et robustes basées sur la programmation DC et DCA. Dans les chapitres 3 et 4, nous reformulons les problèmes étudiés sous forme d'un programme DC standard (minimisation d'une fonction DC avec les contraintes convexes). La structure spécifique est bien exploitée afin de concevoir des schémas DCA standard efficaces où les sous-problèmes convexes de ces schémas sont résolus soit explicitement soit de manière peu coûteuse. Les problèmes d'optimisation dans le chapitre 5 sont reformulés comme les programmes DC généraux et les schémas DCA généraux sont développés pour résoudre ces problèmes. Les résultats obtenus montrent la supériorité de nos approches par rapport aux méthodes existantes. La convergence des schémas DCA proposés a été rigoureusement étudiée.

Abstract

Wireless communication plays an increasingly important role in many aspects of life. A lot of applications of wireless communication are exploited to serve people's life such as e-banking, e-commerce and medical service. Therefore, quality of service (QoS) as well as confidentiality and privacy of information over the wireless network are of leading interests in wireless network designs. In this dissertation, we focus on developing optimization techniques to address some problems in two topics: QoS and physical layer security. Our methods are relied on DC (Difference of Convex functions) programming and DCA (DC Algorithms) which are powerful, non-differentiable, non-convex optimization tools that have enjoyed great success over the last two decades in modelling and solving many application problems in various fields of applied science.

Besides the introduction and conclusion chapters, the main content of the dissertation is divided into four chapters: the chapter 2 concerns QoS in wireless networks whereas the next three chapters tackle physical layer security. The chapter 2 discusses a criterion of QoS assessed by the minimum of signal-to-noise (SNR) ratios at receivers. The objective is to maximize the minimum SNR in order to ensure the fairness among users, avoid the case in which some users have to suffer from a very low SNR. We apply DC programming and DCA to solve the derived max-min fairness optimization problem. With the awareness that the efficiency of DCA heavily depends on the corresponding DC decomposition, we recast the considered problem as a general DC program (minimization of a DC function on a set defined by some convex constraints and some DC constraints) using a DC decomposition different from the existing one and design a general DCA scheme to handle that problem. The numerical results reveal the efficiency of our proposed DCA compared with the existing DCA and the other methods.

In addition, we rigorously prove the convergence of the proposed general DCA scheme.

The common objective of the next three chapters (Chapter 3,4,5) is to guarantee security at the physical layer of wireless communication systems based on maximizing their secrecy rate. Three different architectures of the wireless system using various cooperative techniques are considered in these three chapters. More specifically, a point-to-point wireless system including single eavesdropper and employing cooperative jamming technique is considered in the chapter 3. Chapter 4 is about a relay wireless system including single eavesdropper and using a combination of beamforming technique and cooperative relaying technique with two relaying protocols Amplify-and-Forward (AF) and Decode-and-Forward (DF). Chapter 5 concerns a more general relay wireless system than the chapter 4, in which multiple eavesdroppers are considered instead of single eavesdropper. The difference in architecture of wireless systems as well as in the utilized cooperative techniques result in three mathematically different optimization problems. The unified approach based on DC programming and DCA is proposed to deal with these problems. The special structures of the derived optimization problems in the chapter 3 and the chapter 4 are exploited and explored to design efficient standard DCA schemes in the sense that the convex subproblems in these schemes are solved either explicitly or in an inexpensive way. The max-min forms of the optimization problems in the chapter 5 are reformulated as the general DC programs with DC constraints and the general DCA schemes are developed to address these problems. The results obtained by DCA show the efficiency of our approach in comparison with the existing methods. The convergence of the proposed general DCA schemes is thoroughly shown.

Introduction générale

Cadre général et motivations

L'augmentation exponentielle des techniques de diffusions sans fil a permis de nombreux protocoles de communications et services afin d'améliorer la vie des gens. Cette augmentation abouti à une plus grande demande de qualité de service (QoS), de sécurité et de confidentialité des données utilisateurs. Par suite, ces problèmes sont au premier plan dans la conception de systèmes sans fil.

Plusieurs critères permettent de mesurer quantitativement la qualité de service : taux d'erreur, bande passante, débit, rapport signal/bruit, etc. Dans cette thèse, la qualité de service est évaluée par le rapport signal sur bruit minimum des récepteurs. Cette quantité est à maximiser sous contraintes de puissances, dans le but d'éviter les cas ou des utilisateurs souffriraient d'un rapport signal/bruit extrêmement faible. Ce critère nous permet d'assurer une équité entre utilisateurs.

Assurer la confidentialité des données est un aspect important de la communication sans fil. La nature du medium permet l'écoute clandestine ("eavedropping" en anglais) et les attaques indues. Traditionnellement, ce problème est traité par des méthodes cryptographiques. Ces méthodes supposent qu'il est impossible pour un espion ("eavesdropper" en anglais) de décrypter les données qui ont été chiffrées sans en posséder la clé. Cependant, l'augmentation rapide de la puissance de calcul et l'arrivée des ordinateurs quantiques sont des menaces pour les systèmes de cryptographie. Par suite, il est nécessaire de créer et développer des méthodes alternatives à la cryptographie pour assurer la sécurité des systèmes.

Dans ce contexte, la sécurité de la couche physique attire une grande attention de la communauté scientifique. Le principal objectif de ces approches est de rendre les écoutes illicites impossibles. Ces dernières sont des attaques passives consistant à espionner le canal de communication pour capter de l'information confidentielle. Wyner est un des pionniers à avoir mis en place les fondements théoriques de ces approches avec son étude révolutionnaire ([114]). Dans cette étude, il considère une situation dans laquelle le canal d'écoute d'un espion est plus bruité que le canal d'agent autorisé. Il montre qu'il est possible d'obtenir un taux de secret ("secrecy rate" en anglais) strictement positif sans avoir recours à la cryptographie.

Plusieurs travaux ont étendu les résultats de Wyner à des problèmes plus généraux comme des canaux d'écoute à bruit gaussien ([57]), la diffusion sur canaux non dégradés ([13]), canaux à amortissement lents ([26]) et systèmes multi-utilisateurs de type MIMO (Multiple Input, Multiple Output) ([19]). En parallèle à ces techniques, d'autres sont mises en place pour améliorer la confidentialité des échanges, par exemple par un jeu de coopération entre utilisateurs autorisés. De telles techniques de coopération sont par exemple le brouillage coopératif ("cooperative jamming" en anglais), ou les relais coopératifs ("cooperative relaying" en anglais), avec deux protocoles bien connus ; Amplifier-et-Transmettre ("Amplify-and-forward" (AF) en anglais) et Décoder-et-Transmettre ("Decode-and-Forward" (DF) en anglais) ainsi que la formation de faisceau coopérative ("cooperative beamforming" en anglais).

Dans cette thèse, nous nous concentrons sur l'utilisation de techniques d'optimisation pour résoudre efficacement les problèmes issus des domaines de la qualité de service et de la sécurité de la couche physique. Pour la qualité de service, nous traitons le problème de maximisation du rapport signal/bruit (SNR) minimum, sous contraintes de puissances dans un réseau multi-diffusion ("multicasting" en anglais). Pour la sécurité de la couche physique, nous étudions le problème de maximisation du taux de secret sous certaines contraintes de puissances pour plusieurs architectures de communication sans fil et avec plusieurs techniques de coopérations. Tout au long de cette thèse, nous considérons un système de communication sans fil où une ou plusieurs sources S envoient un message à une ou plusieurs destinations D. La communication peut ne pas être directe entre S et D, mais passer par des relais R de communications. Ces relais peuvent coopérer pour retransmettre le signal vers les destinations avec une bonne qualité de service. Pour ce faire, les relais utilisent des méthodes comme la formation de faisceaux et utilisent des protocoles de transmissions de type Amplifier-et-Transmettre ou Décoder-et-Transmettre. De même, ces relais peuvent jouer le rôle de brouilleurs alliés pour empêcher un espion de recevoir correctement le message transmis.

En général, ces problèmes d'optimisation sont non convexes et/ou non différentiables et par suite, difficiles à traiter. Nous étudions une approche unifiée basée sur la programmation DC (Difference of Convex Function) et DCA (Difference of Convex function Algorithm) pour résoudre ces problèmes.

La programmation DC et DCA (DC Algorithm) sont des outils puissants d'optimisation non convexe. Ces outils connaissent un grand succès, au cours des deux dernières décennies, dans la résolution de nombreux problèmes d'application dans divers domaines de sciences appliquées en général ([48], [49], [55], [56], [78], [79], [80] and references therein), et des systèmes de communication en particulier (voir par exemple ([106], [128], [104], [36], [4], [50], [52], [93], [94], [95], [53] et la liste des références dans [Le Thi]). De nombreuses expérimentations numériques réalisées dans cette thèse ont prouvé l'efficacité, la scalabilité, la rapidité des algorithmes proposés et leur supériorité par rapports aux méthodes standards. La programmation DC et DCA considèrent le problème DC de la forme

$$\alpha = \inf\{f(x) := g(x) - h(x) \ : \ x \in \mathbb{R}^n\} \quad (P_{dc}),$$

où g et h sont des fonctions convexes définies sur $\mathbb{R}^n$ et à valeurs dans $\mathbb{R} \cup \{+\infty\}$, semi-continues inférieurement et propres. La fonction f est appelée fonction DC avec les composantes DC g et h, et $g-h$ est une décomposition DC de f. DCA est basé sur la dualité DC et des conditions d'optimalité locale. La construction de DCA implique les composantes DC g et h et non la fonction DC f elle-même. Or chaque fonction DC admet une infinité de décompositions DC qui influencent considérablement sur la qualité (la rapidité, l'efficacité, la globalité de la solution obtenue,...) de DCA. Ainsi, au point de vue algorithmique, la recherche d'une "bonne" décomposition DC et d'un "bon" point initial est très importante dans le développement de DCA pour la résolution d'un programme DC.

L'utilisation de la programmation DC et DCA dans cette thèse est justifiée par de multiples arguments ([80]):

- On a assisté ces derniers temps à une augmentation de l'utilisation de la programmation DC et DCA pour résoudre des modèles d'optimisation non convexes non différentiables difficiles dans les systèmes de communication sans fil. En particulier, l'étude récente des DCAs généraux, qui sont des extensions des DCAs standard permet de traiter une classe plus large de problèmes d'optimisation non convexes, ce qui amène des applications plus larges de ces outils dans ce domaine. Avec des techniques appropriées, la plupart des modèles d'optimisation dans les systèmes de communication sans fil peuvent être reformulés comme des programmes DC standard ou généraux et être ainsi éventuellement résolus par des schémas DCA standard ou généraux. L'efficacité de la programmation DC et DCA a été démontrée dans de nombreux travaux dans ce domaine.
- DCA est une philosophie plutôt qu'un algorithme. Pour chaque problème, nous pouvons concevoir une famille d'algorithmes basés sur DCA. La flexibilité de DCA sur le choix de décomposition DC peut offrir des schémas DCA plus performants que des méthodes standards.
- L'analyse convexe fournit des outils puissants pour prouver la convergence de DCA dans un cadre général. Ainsi tous les algorithmes basés sur DCA bénéficient (au moins) des propriétés de convergence générales du schéma DCA générique qui ont été démontrées.

Il est important de noter qu'avec les techniques de reformulation en programmation DC et les décompositions DC appropriées, on peut retrouver la plupart des algorithmes existants en programmation convexe/non convexe comme cas particuliers de DCA.

Nos contributions

Les principales contributions de la thèse résident dans le développement de techniques d'optimisation pour résoudre certaines classes de problèmes dans les systèmes de communication sans fil. Nous développons la programmation DC et DCA pour aborder les problèmes issus des domaines de la qualité de service et de la sécurité de la couche physique. Tout au long de la thèse, les deux questions cruciales dans le développement de DCA ont bien été étudiées pour chaque problème considéré, à savoir la recherche de bonnes décompositions DC et la résolution des sous-problèmes

convexes dans les schémas DCAs. Plus précisement, nous analysons la structure particulière des problèmes d'optimisation dans les deux chapitres 3 et 4 pour introduire les décompositions DC efficaces dans le sens où elles conduisent à des sous-problèmes explicitement résolus ou faciles à résoudre. En outre, nous proposons le schéma DCA distribué dans le chapitre 3 qui permet de résoudre les problèmes d'optimisation de grande taille souvent rencontrés dans les systèmes de communication, qui restent un défi pour la plupart des solveurs disponibles. De plus, les deux chapitres 2 et 5 consistent à développer les nouvelles approches de la programmation DC, appelées "DCA général" pour aborder des problèmes DC généraux qui concernent la minimisation d'une fonction DC sur un ensemble convexe avec en plus des contraintes DC. Etant une extension de DCA standard, cette nouvelle approche DCA général, qui émerge depuis quelques années de par ses applications dans nombreux domaines, ouvre des voies prometteuses. Ces deux chapitres de la thèse comportent des contributions significatives au développement de DCA général et ses applications dans les systèmes de communication sans fil.

Plus en détail, au chapitre 2, nous étudions le problème d'optimisation de l'équité max-min issu du domaine de la qualité de service. Ce problème est non convexe et non différentiable. Nous reformulons ce problème comme un programme DC général avec une nouvelle décomposition DC et nous proposons un schéma DCA général pour le résoudre. La décomposition DC proposée amène à des sous problèmes plus simples parce que les contraintes DC sont approximées par des contraintes quadratiques convexes à la place de contraintes fractionnaires convexes comme dans le schéma DCA existant. Nous prouvons la convergence globale du schéma DCA proposé. Nous adaptons la preuve de convergence d'un schéma DCA général générique à notre situation. Plus particulièrement, dans le schéma DCA général générique, pour éviter l'infaisabilité des sous-problèmes susceptibles d'être causés par l'approximation des contraintes DC, une variable d'écart est introduite et pénalisée à la fonction objectif pour résoudre les sous-problèmes résultant. La mise à jour de ces coefficients de pénalités est alors requise. Cependant, dans notre schéma DCA, la faisabilité des sous-problèmes est assurée sans avoir à introduire une variable d'écart. Par suite, la preuve de convergence de notre schéma DCA est réduite comparativement à celle d'un schéma DCA général générique. Cependant, la difficulté d'adaptation de la preuve à notre situation vient de la perte de la forte convexité des composants DC. En effet, pour prouver la convergence d'un schéma DCA général générique, il faut au moins que l'un des composants DC de la fonction objectif ou que les contraintes DC soient fortement convexe, ce qui n'est pas le cas dans notre problème. Pour surmonter cette difficulté, nous exploitons la propriété de forte convexité dans seulement une partie des variables des décompositions DC dans les contraintes DC pour parvenir à un résultat similaire à celui du lemme 1 de [47]. Il s'agit d'une étape importante pour montrer la convergence du schéma DCA présenté. Il est à noter que, bien que la convergence du système DCA existant ait été proposée, nous nous rendons compte d'un argument lâche pour montrer la propriété de clôture de l'algorithme, qui est l'une des trois conditions de théorème de convergence globale de Zangwill. Ainsi, la convergence du schéma DCA existant devra être examinée très attentivement.

Dans le chapitre 3, nous traitons le problème de maximisation du taux de secret ("se-

crecy rate maximization (SRM)" en anglais) dans un système de communication sans fil point-à-point composé de l'espion unique et de multiples brouilleurs alliés, qui utilisent le brouillage coopératif. Nous exploitons la structure particulière de la fonction objectif d'une manière appropriée afin de proposer une nouvelle décomposition DC. Le DCA résultant implique un sous-problème convexe qui est un programme quadratique fortement convexe. Il peut donc être efficacement résolu de manière centralisée par les logiciels standards. En outre, la fontion objectif quadratique convexe est séparée sur ses variables, et plusieurs contraintes sont également séparées sur ces variables. Ces belles propriétés facilitent l'utilisation des algorithmes distribués. La méthode distribuée est considérée comme un outil efficace pour traiter le problème d'optimisation à grande échelle souvent rencontré dans les systèmes de communication. Cette méthode permet en effet des approches diviser pour régner, scindant un problème en plusieurs problèmes plus petits. L'une des principales contributions de ce chapitre est de développer un algorithme du gradient projeté distribué à la base du dual très efficace pour résoudre le sous-problème convexe dans le schéma DCA en explorant et en exploitant la structure particulière de ce problème de façon très efficace. Notre schéma DCA distribué calcule itérativement la projection de points à l'intersection d'une boîte et d'un demi-espace pouvant être déterminé très efficacement. Cela améliore très sensiblement la vitesse de l'algorithme dual proposé, réduisant drastiquement le temps d'exécution du schéma DCA distribué. Les expérimentations montrent que notre version DCA distribuée est extrêmement plus rapide que la version distribuée SCA. Le ratio de gain de rapidité peut atteindre 970.

Nous traitons aussi des problèmes SRM dans le chapitre 4, mais dans des réseaux sans fil AF et DF comportant un seul espion et utilisant la technique de relais coopératif et la technique de formation de faisceaux coopératif pour la transmission de données. Les méthodes existantes pour résoudre ces problèmes sont basées sur une technique de relaxation semidéfinie ("semidefinite relaxation" (SDR) en anglais). Nous proposons une nouvelle approche basée sur la programmation DC et DCA pour les résoudre. Nous reformulons tout d'abord ces problèmes non convexes comme programmes DC standards puis développons deux schémas DCA standards pour résoudre le problème d'optimisation dans le scénario AF et un schéma DCA standard pour traiter le problème d'optimisation dans le scénario DF. L'avantage de notre approche est de fournir une solution réalisable alors que les méthodes SDR se contentent d'une relaxation de la solution de part l'élimination des contraintes de rang 1. Dans les méthodes SDR, il faut utiliser des techniques de randomisation sur la solution obtenue pour trouver une solution réalisable ; il s'agit seulement d'une heuristique. A contrario, la convergence de DCA est garantie par la théorie rigoureuse et complète de la programmation DC et DCA. De plus, nous exploitons la structure particulière des problèmes pour fournir une décomposition DC efficace ; elle produit des sous-problèmes convexes qui peuvent être explicitement résolus. Généralement, avec la décomposition DC proposée, le schéma DCA correspondant doit calculer itérativement la projection de points sur une boule euclidienne ou l'intersection de telles boules. Ceci peut être explicitement déterminé. Rechercher une bonne décomposition DC qui résulte en sous problèmes convexes à résolution explicite est hautement recommandé en programmation DC et DCA. Il apporte de bons effets sur la vitesse de convergence du DCA ainsi

que sur les propriétés de la solution trouvée. Les résultats d'expérimentations montrent que notre approche outrepasse SDR à la fois sur le temps de calcul et l'optimisation du taux de secret.

Le chapitre 5 étend le chapitre 4 à des situations où il y a maintenant plusieurs espions dans le réseau de communication. Les problèmes SRM dans ce chapitre sont considérés dans deux scénarios AF et CJ, respectivement. Ils sont sous forme min-max et plus complexes que dans le chapitre précédent. Les méthodes existantes de résolution utilisent les approches SDR. Nous appliquons une extension de la programmation DC et DCA à ces problèmes. Plus spécifiquement, nous reformulons les problèmes SRM originaux, qui sont non différentiables et non convexes comme les programmes DC généraux avec des contraintes DC. Nous développons les schémas DCA généraux pour ces problèmes et prouvons leur convergence. De plus, les résultats expérimentaux montrent que les taus de secret obtenus par les algorithmes DCA sont considérablement meilleurs que ceux obtenus par les méthodes classiques. Pour un réseau AF à espace null, DCA peut donner la solution globale malgré une approche locale.

Organisation de la Thèse

La thèse est composée de six chapitres. Le chapitre 1 présente les concepts et les résultats fondamentaux en analyse convexe ainsi que la programmation DC et DCA, qui crée la base théorique et algorithmique pour les autres chapitres. Le chapitre 2 concerne la qualité de service dans les systèmes de communication tandis que les trois chapitres suivants (Chapitre 3, 4 et 5) s'intéressent à la sécurité de la couche physique des systèmes de communication. Plus précisément, le chapitre 2 présente la façon d'appliquer la programmation DC et DCA pour traiter du problème de maximisation du rapport signal/bruit minimum sous contraintes des puissances. Les trois chapitres 3, 4 et 5 traitent le problème SRM dans diverses architectures de systèmes de communication en utilisant des différentes techniques coopératives. Le chapitre 3 présente une approche DCA pour résoudre le problème SRM dans un système de communication sans fil point-à-point qui se compose d'un espion et déploie la technique de brouillage coopératif dans la transmission de données. Le chapitre 4 considère le problème SRM dans un réseau sans fil de relais comprenant un espion et en utilisant les relais coopératifs et la technique de formation de faisceau coopérative pour la transmission de données. Le chapitre 5 étend les problèmes du chapitre 4 pour le cas où il y a maintenant plusieurs espions sur le réseau. Pour finir, le chapitre 6 donne les conclusions et les perspectives de ce travail.

Chapter 1

Preliminary

Most nonconvex and nondifferentiable optimization problems encountered in applications are formulated as the following forms ([78]).

(1) $\sup\{f(x) : x \in C\}$, where f and C are convex,

(2) $\inf\{g(x) - h(x) : x \in \mathbb{R}^n\}$, where g, h are convex,

(3) $\inf\{g(x) - h(x) : x \in C, f_1(x) - f_2(x) \leq 0\}$, where g, h, f_1, f_2 and C are convex.

It is realized that Problem (1) can be rewritten in the form of Problem (2) with $g = \chi_C$ and $h = f$ and in reserve Problem (2) can be reformulated as $\inf\{t - h(x) : g(x) - t \leq 0\}$, which is equivalent to the form of Problem (1). Problem (3) can be transformed to the form of Problem (2) by using exact penalty related to the DC constraint $f_1(x) - f_2(x) \leq 0$.

Problem (2) is called a standard DC program whereas Problem (3) is called a general DC program. It is apparent that the third class of nonconvex programs is the most general in DC programming and thus it is more challenging to deal with than standard DC programs.

DC programming and DCA (DC Algorithms) are effective tools for solving both the standard DC program (2) and the general DC programs (3). They were introduced by Pham Dinh Tao in their preliminary form in 1985. The important developments and improvements on both theoretical and computational aspects have been completed since 1993 throughout the joint works of Le Thi Hoai An and Pham Dinh Tao. In what follows, we present some main points of DC programming and DCA, which is a methodology of this dissertation. We first recall some notions from Convex Analysis and Nonsmooth Analysis and then present the core contents of standard DC optimization. The general DC optimization is presented in the last section of this chapter. The materials of this chapter are extracted from [46, 78, 51, 80, 47].

Throughout this section, X denotes the Euclidean space $\mathbb{R}^n$ and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is the set of extended real numbers.

1.1 Fundamental Convex Analysis

This section is dedicated to recall some notions and results in convex analysis and nonsmooth analysis that are related to the dissertation. The readers are referred to the work of Rockafellar [83] and of Stephan Boyd [10] for more details.

A subset C of X is said to be *convex* if $(1-\lambda)x + \lambda y \in C$ whenever $x, y \in C$ and $\lambda \in [0,1]$.

Let f be a function whose values are in $\overline{\mathbb{R}}$ and whose domain is a subset S of X. The set

$$\{(x,t) : x \in S, t \in \mathbb{R}, f(x) \leq t\}$$

is called the *epigraph* of f and denoted by epi$f$.

We define f to be a *convex function* on S if epi$f$ is a convex set in $X \times \mathbb{R}$. This is equivalent to that S is convex and

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y), \quad \forall x, y \in S, \forall \lambda \in [0,1]$$

The function f is *strictly convex* if the inequality above holds strictly whenever x and y are distinct in S and $0 < \lambda < 1$.

The *effective domain* of a convex function f on S, denoted by dom$f$, is the projection on X of the epigraph of f

$$\text{dom}f = \{x : \exists t \in \mathbb{R}, (x,t) \in \text{epi}f\} = \{x \mid f(x) < +\infty\}$$

and it is convex.

The convex function f is called *proper* if dom$f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in S$.

The function f is said to be *lower semi-continuous* at a point x of S if

$$f(x) \leq \liminf_{y \to x} f(y)$$

Denote by $\Gamma_0(X)$ the set of all proper lower semi-continuous convex function on X.

Let $\rho \geq 0$ and C be a convex subset of X. One says that a function $\theta : C \mapsto \mathbb{R} \cup \{+\infty\}$ is ρ–convex if

$$\theta[\lambda x + (1-\lambda)y] \leq \lambda\theta(x) + (1-\lambda)\theta(y) - \frac{\lambda(1-\lambda)}{2}\rho\|x-y\|^2$$

for all $x, y \in C$ and $\lambda \in (0,1)$. It is equivalent to say that $\theta - (\rho/2)\|.\|^2$ is convex on C. The modulus of strong convexity of θ on C, denoted by $\rho(\theta, C)$ or $\rho(\theta)$ if $C = X$, is given by

$$\rho(\theta, C) = \sup\{\rho \geq 0 : \theta - (\rho/2)\|.\|^2 \text{ is convex on } C\}$$

One says that θ is *strongly convex* on C if $\rho(\theta, c) > 0$.

A vector y is said to be a *subgradient* of a convex function f at a point x^0 if

$$f(x) \geq f(x^0) + \langle x - x^0, y\rangle, \qquad \forall x \in X$$

The set of all subgradients of f at x^0 is called the *subdifferential* of f at x^0 and is denoted by $\partial f(x^0)$. If $\partial f(x)$ is not empty, f is said to be *subdifferentiable* at x.

For $\varepsilon > 0$, a vector y is said to be a *ε–subgradient* of a convex function f at a point x^0 if

$$f(x) \geq (f(x^0) - \varepsilon) + \langle x - x^0, y\rangle, \qquad \forall x \in X$$

The set of all ε–subgradients of f at x^0 is called the *ε–subdifferential* of f at x^0 and is denoted by $\partial_\varepsilon f(x^0)$.

We also have notations

$$\text{dom } \partial f = \{x \in X : \partial f(x) \neq \emptyset\} \text{ and range } \partial f(x) = \cup\{\partial f(x) : x \in \text{dom } \partial f\}$$

Proposition 1.1. *Let f be a proper convex function. Then*

1. *$\partial_\varepsilon f(x)$ is a closed convex set, for any $x \in X$ and $\varepsilon \geq 0$.*
2. $\text{ri}(\text{dom} f) \subset \text{dom } \partial f \subset \text{dom} f$
   where $\text{ri}(\text{dom} f)$ *stands for the* relative interior *of* $\text{dom} f$.
3. *If f has a unique subgradient at x, then f is differentiable at x, and $\partial f(x) = \{\nabla f(x)\}$.*
4. *$x_0 \in \text{argmin}\{f(x) : x \in X\}$ if and only if $0 \in \partial f(x_0)$.*

Let C be a nonempty closed subset of $\mathbb{R}^n$. The indicator function $\chi_C(x) = 0$ if $x \in C, +\infty$ otherwise. For a closed subset C of $\mathbb{R}^n$, the normal cone of C, denoted by $N(C, x)$, is given by

$$N(C, x) = \partial\chi_C(x) = \{u \in \mathbb{R}^n : \langle u, y - x\rangle \leq 0 \ \ \forall y \in C\}.$$

The function f is said to be λ-Lipschitz if

$$\|f(x_1) - f(x_2)\| \leq \lambda\|x_1 - x_2\| \ \ \forall x_1, x_2 \in S.$$

The function f is called locally Lipshitz if for every $x \in S$ there exists a neiborhood U_x of x such that the restriction of f to U_x is Lipschitz.

Suppose that f is a locally Lipschitz function at a given $x \in \mathbb{R}^n$. The *Clark direction derivative* and the *Clark subdifferential* of f at x is given by the following formulas, respectively.

$$\begin{aligned} f^\uparrow(x, v) &= \limsup_{(t,y)\to(0^+,x)} \frac{f(y + tv) - f(y)}{t}, \\ \partial^\uparrow f(x) &= \left\{x^* \in \mathbb{R}^n : \langle x^*, v\rangle \leq f^\uparrow(x, v) \ \ \forall v \in \mathbb{R}^n\right\}. \end{aligned}$$

If f is continuously differentiable at x then $\partial^{\uparrow} f(x) = \nabla f(x)$. When f is a convex function, then $\partial^{\uparrow} f(x)$ coincides with the subdifferential $\partial f(x)$.

Conjugates of convex functions

The *conjugate* of a function $f : X \mapsto \overline{\mathbb{R}}$ is the function $f^* : X \mapsto \overline{\mathbb{R}}$ defined by

$$f^*(y) = \sup_{x \in X} \{\langle x, y\rangle - f(x)\}$$

Proposition 1.2. *Let $f \in \Gamma_0(X)$. Then we have*

1. *$f^* \in \Gamma_0(X)$ and $f^{**} = f$.*
2. *$f(x) + f^*(y) \geq \langle x, y\rangle$, for any $x, y \in X$.*
   Equality holds if and only if $y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y)$.
3. *$y \in \partial_\varepsilon f(x) \Longleftrightarrow x \in \partial_\varepsilon f^*(y) \Longleftrightarrow f(x) + f^*(y) \leq \langle x, y\rangle + \varepsilon$, for all $\varepsilon > 0$.*

Difference of convex (DC) functions

A function f is called DC function on X if it has the form

$$f(x) = g(x) - h(x), \quad x \in X$$

where g and h belong to $\Gamma_0(X)$. One says that $g - h$ is a *DC decomposition* of f and g, h are its *DC components*. If g and h are in addition finite on all of X then one says that $f = g - h$ is a finite DC function on X. The set of DC functions (resp. finite DC functions) on X is denoted by $\mathcal{DC}(X)$ (resp. $\mathcal{DC}_f(X)$).

Remark 1.1. *Given a DC function f with a DC decomposition $f = g - h$. Then for every $\theta \in \Gamma_0(X)$ finite on the whole X, $f = (g + \theta) - (h + \theta)$ is another DC decomposition of f. Thus, a DC function f has infinitely many DC decompositions.*

1.2 DC Programming and DCA

In this section, we briefly point out the main results in both the standard and general DC optimization, which create the theoretical bases for our methodology. First of all, we show some classic results in the standard DC optimization, which were completely presented in the works of Hoai An LE THI and Tao PHAM DINH such as [46], [78], [51], [80].

1.2.1 Standard DC optimization

1.2.1.1 Standard DC program

In the sequel, we use the convention $+\infty - (+\infty) = +\infty$.

For $g, h \in \Gamma_0(X)$, a standard *DC program* is that of the form

$$(P) \qquad \alpha = \inf\{f(x) = g(x) - h(x) : x \in X\}$$

and its dual counterpart

$$(D) \qquad \alpha^* = \inf\{h^*(y) - g^*(y) : y \in X\}$$

There is a perfect symmetry between primal and dual programs (P) and (D): the dual program to (D) is exactly (P), moreover, $\alpha = \alpha^*$.

Remark 1.2. *Let C be a nonempty closed convex set. Then, the constrained problem*

$$\inf\{f(x) = g(x) - h(x) : x \in C\}$$

can be transformed into an unconstrained DC program by using the indicator function χ_C, i.e.,

$$\inf\{f(x) = \phi(x) - h(x) : x \in X\}$$

where $\phi := g + \chi_C$ is in $\Gamma_0(X)$.

We will always keep the following assumption that is deduced from the finiteness of α

$$\operatorname{dom} g \subset \operatorname{dom} h \quad \text{and} \quad \operatorname{dom} h^* \subset \operatorname{dom} g^*. \tag{1.1}$$

Optimality conditions for standard DC optimization

A point x^* is said to be a *local minimizer* of $g-h$ if $x^* \in \operatorname{dom} g \cap \operatorname{dom} h$ (so, $(g-h)(x^*)$ is finite) and there is a neighborhood U of x^* such that

$$g(x) - h(x) \geq g(x^*) - h(x^*), \qquad \forall x \in U. \tag{1.2}$$

A point x^* is said to be a *critical point* of $g-h$ if it verifies the generalized Kuhn–Tucker condition

$$\partial g(x^*) \cap \partial h(x^*) \neq \emptyset \tag{1.3}$$

Let $\mathcal{P}$ and $\mathcal{D}$ denote the solution sets of problems (P) and (D) respectively, and let

$$\mathcal{P}_\ell = \{x^* \in X : \partial h(x^*) \subset \partial g(x^*)\}, \quad \mathcal{D}_\ell = \{y^* \in X : \partial g^*(y^*) \subset \partial h^*(y^*)\}$$

Below, we present some fundamental results on DC programming [78].

Theorem 1.1. i) *Global optimality condition: $x \in \mathcal{P}$ if and only if $\partial_\varepsilon h(x) \subset \partial_\varepsilon g(x),\ \forall \varepsilon > 0$.*

ii) *Transportation of global minimizers: $\cup\{\partial h(x) : x \in \mathcal{P}\} \subset \mathcal{D} \subset \operatorname{dom} h^*$.*
The first inclusion becomes equality if g^* is subdifferentiable in $\mathcal{D}$. In this case $\mathcal{D} \subset (\operatorname{dom} \partial g^* \cap \operatorname{dom} \partial h^*)$.

iii) *Necessary local optimality: if x^* is a local minimizer of $g - h$, then $x^* \in \mathcal{P}_\ell$.*

iv) *Sufficient local optimality: Let x^* be a critical point of $g-h$ and $y^* \in \partial g(x^*) \cap \partial h(x^*)$. Let U be a neighborhood of x^* such that $(U \cap \operatorname{dom} g) \subset \operatorname{dom} \partial h$. If for any $x \in U \cap \operatorname{dom} g$, there is $y \in \partial h(x)$ such that $h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*)$, then x^* is a local minimizer of $g-h$. More precisely,*

$$g(x) - h(x) \geq g(x^*) - h(x^*), \qquad \forall x \in U \cap \operatorname{dom} g$$

iv) *Transportation of local minimizers: Let $x^* \in \operatorname{dom} \partial h$ be a local minimizer of $g-h$. Let $y^* \in \partial h(x^*)$ and a neighborhood U of x^* such that $g(x) - h(x) \geq g(x^*) - h(x^*),\ \forall x \in U \cap \operatorname{dom} g$. If*

$$y^* \in \operatorname{int}(\operatorname{dom} g^*) \ \text{ and } \ \partial g^*(y^*) \subset U$$

then y^* is a local minimizer of $h^* - g^*$.

Remark 1.3. *a) By the symmetry of the DC duality, these results have their corresponding dual part. For example, if y is a local minimizer of $h^* - g^*$, then $y \in \mathcal{D}_\ell$.*

b) The properties ii), iv) and their dual parts indicate that there is no gap between the problems (P) and (D). They show that globally/locally solving the primal problem (P) implies globally/locally solving the dual problem (D) and vice–versa. Thus, it is useful if one of them is easier to solve than the other.

c) The necessary local optimality condition $\partial h^*(x^*) \subset \partial g^*(x^*)$ is also sufficient for many important classes of programs, for example [51], if h is polyhedral convex, or when f is locally convex at x^*, i.e. there exists a convex neighborhood U of x^* such that f is finite and convex on U. We know that a polyhedral convex function is almost everywhere differentiable, that is it is differentiable everywhere except on a set of measure zero. Thus, if h is a polyhedral convex function, then a critical point of $g-h$ is almost always a local solution to (P).

d) If f is actually convex on X, we call (P) a "false" DC program. *In addition, if $\operatorname{ri}(\operatorname{dom} g) \cap \operatorname{ri}(\operatorname{dom} h) \neq \emptyset$ and $x^0 \in \operatorname{dom} g$ such that g is continuous at x^0, then $0 \in \partial f(x^0) \Leftrightarrow \partial h(x^0) \subset \partial g(x^0)$ [51]. Thus, in this case, the local optimality is also sufficient for the global optimality. Consequently, if in addition h is differentiable, a critical point is also a global solution.*

1.2.1.2 Standard DC Algorithm (DCA)

The DCA consists in the construction of the two sequences $\{x^k\}$ and $\{y^k\}$ (candidates for being primal and dual solutions, respectively) which are easy to calculate and satisfy the following properties:

i) The sequences $(g-h)(x^k)$ and $(h^* - g^*)(y^k)$ are decreasing.

ii) Their corresponding limits x^∞ and y^∞ satisfy the local optimality condition $(x^\infty, y^\infty) \in \mathcal{P}_\ell \times \mathcal{D}_\ell$ or are critical points of $g-h$ and $h^* - g^*$, respectively.

From a given point $x^0 \in \operatorname{dom} g$, DCA generates these sequences by the scheme

$$y^k \in \partial h(x^k) = \arg\min\{h^*(y) - \langle y, x^k \rangle : y \in X\} \tag{1.4a}$$

$$x^{k+1} \in \partial g^*(y^k) = \arg\min\{g(x) - \langle x, y^k\rangle : x \in X\}. \tag{1.4b}$$

The interpretation of the above scheme is simple. At iteration k of DCA, we replace the second component h in the primal DC program by its affine minorant

$$h_k(x) = h(x^k) + \langle x - x^k, y^k\rangle, \tag{1.5}$$

where $y^k \in \partial h(x^k)$. Then the original DC program reduces to the *convex program*

$$(P_k) \qquad \alpha_k = \inf\{f_k(x) := g(x) - h_k(x) : x \in X\}$$

that is equivalent to (1.4a). It is easy to see that f_k is a majorant of f at x^k. Similarly, by replacing g^* with its affine minorant

$$g_k^*(y) = g^*(y^{k-1}) + \langle y - y^{k-1}, x^k\rangle, \tag{1.6}$$

where $x^k \in \partial g^*(y^{k-1})$, we lead to the convex problem

$$(D_k) \qquad \inf\{h^*(y) - g_k^*(y) : y \in X\}$$

whose solution set is $\partial h(x^k)$.

Remark 1.4. *a) Finding y^k, x^{k+1} by the scheme 1.4 is equivalent to solving the problems (D_k) and (P_k). Thus, DCA works by reducing a DC program to a sequence of convex programs that can be solved efficiently.*

b) In practice, the calculation of the subgradient of the function h at a point x is usually easy if we know its explicit expression. But, the explicit expression of the conjugate of a given function g is unknown, so calculating x^{k+1} is done by solving the convex problem (P_k).

c) DCA is constructed from DC convex components g and h and their conjugates but not from the DC function f itself, while a DC function has finitely many DC decompositions. Thus, it is useful to find a suitable DC decomposition since it may have crucial impacts on the efficiency of DCA.

Well definiteness of DCA

DCA is well defined if one can construct two sequences $\{x^k\}$ and $\{y^k\}$ as above from an arbitrary initial point x^0. The following Lemma is the necessary and sufficient condition for this property

Lemma 1.1 ([78])**.** *The sequences $\{x^k\}$ and $\{y^k\}$ in DCA are well defined if and only if*

$$\operatorname{dom}\partial g \subset \operatorname{dom}\partial h \quad \text{and} \quad \operatorname{dom}\partial h^* \subset \operatorname{dom}\partial g^*$$

Since for $\varphi \in \Gamma_0(X)$ we have $\operatorname{ri}(\operatorname{dom}\varphi) \subset \operatorname{dom}\partial\varphi \subset \operatorname{dom}\varphi$ (Proposition 1.1). Moreover, we also keep the assumptions $\operatorname{dom} g \subset \operatorname{dom} h$, $\operatorname{dom} h^* \subset \operatorname{dom} g^*$. So, we can say that DCA in general is well defined.

Convergence properties of standard DCA

Complete convergence of standard DCA is given in the following results ([78]).

Theorem 1.2. *Suppose that the sequences $\{x^k\}$ and $\{y^k\}$ are generated by the DCA. Then we have*

i) *The sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing and*

- *$g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ if and only if $\{x^k, x^{k+1}\} \subset \partial g^*(y^k) \cap \partial h^*(y^k)$ and $[\rho(h) + \rho(g)]\|x^{k+1} - x^k\| = 0$.*
- *$h^*(y^{k+1}) - g^*(y^{k+1}) = h^*(y^k) - g^*(y^k)$ if and only if $\{y^k, y^{k+1}\} \subset \partial g(x^k) \cap \partial h(x^k)$ and $[\rho(h^*) + \rho(g^*)]\|y^{k+1} - y^k\| = 0$.*

DCA terminates at the kth iteration if either of the above equalities holds.

ii) *If $\rho(h) + \rho(g) > 0$ (resp. $\rho(h^*) + \rho(g^*) > 0$), then the sequences $\{\|x^{k+1} - x^k\|^2\}$ (resp. $\{\|y^{k+1} - y^k\|^2\}$) converge.*

iii) *If the optimal value α is finite and the sequences $\{x^k\}$ and $\{y^k\}$ are bounded, then every limit point x^∞ (resp. y^∞) of the sequence $\{x^k\}$ (resp. $\{y^k\}$) is critical point of $g - h$ (resp. $h^* - g^*$).*

iv) *DCA has a linear convergence for general DC program.*

1.2.1.3 An useful DC decomposition and the corresponding DCA

As mentioned above, each DC function f has infinitely many DC decompositions which have critical effects on the qualities of DCA such as convergence speed, robustness, efficiency. How to choose a good DC decomposition for a given DC program remains an open question. It requires to exploit the special structure of the considered problems to introduce the DC components g and h such that the sequences $\{x^k\}$ and $\{y^k\}$ can be easily calculated, i.e., their computations are either explicit or inexpensive. Normally, the sequence $\{y^k\}$ is explicitly computed with the use of normal rules for calculating subdifferential of convex functions. Thus, the focus is on the computation of the sequence $\{x^k\}$. In what follows, we present a way to construct a DC decomposition which leads to explicitly-solved convex subproblems in some special cases of the feasible set.

Consider a nonconvex program $\min\{f(x) : x \in C\}$, where C is a convex set. Assume that there exists a nonnegative ρ such that the function $\frac{1}{2}\rho\|x\|^2 - f(x)$ is convex. The function f now can be decomposed into the difference of two convex fucntions $g(x) = \frac{1}{2}\rho\|x\|^2$ and $h(x) = \frac{1}{2}\rho\|x\|^2 - f(x)$. The DCA applied to this problem can be described as follows.

Initialization: Let $x^0 \in \mathbb{R}^n$. Set $k := 0$.
Repeat
- Compute $y^k \in \partial h(x^k)$.
- Compute x^{k+1} by solving the convex program

$$\min\left\{\frac{1}{2}\rho\|x\|^2 - \langle x, y^k\rangle : x \in C\right\}.$$

- Set $k := k + 1$.
Until convergence of $\{x^k\}$.

In the above DCA scheme, solving the convex program is actually finding a projection of $\frac{y^k}{\rho}$ onto C. If C is a box or a ball, the computation of this projection is explicit. Therefore, the use of this DCA scheme is highly recommended in these special cases. However, there is still a crucial question, that is how to determine ρ such that the function $h(x) = \frac{1}{2}\rho\|x\|^2 - f(x)$ is convex. There is not a common answer for all cases since it depends on the properties of the function f. For a special class of functions which is smooth and have Lipschitz continuous gradient, ρ is nothing but the Lipschitz constant of the function's gradient ([53]).

To terminate Section 1.2 , we show two approaches to deal with the general DC programs in general DC optimization. This content is drawn from [80], [47].

1.2.2 General DC optimization

1.2.2.1 General DC program

A general DC program is of the form

$$\begin{array}{ll} \min\limits_{x} & f_0(x) \\ \text{s.t} & f_i(x) \leq 0 \ \forall i = 1, ..., m, \\ & x \in C, \end{array} \tag{1.7}$$

where $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set; $f_i : \mathbb{R}^n \to \mathbb{R}(i = 0, 1, ..., m)$ are DC functions.

Two approaches for general DC programs were proposed in [80], [47] to overcome the difficulty caused by the nonconvexity of the constraints. Both approaches are built on the main idea of the philosophy of DC programming and DCA, that is approximating (1.7) by a sequence of convex programs. The former was based on penalty techniques in DC programming while the latter was relied on the convex inner approximation method.

Denote F as the feasible set of (1.7). A point $x^* \in F$ is a Karush-Kuhn-Tucker (KKT) point for the problem (1.7) if there exist nonnegative scalars $\lambda_i, i = 1, ..., m$ such that

$$\begin{cases} 0 \in \partial f_0(x^*) + \sum_{i=1}^{m} \lambda_i \partial f_i(x^*) + N(C, x), \\ \lambda_i f_i(x^*) = 0, \quad i = 1, \ldots, m. \end{cases} \tag{1.8}$$

Denote

$$p(x) = \max\{f_1(x), f_2(x), ..., f_m(x)\},$$
$$I(x) = \{i \in \{1, ..., m\} : f_i(x) = p(x)\}; p^+(x) = \max\{p(x), 0\}.$$

It is said that the *extended Mangasarian-Fromowitz constraint qualification (EMFCQ)* is satisfied at $x^* \in F$ with $I(x^*) \neq \emptyset$ if

(EMFCQ) There is a vector $d \in \text{cone}(C - \{x^*\})$ (the cone hull of $C - \{x^*\}$) such that $f_i^{\uparrow}(x^*, d) < 0 \ \ \forall i \in I(x^*)$.

When $f_i's$ are continuously differentiable, then $f_i^{\uparrow}(x^*, d) = \langle \nabla f_i(x^*), d\rangle$. Therefore (EMFCQ) becomes the well-known Mangasarian-Fromowitz constraint qualification.

It was shown in [71, 70] that if the (extended) Mangasarian-Fromowitz constraint qualification is satisfied at a local minimizer x^* of (1.7) then the KKT first order necessary conditions (1.8) hold.

1.2.2.2 General DC Algorithm using l_∞-penalty function with updated parameter: DCA1

Consider the following penalty problems

$$\begin{aligned} \min_x \quad & \phi_k(x) = f_0(x) + \beta_k p^+(x) \\ \text{s.t} \quad & x \in C, \end{aligned} \tag{1.9}$$

where β_k are penalty parameters. Since $f_i(x), i = 1, \ldots, m$ are DC functions, so is p^+. Suppose that f_0 and p^+ are decomposed into the difference of two convex functions as below

$$f_0(x) = g_0(x) - h_0(x), p^+(x) = p_1(x) - p_2(x)$$

where g_0, h_0, p_1, p_2 are convex functions defined on the whole space. It results in the following DC decomposition for ϕ_k.

$$\phi_k(x) = g_k(x) - h_k(x), x \in \mathbb{R}^n,$$

where

$$g_k(x) = g_0(x) + \beta_k p_1(x), \quad h_k(x) = h_0(x) + \beta_k p_2(x).$$

The algorithm using l_∞-penalty function with updated parameter is described as follows.

DCA1

Initialization: Take an initial point $x^1 \in C; \delta > 0$; an inital penalty parameter $\beta_1 > 0$ and set $k := 1$.

1. Compute $y^k \in \partial h_k(x^k)$.
2. Compute x^{k+1} by solving the convex program
$$\min \left\{ g_k(x) - \langle x, y^k \rangle : x \in C \right\}.$$
3. Stopping test.
Stop if $x^{k+1} = x^k$ and $p(x^k) \leq 0$.
4. Penalty parameter update.
Compute $r_k = \min \left\{ p(x^k), p(x^{k+1}) \right\}$ and set
$$\beta_{k+1} = \begin{cases} \beta_k & \text{if either } \beta_k \geq \|x^{k+1} - x^k\|^{-1} \text{ or } r_k \leq 0, \\ \beta_k + \delta & \text{if } \beta_k < \|x^{k+1} - x^k\|^{-1} \text{ and } r_k > 0 \end{cases}$$

5. Set $k := k + 1$ and go to Step 1.

The global convergence of DCA1 is given by the theorem below. To attain the assertions in the theorem, three necessary assumptions below are assumed

Assumption 1. *$f_i', i = 0, ..., m$ are locally Lipschitz functions at every point of C.*

Assumption 2. *Either g_k or h_k is differentiable on C, and $\rho(g_0) + \rho(h_0) + \rho(p_1) + \rho(p_2) > 0$.*

Assumption 3. *The (extended) Mangasarian-Fromowitz constraint qualification (EMFCQ) is satisfied at any $x \in \mathbb{R}^n$ with $p(x) \geq 0$.*

Theorem 1.3. *Suppose that $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $f_i, i = 1, \ldots, m$ are DC functions on C. Suppose further that Assumptions 1-3 are verified. Let $\delta > 0, \beta_1 > 0$ be given. Let $\{x^k\}$ be a sequence generated by DCA1. Then DCA1 either stops, after finitely many iterations, at a KKT point x^k for problem (1.7) or generates an infinite sequence $\{x^k\}$ of iterates such that $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ and every limit point x^∞ of the sequence $\{x^k\}$ is a KKT point of problem (1.7).*

The detail of proof of this theorem is shown in [47].

1.2.2.3 Genearal DC Algorithm using slack variables with updated relaxation parameter: DCA2

Since $f_i (i = 0, ..., m)$ are DC functions, they can be decomposed into the difference of two convex functions $f_i(x) = g_i(x) - h_i(x),\ x \in \mathbb{R}^n, i = 0, ..., m$. By linearizing the concave part of DC decompositions of all DC objective function and DC constraints, it raises sequential convex subproblems of the following form:

$$\begin{aligned} \min_x \quad & g_0(x) - \langle y_0^k, x\rangle \\ \text{s.t} \quad & g_i(x) - h_i(x^k) - \langle y_i^k, x - x^k\rangle \ \leq 0 \ \ \forall i = 1, ..., m, \\ & x \in C, \end{aligned} \tag{1.10}$$

where $x^k \in \mathbb{R}^n$ is a point at the current iteration , $y_i^k \in \partial h_i(x^k)\ \forall\ i = 0, ..., m$.
This linearization introduces an inner convex approximation of the feasible set of (1.7). However, it may lead to infeasibility of convex subproblems (1.10). The relaxation technique was proposed to confront this difficulty. Instead of (1.10), the relaxed subproblem below is considered:

$$\begin{aligned} \min_x \quad & g_0(x) - \langle y_0^k, x\rangle + \beta_k t \\ \text{s.t} \quad & g_i(x) - h_i(x^k) - \langle y_i^k, x - x^k\rangle \ \leq t \ \ \forall i = 1, ..., m, \\ & x \in C, t \geq 0, \end{aligned} \tag{1.11}$$

where β_k is a penalty parameter. It is easy to realize that the relaxed subproblem (1.11) is always feasible. Furthermore, the Slater constraint qualification is satisfied for the

constraints of (1.11), thus the Karush-Kuhn-Tucker (KKT) optimality condition holds for some solutions (x^{k+1}, t^{k+1}). Thus, there exits some $\lambda_i^{k+1} \in \mathbb{R}, i = 1, .., m$ and $\mu^{k+1} \in \mathbb{R}$ such that

- $0 \in \partial g_0(x^{k+1}) - y_0^k + \sum_{i=0}^{m} \lambda_i^{k+1}(\partial g_i(x^{k+1}) - y_i^k) + N(C, x^{k+1})$,
- $\beta_k - \sum_{i=1}^{m} \lambda_i^{k+1} - \mu^{k+1} = 0$,
- $g_i(x^{k+1}) - h_i(x^k) - \langle y_i^k, x^{k+1} - x^k \rangle \leq t^{k+1}, \lambda_i^{k+1} \geq 0 \ \forall i = 1, ..., m, x^{k+1} \in C$,
- $\lambda_i^{k+1}(g_i(x^{k+1}) - h_i(x^k) - \langle y_i^k, x^{k+1} - x^k \rangle - t^{k+1}) = 0, \forall i = 1, ..., m$,
- $t^{k+1} \geq 0, \mu^{k+1} \geq 0, t^{k+1}\mu^{k+1} = 0$.

The general DC Algorithm using slack variables with updated relaxation parameter (DCA2) for general DC program (1.7) is proposed as follows:

DCA2

Initialization: Take initial point $x^0; \delta_1, \delta_2 > 0$, an initial penalty parameter $\beta_1 > 0$. Set $k \longleftarrow 1$.

1. Compute $y_i^k \in \partial h_i(x^k), \quad i = 0, .., m$.
2. Compute (x^{k+1}, t^{k+1}) as the solution of (1.11), and the associated Lagrange multipliers $(\lambda^{k+1}, \mu^{k+1})$
3. Stopping test.
   Stop if $x^{k+1} = x^k$ and $t^{k+1} = 0$.
4. Penalty parameter update.
   compute $r_k = \min\{\|x^{k+1} - x^k\|^{-1}, \|\lambda^{k+1}\|_1 + \delta_1\}$
   and set $\beta_{k+1} = \begin{cases} \beta_k & \text{if } \beta_k \geq r_k, \\ \beta_k + \delta_2 & \text{if } \beta_k < r_k. \end{cases}$
5. Set $k := k + 1$ and go to Step 1.

The proof of global convergence of the above algorithm is shown in the theorem below.

Theorem 1.4. *Suppose that $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $f_i, i = 1, \ldots, m$ are DC functions on C such that assumptions 1 and 3 are verified. Suppose further that for each $i = 0, ..., m$ either g_i or h_i is differentiable on C and that*

$$\rho = \rho(g_0) + \rho(h_0) + \min\{\rho(g_i) : i = 1, \ldots, m\} > 0.$$

Let $\delta_1, \delta_2 > 0, \beta_1 > 0$ be given. Let $\{x^k\}$ be a sequence generated by DCA2. Then DCA2 either stops, after finitely many iterations, at a KKT point x^k for problem (1.7) or generates an infinite sequence $\{x^k\}$ of iterates such that $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ and every limit point x^∞ of the sequence $\{x^k\}$ is a KKT point of problem (1.7).

The proof of this theorem is presented in [47].

Chapter 2

DC Programming and DCA for Rank-Two Transmit Beamforming and Power Allocation in Multicasting Relay Network

Abstract: *This chapter concerns a single-group multicasting relay network consisting of multiple amplify-and-forward relays forwarding signal from a single source to multiple destinations. The source, relays and destinations are equipped by single antenna. In this scenario, we deal with the problem of maximizing the minimum Quality of Service (QoS) assessed by the Signal to Noise Ratio (SNR) of the destinations subject to power constraints. This problem is nonsmooth and nonconvex thus hard to deal with. We first reformulate it as a general DC (Difference of Convex functions) program with DC constraints based on a novel DC decomposition and then develop an efficient general DCA (DC Algorithms) scheme for solving it. Numerical experiments are carried out on several simulated datasets and they show that the performance of the proposed DCA scheme is better than that of the existing ones.*

2.1 Introduction and Related Works

To meet the higher and higher demand of wireless network users for quality of service, the next-generation wireless networks should develop in the direction of offering more new techniques with the aim of achieving a better data rate compared to the currently deployed networks. Nevertheless, the difficulty of obtaining a good data rate is often caused by the interference. Therefore, it is necessary to develop techniques mitigating interference for a better operation of future networks. Some of the techniques exploiting the spatial domain often used recently can be listed here such as multiple-input multiple-output (MIMO) antenna settings, cooperative relays, and beamforming (BF).

Among them, cooperative relaying has been deployed in a lot of recent works because this technique plays an important role in improving three critical parameters of wireless

networks. The path loss attenuation and shadowing are reduced with the assistance of the cooperative relays. Moreover, cooperative relaying brings independent propagation paths, thus leads to an alleviation of multipath fading. In short, through cooperative relaying, coverage is extended, reliability is risen, and diversity can be attained ([103]).

Beamforming is actually a signal processing technique in which BF matrices are used at transmitters and receivers for directional signal transmission and reception. The entries of BF matrices are chosen in such a way to satisfy a particular objective, such as mean square error (MSE) or signal-to-noise ratio (SNR). Beamforming technique is regarded as an efficient approach to receive, transmit, or relay signal in the medium having existence of noise and interference ([22]). It has ever been applied to various relay network architectures, ranging from single-user networks ([11],[29]) to point-to-point networks ([21],[88]) and then to multi-user multicasting networks ([9],[87]). Beamforming technique can be classified into two categories: distributed beamforming through using non-connected relays ([9],[87]) and centralized beamforming through using a connected antenna array ([41],[43]).

A popular technique in transmit beamforming, called rank-one transmit beamforming, is to use a single weight vector at the transmitter to send signal to the intended destination ([9],[123],[120]). Besides, recent works introduce another technique, called rank-two transmit beamforming, in which two weight vectors are deployed to transmit two data symbols at the same time ([112],[113]). The simulation results on these works indicate that with the rank-two beamforming scheme, the performance of system is considerably improved compared with the use of the rank-one beamforming one.

To address the beamforming problem, which is in essence a nonconvex quadratically constrained quadratic optimization problem (QCQP), two approaches in direction of approximating the feasible set have been applied ([113],[112],[81]). The first one tends to narrow the feasible set that may make the resulting problem infeasible. To overcome this difficulty, the technique of searching a feasible initial point by introducing a slack variable to the approximated constraints was mentioned in [12]. Another method is the combination of penalty technique and method of introducing a slack variable mentioned above ([47]). The second approach to solve a QCQP is based on an outer approximation which reformulates the QCQP as a semidefinite program (SDP) after eliminating the rank constraint. However, this semidefinite relaxation (SDR) technique makes the feasible set wider, thus it only provides an upper/lower bound for the objective value of the beamforming problem. In the case when the relaxed solution does not belong to the original feasible set, randomization techniques have been deployed to generate feasible points that are in general suboptimal ([113],[90]).

In this chapter, we take account of a distributed rank-two beamforming scheme for a single-group multicasting network including amplify-and-forward (AF) relays. The objective of the model is not only to design the BF vectors to direct signals to the intended destinations as in the aforementioned works, but also to find the scaling factors to control power between different time slots and between the source and the relays to maximize the minimum SNR subject to power constraints. It should be noted that the max-min fairness used in this model is a well-known criterion and widely

used in many works such as [11],[91],[15]. The joint determination of BF vectors and scaling factors in this model makes it more intractable compared to those without considering scaling factors. This model was introduced in [86] and solved based on both SDR technique and concave-convex procedure (CCCP) that is in fact a DCA based algorithm. We also address this max-min fairness optimization problem via DCA but with a novel DC decomposition. The initial results were reported in [54]. However the efficiency of this DCA in terms of runtime has not been achieved yet. Therefore, we introduce in this version a DC decomposition different from that in [54]. This is meaningful since in DC programming and DCA, the qualities of DCA such as convergence speed, efficiency, property of computed solution depend on the corresponding DC formulation, thus it is worth exploring various DC decompositions for a given DC program to find a good one. Moreover, by treating variables jointly in the proposed DCA scheme, we avoid using a two-dimension search in the outer level followed by solving a convex SDP in the inner level as in the SDR based method, which causes a huge computational burden to this algorithm.

Our contributions are to reformulate the max-min fairness optimization problem as a general DC program by a new DC decomposition and develop an efficient general DCA scheme for dealing with the given problem. The numerical results show that our general DCA outperforms the existing algorithms in terms of both quality of service and runtime. In addition, the convergence property of the proposed general DCA scheme is rigorously proved. The general DCA scheme is regarded as a generalization of the standard DCA. While the standard DCA has been completely studied and successfully applied to solve many optimization problems in various areas for a long time, the general DCA scheme has just been investigated and applied recently. The convergence of a generic general DCA scheme was shown in [47] under some strong assumptions, however not all DC programs satisfy them. Therefore, the proof of convergence for a general DCA scheme designed for a specific DC program is not always straightforward from the convergence of the generic general DCA. In our situation, the DC components are not strongly convex in all variables, which violates one of those assumptions. Nonetheless, we exploit the partially strongly convex property of DC components to obtain a similar result as in Lemma 1 of [47], which is an important step to show the convergence property. Besides, in our general DCA, we do not need to introduce a slack variable while still ensuring the feasibility of convex subproblems. Meanwhile, in the generic general DCA, a slack variable was introduced and penalized to the objective function to avoid the case of infeasibility for subproblems and the update of penalty coefficients was required. As a consequence, our proof of convergence is different from that of the generic general DCA.

It should be noted that the convergence of the existing DCA although was shown in [86] based on Theorem 10 of [92], we realize a loose argument in the proof of this theorem. More specifically, to prove Theorem 10, it requires to verify three conditions of Zangwill's convergence theorem ([122]) including the condition of closeness of the algorithm. For the case of standard DC program, this condition was shown by Lemma 6 ([92]). Nevertheless, for the case of general DC program with DC constraints, there are not sufficient grounds to infer the closeness of the algorithm from this lemma. It is because the feasible sets of subproblems in the general DCA scheme are different across

iterations while this lemma requires the feasible sets of subproblems unchanged after each iteration. Therefore the convergence of the existing DCA need to be reconsidered.

2.2 Transmit Beamforming and Power Allocation in Multicasting Relay Networks

2.2.1 Rank-two beamforming model

In this section, we briefly restate the problem formulated in [86]. Consider a wireless system comprising a single source, R relays and M destinations. The source, each relay and each destination are equipped by single antenna. In this model, two data symbols are simultaneously processed in a four time slot scheme. All channels in the network are supposed to be frequency flat and constant over the considered four time slots.
In the first and second time slot, the source transmits the data symbols s_1 and s_2^* (conjugate of s_2) to the relays and the destinations respectively. Both symbols are multiplied by the same coefficient $p_1 \in \mathbb{R}$ before being sending. The relays, in the first and second time slot, receive the following signals

$$\mathbf{r}_1 = \mathbf{f}p_1 s_1 + \mathbf{n}_{R,1},\ \mathbf{r}_2 = \mathbf{f}p_1 s_2^* + \mathbf{n}_{R,2}, \tag{2.1}$$

where $\mathbf{n}_{R,1} \in \mathbb{C}^R$ and $\mathbf{n}_{R,2} \in \mathbb{C}^R$ are the relay noise vectors in the first and the second time slot, respectively, and $\mathbf{f} \in \mathbb{C}^R$ is the vector of the channel coefficients between the source and the relays. The signal $d_{m,1}$ and $d_{m,2}$ received by the mth destination in the first and second time slot are respectively computed by

$$d_{m,1} = h_m p_1 s_1 + n_{D,m,1},\ d_{m,2} = h_m p_1 s_2^* + n_{D,m,2}, \tag{2.2}$$

where h_m is the channel coefficient from the source to the mth destination and $n_{D,m,1}, n_{D,m,2}$ are the noise at the mth destination in the first and second time slot, respectively. The signals transmitted by relays in the third and fourth time slot are given by

$$\mathbf{t}_3 = \mathbf{W}_1\mathbf{r}_1 + \mathbf{W}_2\mathbf{r}_2^*,\ \mathbf{t}_4 = -\mathbf{W}_2\mathbf{r}_1^* + \mathbf{W}_1\mathbf{r}_2, \tag{2.3}$$

where $\mathbf{W}_1 \triangleq \text{diag}(\mathbf{w}_1^\dagger), \mathbf{W}_2 \triangleq \text{diag}(\mathbf{w}_2^\dagger)$, and $\mathbf{w}_1 = [w_{1,1}, ..., w_{R,1}]^T$, $\mathbf{w}_2 = [w_{1,2}, ..., w_{R,2}]^T$ are the complex $R \times 1$ beamforming vectors. At the same time, the source sends the signals $p_3 s_1 + p_4 s_2$ and $-p_4 s_1^* + p_3 s_2^*$, respectively to the destinations, where p_3, p_4 are complex weights. The signals received by the mth destination in the third and fourth time slot are calculated by

$$\begin{aligned} d_{m,3} &= \mathbf{g}_m^T\mathbf{t}_3 + h_m(p_3 s_1 + p_4 s_2) + n_{D,m,3}, \\ d_{m,4} &= \mathbf{g}_m^T\mathbf{t}_4 + h_m(-p_4 s_1^* + p_3 s_2^*) + n_{D,m,4}, \end{aligned} \tag{2.4}$$

where $\mathbf{g}_m \in \mathbb{C}^R$ is the vector of the complex channel coefficients between the relays and the mth destination and $n_{D,m,3}$, $n_{D,m,4}$ are the receiver noise at the mth destination in

the third and fourth time slot, respectively. It is assumed that the noise processes in the network are spatially and temporally independent and complex Gaussian distributed. The noise power at the destinations equals to $E\{|n_{D,m,q}|^2\} = \sigma_D^2,\ q \in \{1,2,3,4\}$, and the noise at the relays has distribution $\mathbf{n}_{R,1} \sim \mathcal{CN}(0_R, \sigma_R^2\mathbf{I}_R), \mathbf{n}_{R,2} \sim \mathcal{CN}(0_R, \sigma_R^2\mathbf{I}_R)$. Let us denote the vector of the received signals at the mth destination by $\mathbf{d}_m \triangleq \left[d_{m,1}, d_{m,2}^*, d_{m,3}, d_{m,4}^*, \right]^T$, the vector of the noise at the mth destination by $\mathbf{n}_m$, the equivalent channel matrix by $\mathbf{Z}_m$ and using the equations (2.1),(2.3) and (2.4), the received signals of the four time slots can be jointly written as

$$\mathbf{d}_m = \mathbf{Z}_m\mathbf{s} + \mathbf{n}_m \tag{2.5}$$

where

$$\mathbf{s} = [s_1, s_2]^T, \mathbf{n}_m \triangleq \begin{bmatrix} n_{D,m,1} \\ n_{D,m,2}^* \\ \mathbf{w}_1^\dagger\mathbf{G}_m\mathbf{n}_{R,1} + \mathbf{w}_2^\dagger\mathbf{G}_m\mathbf{n}_{R,2}^* + n_{D,m,3} \\ -\mathbf{w}_2^T\mathbf{G}_m^\dagger\mathbf{n}_{R,1} + \mathbf{w}_1^T\mathbf{G}_m^\dagger\mathbf{n}_{R,2}^* + n_{D,m,4}^* \end{bmatrix}, \mathbf{Z}_m \triangleq \begin{bmatrix} p_1h_m & 0 \\ 0 & (p_1h_m)^* \\ z_{m,1} & z_{m,2} \\ -z_{m,2}^* & z_{m,1}^* \end{bmatrix} \tag{2.6}$$

with $z_{m,1} \triangleq p_1\mathbf{w}_1^\dagger\mathbf{G}_m\mathbf{f} + p_3h_m,\ z_{m,2} \triangleq p_1\mathbf{w}_2^\dagger\mathbf{G}_m\mathbf{f}^* + p_4h_m,\ \mathbf{G}_m \triangleq \text{diag}(\mathbf{g}_m)$.
It can be easily verified that $E(\mathbf{n}_m\mathbf{n}_m^\dagger) = \text{blkdiag}(\left[\sigma_D^2\mathbf{I}_2, \sigma_{m,34}^2\mathbf{I}_2\right])$, where

$$\sigma_{m,34}^2 \triangleq \sigma_R^2(\mathbf{w}_1^\dagger\mathcal{G}_m\mathbf{w}_1 + \mathbf{w}_2^\dagger\mathcal{G}_m\mathbf{w}_2) + \sigma_D^2, \quad \mathcal{G}_m \triangleq \mathbf{G}_m\mathbf{G}_m^\dagger.$$

The signal-to-noise ratio for both data symbols at the mth destination is shown in [86] by the formula below

$$\text{SNR}_m = \frac{p_1^2|h_m|^2}{\sigma_D^2} + \frac{|z_{m,1}|^2 + |z_{m,2}|^2}{\sigma_{m,34}^2}. \tag{2.7}$$

By denoting $p = \frac{1}{p_1^2}$; $\mathbf{w} = [\hat{\mathbf{w}}_1^T, \hat{\mathbf{w}}_2^T]^T$ with $\hat{\mathbf{w}}_1 = [\mathbf{w}_1^T, p_3^*/p_1]^T, \hat{\mathbf{w}}_2 = [\mathbf{w}_2^T, p_4^*/p_1]^T$; $\mathbf{B}_m = \text{blkdiag}([\hat{\mathbf{B}}_m, \hat{\mathbf{B}}_m])$ with $\hat{\mathbf{B}}_m = [\sigma_R^2\mathcal{G}_m, 0]$; $\mathbf{A}_m = \text{blkdiag}([\hat{\mathbf{A}}_{m,1}, \hat{\mathbf{A}}_{m,2}])$, where $\hat{\mathbf{A}}_{m,1} = \mathbf{a}_{m,1}\mathbf{a}_{m,1}^\dagger, \hat{\mathbf{A}}_{m,2} = \mathbf{a}_{m,2}\mathbf{a}_{m,2}^\dagger$ with $\mathbf{a}_{m,1} = \begin{bmatrix} \mathbf{G}_m\mathbf{f} \\ h_m \end{bmatrix}, \mathbf{a}_{m,2} = \begin{bmatrix} \mathbf{G}_m\mathbf{f}^* \\ h_m \end{bmatrix}$ and from the formula of $z_{m,1}, z_{m,2}$, (2.7) can be rewritten in the following form

$$\text{SNR}_m(\mathbf{w}, p) = \frac{\hat{\mathbf{w}}_1^\dagger\hat{\mathbf{A}}_{m,1}\hat{\mathbf{w}}_1 + \hat{\mathbf{w}}_2^\dagger\hat{\mathbf{A}}_{m,2}\hat{\mathbf{w}}_2}{\left(\hat{\mathbf{w}}_1^\dagger\hat{\mathbf{B}}_m\hat{\mathbf{w}}_1 + \hat{\mathbf{w}}_2^\dagger\hat{\mathbf{B}}_m\hat{\mathbf{w}}_2 + \sigma_D^2\right)p} + \frac{|h_m|^2}{\sigma_D^2 p} = \frac{\mathbf{w}^\dagger\mathbf{A}_m\mathbf{w}}{(\mathbf{w}^\dagger\mathbf{B}_m\mathbf{w} + \sigma_D^2)p} + \frac{|h_m|^2}{\sigma_D^2 p}. \tag{2.8}$$

In multicast networks, a trade-off between the transmitted power and the QoS at the intended receivers has to be met. The best trade-off is achieved by solving an optimization problem, where the beamforming weight vectors and power scaling factors are the optimization variables. In this chapter, the objective is to find the beamforming weight vectors as well as power scaling factors to maximize the minimum SNR at the destinations subject to power constraints. The worst SNR is an important limiting value in multicasting application because it determines the common information rate. Maximizing the worst SNR is to ensure fairness among users, avoid the existence of users with very poor SNR. This max-min fairness criterion has been used in many

previous works such as [41], [15], [91]. The max-min fairness optimization problem considered in [86] is of the following form

$$\max_{\mathbf{w},p} \min_{m\in\{1,..,M\}} \mathrm{SNR}_m(\mathbf{w},p) \quad \text{s.t.} \quad (\mathbf{w},p)\in\Omega, \tag{2.9}$$

where Ω is a set of $(\mathbf{w},p)$ satisfying the constraints below

$$\begin{aligned}
&\text{positivity} : p>0\\
&\text{individual relay power} : p_r(\mathbf{w},p)=\mathbf{w}^\dagger\mathbf{D}_r\mathbf{w}/p+\mathbf{w}^\dagger\mathbf{E}_r\mathbf{w}\le p_{r,max}\ \forall r\in\{1,...,R\},\\
&\text{relay sum power} : \sum_{r=1}^{R} p_r(\mathbf{w},p)=\sum_{r=1}^{R}(\mathbf{w}^\dagger\mathbf{D}_r\mathbf{w}/p+\mathbf{w}^\dagger\mathbf{E}_r\mathbf{w})\le P_{R,max},\\
&\text{source power} : P_S(\mathbf{w},p)=2/p+\mathbf{w}^\dagger\mathbf{S}\mathbf{w}/p\le P_{S,max},\\
&\text{total power} : P_T(\mathbf{w},p)=2/p+\mathbf{w}^\dagger\mathbf{S}\mathbf{w}/p+2\sum_{r=1}^{R}(\mathbf{w}^\dagger\mathbf{D}_r\mathbf{w}/p+\mathbf{w}^\dagger\mathbf{E}_r\mathbf{w})\le P_{T,max},
\end{aligned}$$

where $\mathbf{D}_r \triangleq \mathrm{blkdiag}\left(\left[\widehat{\mathbf{D}}_r,\widehat{\mathbf{D}}_r\right]\right)$, $\mathbf{E}_r \triangleq \mathrm{blkdiag}\left(\left[\widehat{\mathbf{E}}_r,\widehat{\mathbf{E}}_r\right]\right)$ and $\mathbf{S} \triangleq \mathrm{blkdiag}\left(\left[\widehat{\mathbf{S}},\widehat{\mathbf{S}}\right]\right)$, in which $\widehat{\mathbf{D}}_r$ is a $(R+1)\times(R+1)$ matrix with all entries equal to zero except (r,r)-entry equals to $|f_r|^2$, $\widehat{\mathbf{E}}_r$ is a $(R+1)\times(R+1)$ matrix having σ_R^2 as its rth diagonal entry and zeros elsewhere, $\widehat{\mathbf{S}}$ is a $(R+1)\times(R+1)$ diagonal matrix with $(R+1,R+1)$-entry equals to 2 and the others equal to zeros.
Note that the quadratic form $\mathbf{w}^\dagger\mathbf{A}\mathbf{w}$ and the fraction of the quadratic form $\mathbf{w}^\dagger\mathbf{A}\mathbf{w}$ and the linear term a are convex provided that $\mathbf{A}$ is a positive semidefinite Hermitian matrix and $a>0$. Therefore the constraint set Ω mentioned above is convex.

2.2.2 Rank-one beamforming model

The Rank-One model is actually a special case of the Rank-Two model in which symbols are forwarded by relays with a single beamformer. It means that the second beamforming vector is assigned by $\mathbf{w}_2=0$. In addition, each symbol is processed in two time slots instead of four time slots as in the Rank-Two model, therefore the complex scaling factors p_3,p_4 are regarded as zero. More specifically, in the first time slot the source transmits the signal to the relays and in the second time slot, these received signals are forwarded to the destination after being multiplied by the beamforming vector.

2.3 Solution Method Based on DC Programming and DCA

2.3.1 The real form of the problem (2.9)

Define $\bar{\mathbf{A}}_m = \begin{bmatrix} \mathrm{Re}(\mathbf{A}_m) & -\mathrm{Im}(\mathbf{A}_m) \\ \mathrm{Im}(\mathbf{A}_m) & \mathrm{Re}(\mathbf{A}_m) \end{bmatrix}, \bar{\mathbf{B}}_m = \begin{bmatrix} \mathrm{Re}(\mathbf{B}_m) & -\mathrm{Im}(\mathbf{B}_m) \\ \mathrm{Im}(\mathbf{B}_m) & \mathrm{Re}(\mathbf{B}_m) \end{bmatrix}, \bar{\mathbf{D}}_r = \begin{bmatrix} \mathrm{Re}(\mathbf{D}_r) & -\mathrm{Im}(\mathbf{D}_r) \\ \mathrm{Im}(\mathbf{D}_r) & \mathrm{Re}(\mathbf{D}_r) \end{bmatrix}, \bar{\mathbf{E}}_r = \begin{bmatrix} \mathrm{Re}(\mathbf{E}_r) & -\mathrm{Im}(\mathbf{E}_r) \\ \mathrm{Im}(\mathbf{E}_r) & \mathrm{Re}(\mathbf{E}_r) \end{bmatrix}, \bar{\mathbf{S}} = \begin{bmatrix} \mathrm{Re}(\mathbf{S}) & -\mathrm{Im}(\mathbf{S})) \\ \mathrm{Im}(\mathbf{S}) & \mathrm{Re}(\mathbf{S}) \end{bmatrix}, \mathbf{x} = [\mathrm{Re}(\mathbf{w}^T)\ \mathrm{Im}(\mathbf{w}^T)]^T$, the problem (2.9) can be rewritten in the real form as below

$$\begin{aligned} \max_{\mathbf{x},p} \min_{m\in\{1,..,M\}} & \mathrm{SNR}_m(\mathbf{x},p) \\ \text{s.t.} \quad & (\mathbf{x},p) \in \bar{\Omega}, \end{aligned} \tag{2.10}$$

where

$$\mathrm{SNR}_m(\mathbf{x},p) = \frac{\mathbf{x}^T\bar{\mathbf{A}}_m\mathbf{x}}{(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)p} + \frac{|h_m|^2}{\sigma_D^2 p} = \frac{\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2}{(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)p}, \tag{2.11}$$

with $\bar{\mathbf{T}}_m = \bar{\mathbf{A}}_m + \bar{\mathbf{B}}_m\frac{|h_m|^2}{\sigma_D^2}$ and

$$\bar{\Omega} = \left\{ (\mathbf{x},p): \begin{array}{l} p>0, \\ \mathbf{x}^T\bar{\mathbf{D}}_r\mathbf{x}/p + \mathbf{x}^T\bar{\mathbf{E}}_r\mathbf{x} \le p_{r,max}\ \forall r\in\{1,...,R\} \\ \sum_{r=1}^R(\mathbf{x}^T\bar{\mathbf{D}}_r\mathbf{x}/p + \mathbf{x}^T\bar{\mathbf{E}}_r\mathbf{x}) \le P_{R,max}, \\ 2/p + \mathbf{x}^T\bar{\mathbf{S}}\mathbf{x}/p \le P_{S,max}, \\ 2/p + \mathbf{x}^T\bar{\mathbf{S}}\mathbf{x}/p + 2\sum_{r=1}^R(\mathbf{x}^T\bar{\mathbf{D}}_r\mathbf{x}/p + \mathbf{x}^T\bar{\mathbf{E}}_r\mathbf{x}) \le P_{T,max} \end{array} \right\}.$$

2.3.2 The Rank-two DCA scheme for solving the problem (2.10)

First of all, the problem (2.10) can be equivalently rewritten as follows

$$\begin{aligned} \min_{\mathbf{x},p} \quad & \max_{m\in\{1,..,M\}} \ln\left(\frac{(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)p}{\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2}\right) \\ \text{s.t.} \quad & (\mathbf{x},p)\in\bar{\Omega}, \end{aligned} \tag{2.12}$$

By introducing a variable t, the problem (2.12) can be equivalently reformulated as

$$\min_{\mathbf{x},p,t} \quad t \tag{2.13}$$

$$\text{s.t.} \quad \ln\left(\frac{(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)p}{\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2}\right) \le t \ \ \forall m\in\{1,...,M\}, \tag{2.14}$$

$$(\mathbf{x},p)\in\bar{\Omega}.$$

The above problem is nonconvex because the constraint (2.14) is nonconvex. In what follows, we formulate this problem as a general DC program and apply DCA for solving it.

DC decomposition for the nonconvex constraints (2.14)
We have

$$\begin{aligned}
&\ln\left(\frac{(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)p}{\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2}\right)\le t\\
&\Leftrightarrow \ln(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)+\ln(p)-\ln(\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2)\le t\\
&\Leftrightarrow G_m(\mathbf{x},p)-H_m(\mathbf{x},p)\le t,
\end{aligned}$$

where $G_m(\mathbf{x},p)=\frac{\rho}{2}\|\mathbf{x}\|^2$ and $H_m(\mathbf{x},p)=\frac{\rho}{2}\|\mathbf{x}\|^2-\ln(\mathbf{x}^T\bar{\mathbf{B}}_m\mathbf{x}+\sigma_D^2)+\ln(\mathbf{x}^T\bar{\mathbf{T}}_m\mathbf{x}+|h_m|^2)-\ln(p)$. It is obviously that G_m is convex for every value of m if $\rho>0$. However, this condition of ρ does not ensure the convexity of H_m. The following theorem shows a sufficient condition of ρ such that H_m is convex.

Theorem 2.1. *Denote ρ_m as the largest eigenvalue of matrix $\frac{2\bar{\boldsymbol{B}}_m}{\sigma_D^2}+\frac{\bar{\boldsymbol{T}}_m}{2|h_m|^2}$ and $\rho_0=\max_{m=1,...,M}\rho_m$. If $\rho\ge\rho_0$ all the function $G_m(\boldsymbol{x},p)$ and $H_m(\boldsymbol{x},p)$ $(m=1,...,M)$ are convex.*

The proof of this theorem is straightforwardly deduced from the following proposition.

Proposition 2.1. *Let $\boldsymbol{B}$ be a $n\times n$ symmetric and positive semidefinite matrix and σ_0 is a constant.*
(i) If ρ is greater than the largest eigenvalue of matrix $\frac{2\boldsymbol{B}}{\sigma_0^2}$ then the function $v_1(\boldsymbol{x})=\frac{1}{2}\rho\|\boldsymbol{x}\|^2-\ln(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}+\sigma_0^2)$ is convex.
(ii) If ρ is greater than the largest eigenvalue of matrix $\frac{\boldsymbol{B}}{2\sigma_0^2}$ then the function $v_2(\boldsymbol{x})=\frac{1}{2}\rho\|\boldsymbol{x}\|^2+\ln(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}+\sigma_0^2)$ is convex.

Proof. (i) The necessary and sufficient condition for the function $v_1(\mathbf{x})$ to be convex is that $\nabla^2 v_1(\mathbf{x})\succeq 0\ \forall\mathbf{x}\in\mathbb{R}^n$.

We have $\nabla^2 v_1(\mathbf{x})=\rho\mathbf{I}-\frac{2\mathbf{B}}{\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2}+\frac{4\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T}{(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)^2}$. Thus

$$\begin{aligned}
&\nabla^2 v_1(\mathbf{x})\succeq 0\\
\Leftrightarrow\ &\rho\|\mathbf{y}\|^2-\frac{2\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2}+\frac{4\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y}}{(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)^2}\ge 0\ \forall\mathbf{y},\ \mathbf{x}\in\mathbb{R}^n.
\end{aligned}$$

Since ρ is greater than the largest eigenvalue of matrix $\frac{\mathbf{2B}}{\sigma_0^2}$, $\rho\mathbf{I}\succeq\frac{\mathbf{2}B}{\sigma_0^2}$. Thus $\mathbf{y}^T\left(\rho\mathbf{I}-\frac{\mathbf{2B}}{\sigma_0^2}\right)\mathbf{y}\ge 0\ \forall\mathbf{y}\in\mathbb{R}^n$. In addition, $\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y}=(\mathbf{y}^T\mathbf{B}\mathbf{x})^2\ge 0\ \forall\mathbf{y},\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{x}^T\mathbf{B}\mathbf{x}\ge 0\ \forall\mathbf{x}\in\mathbb{R}^n$ due to the positive semidefinite property of $\mathbf{B}$. Therefore

$$\rho\|\mathbf{y}\|^2-\frac{2\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2}+\frac{4\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y}}{(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)^2}\ge\frac{2\mathbf{y}^T\mathbf{B}\mathbf{y}}{\sigma_0^2}-\frac{2\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2}\ge 0.$$

(ii) Similarly to the part (i), the function $v_2(\mathbf{x})$ is convex if and only if

$$\begin{aligned}
&\nabla^2 v_2(\mathbf{x})\succeq 0\ \ \forall\mathbf{x}\in\mathbb{R}^n\\
\Leftrightarrow\ &\rho\|\mathbf{y}\|^2+\frac{2\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2}-\frac{4\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y}}{(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)^2}\ge 0\ \forall\mathbf{y},\ \mathbf{x}\in\mathbb{R}^n\\
\Leftrightarrow\ &\rho\|\mathbf{y}\|^2(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)^2+2\mathbf{y}^T\mathbf{B}\mathbf{y}(\mathbf{x}^T\mathbf{B}\mathbf{x}+\sigma_0^2)-4\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y}\ge 0\ \forall\mathbf{y},\ \mathbf{x}\in\mathbb{R}^n.
\end{aligned}$$

The Cauchy-Schwarz inequality implies that

$$\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y} \leq (\mathbf{x}^T\mathbf{B}\mathbf{x})(\mathbf{y}^T\mathbf{B}\mathbf{y}) \ \forall \mathbf{y},\ \mathbf{x} \in \mathbb{R}^n.$$

Moreover, the Cauchy inequality shows that

$$(\mathbf{x}^T\mathbf{B}\mathbf{x} + \sigma_0^2)^2 \geq 4\sigma_0^2(\mathbf{x}^T\mathbf{B}\mathbf{x}) \ \forall \mathbf{x} \in \mathbb{R}^n.$$

Therefore

$$\begin{aligned} & \rho\|\mathbf{y}\|^2(\mathbf{x}^T\mathbf{B}\mathbf{x} + \sigma_0^2)^2 + 2\mathbf{y}^T\mathbf{B}\mathbf{y}(\mathbf{x}^T\mathbf{B}\mathbf{x} + \sigma_0^2) - 4\mathbf{y}^T\mathbf{B}\mathbf{x}(\mathbf{B}\mathbf{x})^T\mathbf{y} \\ \geq\ & 4\sigma_0^2(\rho\|\mathbf{y}\|^2)(\mathbf{x}^T\mathbf{B}\mathbf{x}) - 2(\mathbf{y}^T\mathbf{B}\mathbf{y})(\mathbf{x}^T\mathbf{B}\mathbf{x}) \\ \geq\ & 0 \ \forall \mathbf{y},\ \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

The last inequality is deduced from the fact that ρ is greater than the greatest eigenvalue of matrix $\frac{\mathbf{B}}{2\sigma_0^2}$, hence $\rho\mathbf{I} \succeq \frac{\mathbf{B}}{2\sigma_0^2}$ that implies $4\sigma_0^2\rho\|\mathbf{y}\|^2 \geq 2\mathbf{y}^T\mathbf{B}\mathbf{y}$ and $\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0$ since $\mathbf{B} \succeq 0$. □

In summary, with $\rho = \rho_0$ indicated in Theorem 2.1, the DC formulation of the problem (2.13) is of the following form

$$\begin{aligned} \min_{\mathbf{x},p,t} \quad & t \\ \text{s.t.} \quad & G_m(\mathbf{x},p) - H_m(\mathbf{x},p) \leq t \\ & (\mathbf{x},p) \in \bar{\Omega}. \end{aligned} \tag{2.15}$$

Following the idea of DCA, at the kth iteration, the second DC component H_m is approximated by its linear minorant at the iterate $(\mathbf{x}^k, p^k)$, which is given by

$$H_m^k(\mathbf{x},p) = H_m(\mathbf{x}^k,p^k) + \langle \mathbf{y}_m^k, \mathbf{u} - \mathbf{u}^k \rangle,$$

where $\mathbf{u}^k = \left[(\mathbf{x}^k)^T\ p^k\right]^T, \mathbf{u} = \left[\mathbf{x}^T\ p\right]^T$ and $\mathbf{y}_m^k \in \partial H_m(\mathbf{x}^k,p^k)$. Since the function $H_m(\mathbf{x},p)$ is differentiable, its subgradient at a point $(\mathbf{x}^k,p^k)$ is computed by

$$\mathbf{y}_m^k = \nabla H_m(\mathbf{x}^k,p^k) = \left[\left(\rho\mathbf{x}^k - \frac{2\bar{\mathbf{B}}_m\mathbf{x}^k}{(\mathbf{x}^k)^T\bar{\mathbf{B}}_m\mathbf{x}^k + \sigma_D^2} + \frac{2\bar{\mathbf{T}}_m\mathbf{x}^k}{(\mathbf{x}^k)^T\bar{\mathbf{T}}_m\mathbf{x}^k + |h_m|^2}\right)^T - \frac{1}{p^k}\right]^T.$$

DCA applied to (2.15) involves solving a sequential convex subproblems of the following form

$$\min_{\mathbf{x},p,t} \quad t \tag{2.16}$$

$$\begin{aligned} \text{s.t.} \quad & G_m(\mathbf{x},p) - H_m^k(\mathbf{x},p) \leq t \ \ \forall m \in \{1,...,M\}, \\ & (\mathbf{x},p) \in \bar{\Omega}. \end{aligned} \tag{2.17}$$

The general DCA scheme for the Rank-two max-min fairness optimization problem is described in the scheme below.

The Rank-Two DCA scheme for DC program (2.15) (DCA-R2)

- **Initialization.** Choose an initial point $\mathbf{u}^0 = \left[(\mathbf{x}^0)^T \; p^0\right]^T$, t^0 and the tolerance ϵ_1. $k \longleftarrow 0$.

- **Repeat.**
  Step 1. For each $k, \mathbf{u}^k = \left[(\mathbf{x}^k)^T \; p^k\right]^T$ is known, solving the convex subproblem (2.16) to find $\mathbf{u}^{k+1} = \left[(\mathbf{x}^{k+1})^T \; p^{k+1}\right]^T$ and t^{k+1}.
  Step 2. $k \leftarrow k + 1$.

- **Until** either $\frac{t^{k-1}-t^k}{t^{k-1}+1} < \epsilon_1$ or $\frac{\|\mathbf{u}^{k-1}-\mathbf{u}^k\|}{\|\mathbf{u}^{k-1}\|+1} < \epsilon_1$.

From the description of the feasible set $\bar{\Omega}$, it can be realized that the variable $\mathbf{x}$ is bounded and the variable p is bounded from below by $\frac{2}{P_{S,max}}$. Furthermore, to avoid the case when the power scaling factor at the source p_1 approaches zero, the value of p should not be too large since $|p_1|^2 = \frac{1}{p}$. Therefore, we can assume that p is also bounded from above by a large enough number, say $p_{max} = 10^6$. This combined with the continuity of the functions defining $\bar{\Omega}$ results in the compactness of $\bar{\Omega}$. The following Theorem shows the convergence of the algorithm DCA-R2.

Theorem 2.2.

(1) DCA-R2 generates a sequence $\{(\boldsymbol{x}^k, p^k, t^k)\}$ such that the sequence of the corresponding objective function values $\{t^k\}$ is decreasing.

(2) Every limit point of the sequence $\{(\boldsymbol{x}^k, p^k, t^k)\}$ generated by DCA-R2 is a critical point to the problem (2.15).

To prove this theorem, we first prove the following lemma.

Lemma 2.1. *Assume that $\{(\boldsymbol{x}^{k_j}, p^{k_j}, t^{k_j})\}$ is a subsequence of the sequence generated by DCA-R2, there exists a subsequence $\{(\boldsymbol{x}^{k_{j_s}}, p^{k_{j_s}}, t^{k_{j_s}})\}$ satisfying $\lim_{s\to\infty} \|\boldsymbol{u}^{k_{j_s}} - \boldsymbol{u}^{k_{j_s}+1}\|^2 = 0$, where $\boldsymbol{u}^{k_{j_s}} = \left[(\boldsymbol{x}^{k_{j_s}})^T \; p^{k_{j_s}}\right]^T$.*

Proof. Because the functions G_m and H_m^k in (2.17) do not depend on t, it is easy to verify that the Slater's constraint qualification is satisfied. Furthermore $(\mathbf{x}^{k_j+1}, p^{k_j+1}, t^{k_j+1})$ is the optimal solution to the problem (2.16), therefore there exist some $\lambda_m^{k_j+1} \in \mathbb{R}, m = 1, ..., M$ such that

$$\bullet\ 0 \in \sum_{m=1}^{M} \lambda_m^{k_j+1}(\nabla G_m(\mathbf{u}^{k_j+1}) - \nabla H_m(\mathbf{u}^{k_j})) + N(\bar{\Omega}, \mathbf{u}^{k_j+1}), \tag{2.18}$$

$$\bullet\ 1 - \sum_{m=1}^{M} \lambda_m^{k_j+1} = 0, \quad \mathbf{u}^{k_j+1} \in \bar{\Omega},$$

$$\bullet\ G_m(\mathbf{u}^{k_j+1}) - H_m(\mathbf{u}^{k_j}) - \langle \nabla H_m(\mathbf{u}^{k_j}), \mathbf{u}^{k_j+1} - \mathbf{u}^{k_j} \rangle \le t^{k_j+1},\ \lambda_m^{k_j+1} \ge 0 \ \forall m = 1, ..., M,$$

$$\bullet\ \lambda_m^{k_j+1}(G_m(\mathbf{u}^{k_j+1}) - H_m(\mathbf{u}^{k_j}) - \langle \nabla H_m(\mathbf{u}^{k_j}), \mathbf{u}^{k_j+1} - \mathbf{u}^{k_j} \rangle - t^{k_j+1}) = 0 \ \forall m = 1, ..., M,$$

where $\mathbf{u}^{k_j} = \left[(\mathbf{x}^{k_j})^T \ p^{k_j}\right]^T$. Denote $\mathbf{u} = \left[(\mathbf{x})^T \ p\right]^T$.
Since $G_m(\mathbf{u}) = \frac{1}{2}\rho\|\mathbf{x}\|^2, \nabla^2_{\mathbf{x}} G_m = \rho I$, which shows that G_m is strongly convex in $\mathbf{x}$ with modulus ρ. Because of this property we have

$$G_m(\mathbf{u}^{k_j}) \geq G_m(\mathbf{u}^{k_j+1}) + \langle \nabla G_m(\mathbf{u}^{k_j+1}), \mathbf{u}^{k_j} - \mathbf{u}^{k_j+1} \rangle + \frac{\rho}{2}\|\mathbf{x}^{k_j} - \mathbf{x}^{k_j+1}\|^2 \quad (2.19)$$

The function $y = -\ln(p)$ is strongly convex because $p_{max} \geq p \geq \frac{2}{P_{Smax}}$, so $H_m(\mathbf{u})$ is also strongly convex in p. Therefore there exists $\tau_m > 0$ such that

$$H_m(\mathbf{u}^{k_j+1}) \geq H_m(\mathbf{u}^{k_j}) + \langle \nabla H_m(\mathbf{u}^{k_j}), \mathbf{u}^{k_j+1} - \mathbf{u}^{k_j} \rangle + \frac{\tau_m}{2}\|p^{k_j} - p^{k_j+1}\|^2 \quad (2.20)$$

Adding (2.19) to (2.20) we obtain

$$\langle \nabla G_m(\mathbf{u}^{k_j+1}) - \nabla H_m(\mathbf{u}^{k_j}), \mathbf{u}^{k_j} - \mathbf{u}^{k_j+1} \rangle \leq F_m(\mathbf{u}^{k_j}) - F_m(\mathbf{u}^{k_j+1}) - \frac{\rho_m}{2}\|\mathbf{u}^{k_j} - \mathbf{u}^{k_j+1}\|^2 \quad (2.21)$$

where $F_m(\mathbf{u}) = G_m(\mathbf{u}) - H_m(\mathbf{u})$, $\rho_m = \min\{\tau_m, \rho\}$. In addition, it is deduced from the first inclusion of (2.18) that

$$\sum_{m=1}^{M} \lambda_m^{k_j+1} \langle \nabla G_m(\mathbf{u}^{k_j+1}) - \nabla H_m(\mathbf{u}^{k_j}), \mathbf{u}^{k_j} - \mathbf{u}^{k_j+1} \rangle \geq 0,$$

thus for each k_j there exists m_{k_j} such that

$$\langle \nabla G_{m_{k_j}}(\mathbf{u}^{k_j+1}) - \nabla H_{m_{k_j}}(\mathbf{u}^{k_j}), \mathbf{u}^{k_j} - \mathbf{u}^{k_j+1} \rangle \geq 0. \quad (2.22)$$

It is gained from (2.22) and (2.21) that

$$\frac{\rho_{m_{k_j}}}{2}\|\mathbf{u}^{k_j} - \mathbf{u}^{k_j+1}\|^2 \leq F_{m_{k_j}}(\mathbf{u}^{k_j}) - F_{m_{k_j}}(\mathbf{u}^{k_j+1}).$$

Since $m_{k_j} \in \{1, ..., M\}$ that has finite elements while k_j is in an infinite set, there exists m_* such that there are infinitely many indexes $\{k_{j_s}\}$ satisfying

$$\frac{\rho_{m_*}}{2}\|\mathbf{u}^{k_{j_s}} - \mathbf{u}^{k_{j_s}+1}\|^2 \leq F_{m_*}(\mathbf{u}^{k_{j_s}}) - F_{m_*}(\mathbf{u}^{k_{j_s}+1}) \ \forall s. \quad (2.23)$$

This inequality shows that the sequence $\{F_{m_*}(\mathbf{u}^{k_{j_s}})\}$ is decreasing. Moreover $F_{m_*}(\mathbf{u})$ is continuous on the compact set $\bar{\Omega}$ so it is bounded. Therefore the sequence $\{F_{m_*}(\mathbf{u}^{k_{j_s}})\}$ is also bounded and thus convergent. This combined with (2.23) leads to $\lim_{s\to\infty} \|\mathbf{u}^{k_{j_s}} - \mathbf{u}^{k_{j_s}+1}\| = 0$. □

It is now ready to prove the Theorem 2.

Proof.

(1) The decrease of sequence $\{t^k\}$ is straightforwardly deduced due to the fact that $(\mathbf{x}^{k+1}, p^{k+1}, t^{k+1})$ is a minimizer of (2.16) while $(\mathbf{x}^k, p^k, t^k)$ is a feasible point of (2.16).

(2) Assume $(\mathbf{x}^*, p^*, t^*)$ is a limit point of the sequence $\{(\mathbf{x}^k, p^k, t^k)\}$. Therefore there exists a subsequence $\{(\mathbf{x}^{k_j+1}, p^{k_j+1}, t^{k_j+1})\}$ such that

$$\lim_{j\to\infty}(\mathbf{x}^{k_j+1}, p^{k_j+1}, t^{k_j+1}) = (\mathbf{x}^*, p^*, t^*).$$

Apply the above Lemma, for the sequence $\{(\mathbf{x}^{k_j+1}, p^{k_j+1}, t^{k_j+1})\}$ there exists a subsequence $\{(\mathbf{x}^{k_{j_s}+1}, p^{k_{j_s}+1}, t^{k_{j_s}+1})\}$ such that $\lim_{s\to\infty}\|\mathbf{u}^{k_{j_s}} - \mathbf{u}^{k_{j_s}+1}\| = 0$. Moreover $\lim_{s\to\infty}(\mathbf{x}^{k_{j_s}+1}, p^{k_{j_s}+1}) = (\mathbf{x}^*, p^*)$, hence $\lim_{s\to\infty}(\mathbf{x}^{k_{j_s}}, p^{k_{j_s}}) = (\mathbf{x}^*, p^*)$. In addition, the sequence $\{\lambda_m^{k_{j_s}+1}\}$ is bounded for every $m = 1, ..., M$ thus without loss of generality we can assume that $\lim_{s\to\infty}\lambda_m^{k_{j_s}+1} = \lambda_m^*$.

Replace k in (2.18) by k_{j_s} and taking limits as $s \to \infty$, we obtain

$$\begin{aligned}
&\bullet\ 0 \in \sum_{m=0}^{M}\lambda_m^*(\nabla G_m(\mathbf{u}^*) - \nabla H_m(\mathbf{u}^*)) + N(\bar{\Omega}, \mathbf{u}^*),\\
&\bullet\ 1 - \sum_{m=1}^{M}\lambda_m^* = 0, \quad \mathbf{u}^* \in \bar{\Omega},\\
&\bullet\ G_m(\mathbf{u}^*) - H_m(\mathbf{u}^*) \le t^*, \quad \lambda_m^* \ge 0\ \forall m = 1, ..., M,\\
&\bullet\ \lambda_m^*(G_m(\mathbf{u}^*) - H_m(\mathbf{u}^*) - t^*) = 0, \forall m = 1, ..., M,
\end{aligned}$$

where $\mathbf{u}^* = [(\mathbf{x}^*)^T, p^*]^T$. It shows that $(\mathbf{x}^*, p^*, t^*)$ is a critical point of the problem (2.13).

□

DCA scheme for Rank-one model, called DCA-R1, is a particular case of DCA-R2 in which the second beamforming vector $\mathbf{w}_2 = 0_R$ and scaling factors p_3, p_4 are set to zeros.

2.4 Experimental Results

In our experiments, all algorithms were implemented in the Matlab 2013b, and performed on a PC Intel Core i5-2500S CPU 2.70GHz of 4GB RAM.

2.4.1 Comparative algorithms

In this section, we give the numerical performance obtained by the DCA based algorithms (DCA-R2, DCA-R1) and then compare them with those obtained by CCCP-R2, CCCP-R1, SDR2D-R2 and SDR-R1 algorithms, which are mentioned in [86]. In all experiments, the average minimum achieved rate (MAR), which is calculated by $\frac{1}{2}\log_2(1 + \min_{m=1,...,M}\{\mathrm{SNR}_m\})$, is used for comparing among algorithms.

2.4.1.1 CCCP-R2 and CCCP-R1

CCCP-R2 is actually a DCA based algorithm but it is based on a DC decomposition different from our proposed one. First of all, the authors in [86] reformulated the max-min fairness optimization problem (2.9) into the following form

$$\min_{\mathbf{w},p,t} \quad t \tag{2.24}$$

$$\text{s.t.} \quad \text{SNR}_m(\mathbf{w},p) \geq \frac{1}{t} \ \ \forall m \in \{1,...,M\}, \tag{2.25}$$
$$(\mathbf{w},p) \in \Omega, \ t > 0.$$

Afterwards, the SNR constraint (2.25) is decomposed into a difference of two convex functions as below

$$\frac{\mathbf{w}^\dagger \mathbf{B}_m \mathbf{w} + \sigma_D^2}{t} - \frac{\mathbf{w}^\dagger (\mathbf{A}_m + (|h_m|^2/\sigma_D^2)\mathbf{B}_m)\mathbf{w} + |h_m|^2}{p} \leq 0.$$

Apply the idea of DC programming and DCA, a sequence $\{(\mathbf{w}^k, p^k, t^k)\}$ is constructed by iteratively solving the convex subproblems which are obtained by linearizing the second DC component of SNR constraints at each iteration.

CCCP-R1 is a special case of CCCP-R2 in which some elements in the variable $\mathbf{w}$ are zero. More specifically, in the Rank-1 model, single beamformer is used to transmit symbol instead of two beamformers as in the Rank-two model, thus $\mathbf{w}_2 = 0_R$. Furthermore, in the Rank-one model, each symbol is transmitted in only two time slots in which the source sends the signal to the relays in the first time slot and then the relays forward these received signals to the destinations in the second time slot. Therefore, the scaling factors p_3 and p_4 in the third and fourth time slots of the Rank-two model are assigned to 0 in the Rank-one model.

2.4.1.2 SDR2D-R2 and SDR2D-R1

It should be noted that $\mathbf{w}^\dagger \mathbf{A} \mathbf{w} = \text{tr}(\mathbf{A}\mathbf{X})$, where $\mathbf{X} = \mathbf{w}\mathbf{w}^\dagger$ is a positive semidefinite matrix with $\text{rank}(\mathbf{X}) = 1$. Thanks to this property, the problem (2.9) can be reformulated as a semidefinite program (SDP) as follows

$$\min_{\mathbf{X}_1,\mathbf{X}_2,p,t} \quad t \tag{2.26}$$

$$\text{s.t.} \quad \text{tr}((\mathbf{X}_1 + \mathbf{X}_2)\hat{\mathbf{A}}_{m,1}) \geq \left(\frac{p}{t} - \frac{|h_m|^2}{\sigma_D^2}\right)\left(\text{tr}((\mathbf{X}_1 + \mathbf{X}_2)\hat{\mathbf{B}}_m) + \sigma_D^2\right) \ \ \forall m \in \{1,...,M\},$$
$$(\mathbf{X}_1 + \mathbf{X}_2, p) \in \Upsilon, \ t > 0,$$
$$\mathbf{X}_1 \succeq 0, \mathbf{X}_2 \succeq 0,$$

where the constraints $\text{rank}(\mathbf{X}_1) = \text{rank}(\mathbf{X}_2) = 1$ are discarded and

$$\Upsilon = \left\{ (\mathbf{X},p): \begin{array}{l} \text{tr}(\mathbf{X}(\hat{\mathbf{D}}_r/p + \hat{\mathbf{E}}_r)) \leq p_{r,max}, \\ \sum_{r=1}^R \text{tr}(\mathbf{X}(\hat{\mathbf{D}}_r/p + \hat{\mathbf{E}}_r)) \leq P_{R,max} \\ 2/p + \text{tr}(\mathbf{X}\hat{\mathbf{S}})/p \leq P_{S,max}, \\ 2/p + \text{tr}(\mathbf{X}\hat{\mathbf{S}})/p + 2\sum_{r=1}^R \text{tr}(\mathbf{X}(\hat{\mathbf{D}}_r/p + \hat{\mathbf{E}}_r)) \leq P_{T,max}, \end{array} \right\}.$$

In the Rank-one model, $\mathbf{w}_2 = 0$ leads to $\mathbf{X}_2 = 0$, thus the corresponding SDP is reduced to

$$\begin{aligned} \min_{\mathbf{X}_1,p,t} \quad & t \\ \text{s.t.} \quad & \text{tr}(\mathbf{X}_1\hat{\mathbf{A}}_{m,1}) \geq \left(\frac{p}{t} - \frac{|h_m|^2}{\sigma_D^2}\right)\left(\text{tr}(\mathbf{X}_1\hat{\mathbf{B}}_m) + \sigma_D^2\right) \quad \forall m \in \{1,...,M\}, \\ & (\mathbf{X}_1, p) \in \Upsilon,\ t > 0, \\ & \mathbf{X}_1 \succeq 0, \end{aligned} \tag{2.27}$$

SDR2D-R2 and SDR2D-R1 are to solve the above SDP programs, respectively and then generate a feasible point for the original problem. It was proved in [86] that (2.26) and (2.27) are equivalent. The difference between SDR2D-R2 and SDR2D-R1 resides in the step of obtaining a feasible solution from the found matrix $\mathbf{X}_1$.

Note that the problem (2.27) is nonconvex, however for a particular value of t and p the feasibility problem to compute a feasible matrix $\mathbf{X}_1$ is a convex SDP. Therefore, a grid search on p was performed and then for each point of p chosen from this grid, a bisection algorithm was deployed to find the optimum t^* and the matrix $\mathbf{X}_1^*$.

For SDR2D-R2 scheme, if $\text{rank}(\mathbf{X}_1^*) > 2$, it cannot give a solution to the original problem since there does not exist rank-two decomposition $\mathbf{X}_1^* = \hat{\mathbf{w}}_1\hat{\mathbf{w}}_1 + \hat{\mathbf{w}}_2\hat{\mathbf{w}}_2^\dagger$. To achieve a feasible solution to the original problem, one has to employ randomization techniques ([112]) that results in a suboptimal solution, in general.

Similarly, for SDR2D-R1 scheme, if $\text{rank}(\mathbf{X}_1^*) > 1$, randomization techniques ([90]) are used to find a feasible point to the original problem.

2.4.2 Simulated datasets and parameter setting

In our experiments, we consider a network with $R = 10$ relay nodes. The channel coefficients are assumed to be independent from each other. Specifically, $f_i, h_m, g_{m,i}$ are modeled as

$$\begin{aligned} f_i &= \overline{f}_i + \widehat{f}_i\ \forall i \in \{1,..,R\}, \\ h_m &= \overline{h}_m + \widehat{h}_m\ \forall m \in \{1,..,M\}, \\ g_{m,i} &= \overline{g}_{m,i} + \widehat{g}_{m,i}\ \forall m \in \{1,..,M\},\ \forall i \in \{1,..,R\}, \end{aligned}$$

where $\overline{f}_i, \overline{h}_m, \overline{g}_{m,i}$ are complex channel mean and $\widehat{f}_i, \widehat{h}_m, \widehat{g}_{m,i}$ are zero-mean random variables $\forall m \in \{1,..,M\},\ \forall i \in \{1,..,R\}$. According to [29], the channel mean $\overline{f}_i, \overline{h}_m, \overline{g}_{m,i}$ can be modeled, respectively, as

$$\overline{f}_i = \frac{exp(\sqrt{-1}\Theta_i)}{\sqrt{\Gamma_f}}, \overline{h}_m = \frac{exp(\sqrt{-1}\Omega_m)}{\sqrt{\Gamma_h}}, \overline{g}_{m,i} = \frac{exp(\sqrt{-1}\Upsilon_{m,i})}{\sqrt{\Gamma_g}},$$

where the random angles $\Theta_i, \Omega_m, \Upsilon_{m,i}$ are chosen to be uniformly distributed on the interval $[0, 2\pi]$ $\forall m \in \{1,..,M\},\ \forall i \in \{1,..,R\}$, and $\Gamma_f, \Gamma_h, \Gamma_g$ are positive constants,

which indicate the uncertainty in the channel coefficients. Moreover, the variances of the random variables are given by

$$E\{|\widehat{f}_i|^2\} = \frac{\Gamma_f}{\Gamma_f + 1}, E\{|\widehat{h}_m|^2\} = \frac{\Gamma_h}{\Gamma_h + 1}, E\{|\widehat{g}_{m,i}|^2\} = \frac{\Gamma_g}{\Gamma_g + 1}.$$

In this chapter, we choose $\Gamma_f = \Gamma_h = \Gamma_g = 10$. The noise powers at the relays and the destinations are set to $\sigma_R^2 = \sigma_D^2 = 1$.
Based on the above information, we independently generated 50 datasets and used them to test all the algorithms. The mean of minimum achievable rate and computing time of all the algorithms performed on these 50 datasets were recorded to compare. The maximum transmit power values are chosen such that $P_{S,max} = P_{T,max}/2$, $P_{R,max} = P_{T,max}/3$ and $p_{r,max} = P_{T,max}/15$. The total transmit power value, $P_{T,max}$, and the number of destinations, M, in the network are set differently in various experiments. More particularly, in the first experiment $P_{T,max} = 5$ and M is chosen from the set $\{20, 40, 60, 80, 100\}$ whereas in the second experiment $M = 100$ and $P_{T,max}$ is chosen from the set $\{5, 10, 15, 20\}$. The tolerance in the DCA schemes is set to $\epsilon_1 = 10^{-4}$.

2.4.3 Numerical results and comments

2.4.3.1 The first experiment: Minimum achievable rate versus number of destinations

Table 2.1: Comparison of Minimum Achievable Rate(MAR) obtained by all the algorithms versus number M of destinations

M		DCA-R2	CCCP-R2	SDR-R2	DCA-R1	CCCP-R1	SDR-R1
20	MAR	**0.4615**	0.4150	0.2618	**0.4120**	0.2869	0.2208
	CPU(s)	**39.546**	83.298	1151.361	28.483	**22.422**	905.894
40	MAR	**0.3672**	0.3127	0.1795	**0.3025**	0.1896	0.1353
	CPU(s)	**63.461**	125.233	1443.744	**40.681**	41.021	1226.347
60	MAR	**0.3308**	0.2700	0.1444	0.2651	0.1601	0.1181
	CPU(s)	**61.120**	177.521	1479.493	**48.473**	57.316	1165.753
80	MAR	**0.3163**	0.2553	0.1278	**0.2362**	0.1401	0.093
	CPU(s)	**82.166**	197.416	1726.189	**58.170**	98.805	1410.497
100	MAR	**0.3029**	0.2355	0.1216	**0.2205**	0.1351	0.082
	CPU(s)	**112.638**	287.797	1927.019	**64.245**	132.829	1688.824

Table 1 demonstrates the average minimum rate versus the number of destinations M in case of $P_{T,max} = 5$. There is an obvious fact indicated by Table 1 that when the number of destination is increasing, the minimum achievable rate attained from all the algorithms diminishes. Moreover, a noticeable feature can be observed from Table 1 is that DCA based algorithms always bring the superior minimum achievable rate while they spend time far below compared with SDR based ones. The ratio of running time between DCA based schemes and SDR based ones are significant, up to 29 times. In addition, the difference between minimum achievable rate obtained from the Rank-two

and Rank-one models respectively shows efficiency of the former compared with the latter.
For more details, in the Rank-two model, DCA-R2 is leading, respectively followed by CCCP-R2 and SDR-R2 in terms of minimum achievable rate. Our proposed DCA is better than the existing DCA based one CCCP-R2 in both computing time and minimum achievable rate. SDR-R2 is the worst, which not only spends time the most but also returns the smallest minimum achievable rate.
In the Rank-one model, the behavior of algorithms are quite similar to that of the algorithms in the Rank-two model. The proposed DCA is still most successful in gaining the best minimum achievable rate while SDR-R1 is still in the last position. DCA-R1 produces the better minimum achievable rate than CCCP-R1 does whereas it takes a shorter time to run. SDR-R1 is both expensive and ineffective.

2.4.3.2 The second experiment: Minimum achievable rate versus total power

Table 2.2: Comparison of Minimum Achievable Rate (MAR) obtained by all the algorithms versus Total Power P_t

P_t		DCA-R2	CCCP-R2	SDR-R2	DCA-R1	CCCP-R1	SDR-R1
5	MAR	**0.3029**	0.2355	0.1216	**0.2205**	0.1349	0.082
	CPU(s)	**112.638**	287.797	1927.019	**64.245**	78.515	1688.824
10	MAR	**0.5503**	0.4440	0.2362	**0.4009**	0.2602	0.1662
	CPU(s)	**132.134**	289.991	1968.401	81.279	**75.925**	1690.957
15	MAR	**0.7404**	0.6103	0.3522	**0.5693**	0.3657	0.2471
	CPU(s)	**158.338**	262.618	1948.749	**70.261**	91.876	1700.362
20	MAR	**0.8800**	0.7180	0.4258	**0.6739**	0.4557	0.2973
	CPU(s)	**156.730**	266.726	1804.864	**82.304**	90.937	1692.598

Table 2 depicts the average minimum rate versus $P_{T,max}$ in case of $M = 100$. It can be realized from this table that DCA based algorithms return the better minimum achievable rate compared with the others in both the Rank-two and Rank-one models. In addition, the minimum achievable rates obtained by all the algorithms in the Rank-two model are always larger than those of the corresponding algorithms in Rank-one model.
For Rank-two model, DCA-R2 is the best while SDR-R2 is the worst in terms of both minimum achievable rate and running time. The minimum achievable rate obtained by DCA based algorithms is more than twice that achieved by SDR based one while the running time of DCA based schemes is much less than that of SDR based ones. Between DCA based algorithms, CCCP-R2 gives the worse minimum achievable rate than the other does whereas it consumes more time.
For Rank-one model, DCA-R1 provides the best minimum achievable rate followed by CCCP-R1 and SDR-R1, respectively. In computing time aspect, SDR-R1 is so expensive compared with the others due to the use of two-dimension search in combination with iterative algorithm at each point in the searched grid. Between two DCA based schemes, DCA-R1 not only saves time but also produces the superior minimum

achievable rate.
Finally, both Rank-two and Rank-one models show a common trend that the rise of total power results in an increase of minimum achievable rate obtained by all the algorithms.

2.5 Conclusion

In this chapter, we have reformulated a nonconvex max-min fairness optimization problem as a general DC program based on a novel DC decomposition and then designed a general DCA scheme for solving it. The experimental performances reveal efficiency of the proposed general DCA scheme in terms of both the highest minimum achievable rate and running time compared to the existing methods. In comparison with the previous DCA scheme, our DCA scheme furnishes the superior minimum achievable rates while consuming less time. Similarly, the proposed DCA scheme brings higher minimum achievable rate and spend much less time than the SDR based approach does. In addition, we prove rigorously the convergence of the proposed DCA scheme.

The approach based on a general DCA scheme for solving nonconvex optimization problems has just been exploited in some recent works. It permits to solve a wider class of optimization problems compared with the approach based on a standard DCA, thus being a promising optimization tool to deal with the hard problems in various areas of the applied science.

Chapter 3

DC Programming and DCA for Enhancing Physical Layer Security via Cooperative Jamming[1]

Abstract: *The explosive development of computational tools these days is threatening security of cryptographic algorithms, which are regarded as primary traditional methods for ensuring information security. The physical layer security approach is introduced as a method for both improving confidentiality of the secret key distribution in cryptography and enabling the data transmission without basing on higher-layer encryption. In this chapter, the cooperative jamming paradigm - one of the techniques used in the physical layer is studied and the resulting power allocation problem with the aim of maximizing the sum of secrecy rates subject to power constraints is formulated as a nonconvex optimization problem. The objective function is a so-called DC (Difference of Convex functions) function, and some constraints are coupling. We propose a new DC formulation and develop an efficient DCA (DC Algorithm) to deal with this nonconvex program. The DCA introduces the elegant concept of approximating the original nonconvex program by a sequence of convex ones: at each iteration of DCA requires solution of a convex subproblem. The main advantage of the proposed approach is that it leads to strongly convex quadratic subproblems with separate variables in the objective function, which can be tackled by both distributed and centralized methods. One of the major contributions of the chapter is to develop a highly efficient distributed algorithm to solve the convex subproblem. We adopt the dual decomposition method that results in computing iteratively the projection of points onto a very simple structural set which can be determined by an inexpensive procedure. The numerical results show the efficiency and the superiority of the new DCA based algorithm compared with existing approaches.*

1. The material of this chapter is developed from the following works:
[1]. Thi Thuy Tran, Hoai An Le Thi, Tao Pham Dinh. DC programming and DCA for Enhancing Physcial Layer Security via Cooperative Jamming. *Publish online* in Computers and Operations Research .

3.1 Introduction of Physical Layer Security

Wireless communication these days has a great impact on all aspects of life. More and more people use services from applications of wireless communication such as e-banking, e-commerce and medical service. Therefore, confidentiality and privacy of information over the wireless medium are mandatory requirements in wireless network design. Nevertheless, the broadcast nature of wireless medium makes it susceptible to eavesdropping by illegal receivers. Therefore ensuring secure communication is always a big challenge in wireless system design. Traditionally, this security task is mainly relied on cryptographic algorithms. More specifically, in symmetric-key cryptosystems, by using a private key, the data is encrypted before being sent over the channels and then decrypted at receivers. However, an issue arises is how to distribute the secret key securely. In addition, cryptographic algorithms are based on a hypothesis that it is computationally impossible for unauthorized receivers to decrypt data without secret key. Nonetheless, the explosive development of powerful computing tools nowadays, with the advent of quantum computers, shakes the faith in validity of this hypothesis. In fact, the vulnerability of some recent cryptographic systems is a warning to the security of this method. Therefore, it requires to develop new technologies in data transmission to ensure security and confidentiality for transmitted data, besides cryptography. In this context, physical layer security emerges as an effective method for both ensuring secure transmission without using encryption and aiding secret key exchange in cryptography. Its principle is to exploit the physical features of the wireless channel to ensure secure communications. It was first studied by Wyner in [114] based on information theory. In this work, he considered a basic wiretap channel including a source, a destination, and an eavesdropper, which operates as follows. First, a K-message $\mathbf{S}^{(K)} = (S_1, ..., S_K)$ is chosen randomly from a message set $\mathcal{S}^K$, where $S_1, ..., S_K$ are identically distributed random variables taking value in a finite set $\mathcal{S}$ and having entropy $H(S_k) = H_S$. This message then is encoded as a N-sequence by the encoding function

$$f_{enc} : \mathcal{S}^K \to \mathcal{X}^N, \mathbf{S}^{(K)} \to X^{(N)}.$$

In turn, this N-sequence is an input to the channel. The output of the channel at the legitimate receiver is denoted by $Y^{(N)}$. It is then decoded by the function

$$f_{dec} : \mathcal{Y}^N \to \mathcal{S}^K, Y^{(N)} \to \hat{\mathbf{S}}^{(K)} = (\hat{S}_1, ..., \hat{S}_k).$$

The decoding error rate P_e is defined as

$$P_e = \frac{1}{K} \sum_{k=1}^{K} \Pr(\hat{S}_k \neq S_k).$$

The level of secrecy is measured by the uncertainty of the eavesdropper about the message $\mathbf{S}^{(K)}$ sent by the source under the condition that the eavesdropper receives $\mathbf{Z}^{(N)}$. This measure is called equivocation rate, given by

$$R_e = \frac{1}{K} H(\mathbf{S}^{(K)} | \mathbf{Z}^{(N)}).$$

A pair (R, d) is said to be achievable over the wiretap channel if for any $\epsilon > 0$, there exists an encoder-decoder with parameters (N, K, R_e, P_e) such that

$$\begin{aligned} P_e &\leq \epsilon \\ R_e &\geq d - \epsilon \\ \frac{KH_S}{N} &\geq R - \epsilon \end{aligned}$$

The secrecy capacity C_S is the supremum of all achievable secrecy rates R when $d = H_S$. The results in [114] showed that when the channel between the source and the destination is better than the channel between the source and the eavesdropper, a message can be encoded in a way that allows it to be reliably decoded at the destination while causing significant confusion at the eavesdropper. In other words, there exists a $C_S > 0$ such that the message can be reliably transmitted at rates up to C_S while being ensured approximately perfect secrecy. This approach of Wyner has been later extended to Gaussian channel ([57]), parallel channels ([116],[117]), fading channel ([64]), multiple access channel ([96],[97],[98],[99]), broadcast channel with confidential messages ([13],[37]).

In parallel with designing codes for meeting a secrecy rate, one also exploits techniques in the physical layer in order to enhance this rate. The first technique should be mentioned here is the multiple-input multiple-output (MIMO) technique using multiple antennas at transmitters, receivers and eavesdroppers. The advantage of this technique is to offer diversity gain and multiplexing gain which not only result in increased channel capacity but also bring chances for enhanced secrecy. Many research works ([89, 38, 74, 66, 39, 40]) have been proposed to analyse the secrecy capacity of MIMO systems. Initially, the simple scenarios of MIMO systems are studied. For instance, the 2-2-1 MIMO wiretap channel where both transmitter and receiver are equipped with two antennas and the eavesdropper has one antenna is mentioned in [89]. Another simple case is the 2-1-2 wiretap channel in which both transmitter and eavesdropper are equipped with two antennas and the recever is equipped with one antenna ([38]). These two special cases of MIMO systems are generalized later on into multiple-input (transmitter), multiple-output (receiver), single-eavesdropper (MIMOSE) model and multiple-input, single-output, multiple-eavesdropper (MISOME) model ([40]). The most general multiple-input multiple-output multiple-eavesdropper (MIMOME) system has been discussed in [74, 66, 40]. The general form of secrecy capacity is established ([74, 66, 40])) and the closed-form solution is given in some special cases ([63, 89]). In general cases, a global solution is obtained by a global optimization algorithm called brach-and-bound with reformulation and linearization technique (BB/RLT) ([65]).

Besides the multi-antenna mechanism, the cooperative transmission techniques recently used in the literature have contributed to the secrecy rate improvement (see [25] and the references therein). For example, cooperative jamming (CJ) techniques have been employed in some works ([39],[24]). Its nature is using artificial noise to confuse the eavesdroppers, thus limit the amount of information intercepted by them. Cooperative jamming can be performed by different parts of the network, which is

not necessarily to be intermediate or external nodes but the transmitter and receiver themselves. The use of cooperative jamming for secrecy rate was first discussed in [24] in a single antenna system. Afterwards, it is also applied to a lot of multi-antenna systems ([115, 118, 125, 60, 110, 6, 102]) to enhance the physical layer security. In addition, it is shown in some works ([109, 76, 108, 105, 16]) that the combination of this technique with other ones such as beamforming or power allocation brings considerable effects on secrecy. In addition, various channel state information (CSI) conditions, ranging from a complete lack of CSI to a perfect CSI are considered when applying this technique and it is indicated that this technique is particularly effective when CSI of eavesdroppers is unknown or partially known.

Another cooperative technique is cooperative relaying with two relaying protocols amplify-and-forward (AF) and decode-and-forward (DF) ([18], [34]). The role of relays is one of principle concerns in applying this technique. The relays may be untrusted nodes from which the transmitted message must be kept secret ([31], [34]). They can also play a role as traditional trusted relays simply to forward received information to destinations ([18], [23]). They are even regarded as both jamming and relaying devices to facilitate transmission as well as enhance the secure communications ([42]). Besides, relay selection is another important issue when multiple relays are available. An appropriate relay selection strategy might lead to a physical layer security improvement ([3], [67], [129]).

This dissertation focuses on the direction of system designs to improve the secrecy rate in wireless communication systems. More specifically, we concentrate on designing relay weights as well as allocating transmit power at sources and/or jammers to maximize achievable secrecy rate. In fact, when applying the aforementioned cooperative techniques for improving physical layer security, the very often derived optimization problems are to maximize the achievable secrecy rate under some power constraints in which their variables are often powers or relay weights. In general, these problem are nonconvex and thus hard to deal with. Only in some specific cases, the closed form of their optimal solution is given. For other general cases, there exists some methods in the literature to handle these nonconvex secrecy rate maximization problems. The widely-used approach is based on semidefinite relaxation technique in which these problems are first reformulated as semidefinite programs and the rank-one constraint is discarded ([124], [120], [108]). Afterwards, some randomization techniques are used to find the solution to the original problem. One simpler approach to seek a suboptimal solution is based on a null space scheme ([18], [127], [121]). In this scheme, some suitable constraints are added, which make the expression of secrecy rate is simplified, thus the secrecy rate maximization problem becomes easier to address. Recently, the efficient approach based on DC programming and DCA is employed in some works to tackle these problems ([5], [128], [106], [100], [119], [32], [82], [72], [30], [111], [2], [126], [75], [62], [20], [77]).

As mentioned in the introduction part, there are three chapters in this dissertation concerns physical layer security including this chapter and the next two ones. Various architectures of wireless communication systems are considered in these chapters and a variety of cooperative techniques are employed to enhance their secrecy. More

specifically, this chapter takes account of a point-to-point wireless network using cooperative jamming technique. The power allocation at the sources and friendly jammers needs to be computed such that the secrecy rate is maximized subject to some power constraints. Meanwhile, the next chapter (Chapter 4) studies a wireless multi-relay network including an eavesdropper and deploying jointly cooperative relaying and beamforming techniques. Two relaying protocols, AF and DF, are mentioned in this chapter and the purpose is to find beamforming coefficients to maximize secrecy rate under some power constraints. Chapter 5 also considers a wireless multi-relay network but in the presence of multiple eavesdroppers. The combination of AF cooperative relaying and CJ technique respectively with beamforming technique are used to improve secrecy. The aim of this chapter is similar to that of Chapter 4, but the appearance of multiple eavesdroppers makes the secrecy rate maximization problem more difficult to deal with.
In all these three chapters, the tools of DC programming and DCA are developed to solve the considered problems. We design efficient standard DCAs for addressing the problems in the two first chapters. Especially, the highly efficient distributed DCA scheme in Chapter 3 is a good approach for handling the problems arising from multi-user wireless communication systems. In Chapter 5, we propose the general DCA schemes, which is a new approach in DC programming and only studied and applied recently. The convergence of such general DCA schemes are shown.
In the followings, we present the first problem in this dissertation, which is related to physical layer security.

3.2 Related Works and Contributions

In this chapter, we reconsider the optimization model that was introduced in [5]. The cooperative jamming technique was employed in this model in order to enhance secrecy capacity of a point-to-point wireless communication system comprising multiple pairs of user and single eavesdropper. The purpose of this model is to allocate transmit power at the sources of the users and at the friendly jammers in order to maximize the sum of secrecy rates under some power constraints. This raises a nonconvex, nonsmooth optimization problem, which is hard to solve. This model was addressed in [5] by the DCA based algorithm named SCA. This SCA scheme was based on decomposing the objective function into the difference of two convex logarithm functions and was implemented in both centralized and distributed ways. Numerical results in [5] have showed the efficiency of SCA through the comparison of its results with two standard softwares for nonconvex programming which are MINOS solver [73] and PSwarm [101].

As mentioned in Chapter 1, each DC function has an infinite number of DC decompositions which have crucial impacts on the qualities of DCA such as speed of convergence, robustness, efficiency, globality of computed solutions, etc, thus the search for a "good" DC decomposition is vital from algorithmic point of views. Observing that the flexibility of DCA according to the choice of DC decomposition is a crucial point to design efficient DCA based algorithms, we propose in this chapter a new DC decomposition

for the objective function of the above model and develop an efficient DCA scheme for solving it.

Our contributions are twofold.

Firstly, we exploit the special structure of the objective function in a suitable way to propose the new DC decomposition. The resulting DCA involves the convex subproblem which is a strongly convex quadratic program, thus it can be efficiently solved in a centralized way by standard softwares. Furthermore, the convex quadratic objective function is separate in its variables, and some constraints are also separate on those variables. These nice properties facilitate the use of distributed algorithms. The distributed method is regarded as an effective tool to deal with large-scale optimization problem often encountered in communication systems, because it permits to divide such large-scale problems into smaller-scale ones.

Secondly, one of the major contributions of the chapter is to develop a highly efficient distributed dual based gradient projection algorithm to solve the convex subproblem in the DCA scheme by exploring and exploiting the special structure of this problem in a deep and efficient way. It turns out that our distributed DCA scheme requires computing iteratively the projection of points onto the intersection of a box and a half space which can be determined in a very inexpensive way. This significantly increases the speed of the proposed distributed dual algorithm, thereby sharply reducing the runtime of the distributed DCA scheme. The numerical experiments show that our distributed DCA is far faster than the existing distributed SCA, the ratio of gain can up to 970 times.

The rest of this chapter is organized as follows. In Section 3.3, we describe the considered secrecy rate maximization problem and its optimization model. The numerical solution method is studied in Section 3.4, where we show how to apply DCA to solve the considered problem. Numerical experiments are reported in Section 3.5. At last, Section 3.6 concludes the chapter.

3.3 Secrecy Rate Maximization via Cooperative Jamming

In this section, we briefly state the optimization problem formulated in [5]. Consider a wireless communication system comprised of Q pairs of transmitter and receiver-the legitimate users, J friendly jammers, and a single eavesdropper. OFDMA (Orthogonal Frequency Division Multiple Access) transmissions are assumed for the authorized users over flat-fading and quasi-static channels. H_{qq}^{SD}, H_{jq}^{JD}, H_{je}^{JE}, H_{qe}^{SE} are respectively denoted as the channel coefficients of the channel between the source and the destination of the qth legitimate user, the transmitter of the jth jammer and the destination of the qth user, the transmitter of the jth jammer and the eavesdropper, the source of the qth user and the eavesdropper. It is assumed that the perfect CSI is available on the eavesdropper's channels.

We follow the cooperative jamming (CJ) paradigm, in which the friendly jammers and the users cooperate together to provide interference with the aim of confounding the eavesdropper. Denote p_q as the power allocation of source q; p_{jq}^J as power allocation of friendly jammer j over the channel used by user q. $\mathbf{p}_q^J \triangleq (p_{jq}^J)_{j=1,\dots,J}$ is the vector of powers allocated by all the jammers over the channel of user q. Denote $\mathbf{p} = (p_q)_{q=1,\dots,Q}, \mathbf{p}^J = (\mathbf{p}_q^J)_{q=1,\dots,Q}$. The power of the user q and the jammer j do not exceed C_q and C_j^J, respectively.
The maximum achievable rate on the channel of the user q is calculated by

$$r_{qq}(p_q, \mathbf{p}_q^J) \triangleq \log\left(1 + \frac{H_{qq}^{SD} p_q}{\sigma^2 + \sum_{j=1}^J H_{jq}^{JD} p_{jq}^J}\right). \tag{3.1}$$

In a similar way, the maximum achievable rate on the link between the source q and the eavesdropper is given by

$$r_{qe}(p_q, \mathbf{p}_q^J) \triangleq \log\left(1 + \frac{H_{qe}^{SE} p_q}{\sigma^2 + \sum_{j=1}^J H_{je}^{JE} p_{jq}^J}\right). \tag{3.2}$$

The secrecy rate of the user q is defined by ([35])

$$\max\{0, r_{qq}(p_q, \mathbf{p}_q^J) - r_{qe}(p_q, \mathbf{p}_q^J)\}. \tag{3.3}$$

Problem Formulation: The purpose is to find an effectively cooperative strategy between the legitimate users and the jammers to maximize the system secrecy rate. More particularly, each user q together with the jammers try to search the tuple $(p_q, \mathbf{p}_q^J)$ satisfying the optimization problem below:

$$\begin{aligned} \max_{(\mathbf{p},\mathbf{p}^J)\geq 0} \quad & r(\mathbf{p}, \mathbf{p}^J) \triangleq \sum_{q=1}^Q \max\{0, r_{qq}(p_q, \mathbf{p}_q^J) - r_{qe}(p_q, \mathbf{p}_q^J)\} \\ \text{s.t.} \quad & p_q \leq P_q, \ \forall q = 1, \dots, Q, \\ & \sum_{r=1}^Q p_{jr}^J \leq P_j^J, \ \forall j = 1, \dots, J. \end{aligned}$$

Note that the inequality $r_{qq}(p_q, \mathbf{p}_q^J) \geq r_{qe}(p_q, \mathbf{p}_q^J)$ is equivalent to

$$p_q = 0 \text{ or } \frac{H_{qq}^{SD}}{\sigma^2 + \sum_{j=1}^J H_{jq}^{JD} p_{jq}^J} \geq \frac{H_{qe}^{SE}}{\sigma^2 + \sum_{j=1}^J H_{je}^{JE} p_{jq}^J}$$

or again

$$p_q = 0 \text{ or } \sum_{j=1}^J (H_{qq}^{SD} H_{je}^{JE} - H_{qe}^{SE} H_{jq}^{JD}) p_{jq}^J + (H_{qq}^{SD} - H_{qe}^{SE})\sigma^2 \geq 0. \tag{3.4}$$

If the inequality in (3.4) is violated for some profiles $(p_q, \mathbf{p}_q^J)$, then the secrecy rate of the qth user equals zero. Such profiles are insignificant because the users want to

maximize their secrecy rate. Therefore, we can eliminate the feasible users' strategy profiles not satisfying the inequality in (3.4). This leads us to solving the following problem:

$$\begin{aligned} \min_{(\mathbf{p},\mathbf{p}^J)\geq 0} \quad & r_1(\mathbf{p},\mathbf{p}^J) \triangleq \sum_{q=1}^{Q}[-r_{qq}(p_q,\mathbf{p}_q^J)+r_{qe}(p_q,\mathbf{p}_q^J)] \\ \text{s.t.} \quad & p_q \leq C_q,\ \forall q=1,\ldots,Q, \\ & \sum_{r=1}^{Q} p_{jr}^J \leq C_j^J,\ \forall j=1,\ldots,J, \\ & \sum_{j=1}^{J} a_{jq}p_{jq}^J \geq b_q,\ \forall q=1,\ldots,Q \end{aligned} \tag{3.5}$$

where $b_q=-(H_{qq}^{SD}-H_{qe}^{SE})\sigma^2$, $a_{jq}=(H_{qq}^{SD}H_{je}^{JE}-H_{qe}^{SE}H_{jq}^{JD})$.
We will investigate DC programming and DCA for solving this nonconvex program.

3.4 Solution Methods Based on DC Programming and DCA

3.4.1 The new DC decomposition for the objective function of (3.5)

For convenience, we denote $\mathbf{x}=(\mathbf{p},\mathbf{p}^J), \mathbf{x}_q=(p_q,\mathbf{p}_q^J),\ \forall q=1,\ldots,Q$. For any value of ρ, the objective function of the problem (3.5) can be written in the form:

$$r_1(\mathbf{x}) = G(\mathbf{x}) - H(\mathbf{x}),$$

where

$$G(\mathbf{x}) = \frac{\rho}{2}\|\mathbf{x}\|^2 \quad \text{and} \quad H(\mathbf{x}) = \frac{\rho}{2}\|\mathbf{x}\|^2 - \sum_{q=1}^{Q}\left(-r_{qq}(\mathbf{x}_q)+r_{qe}(\mathbf{x}_q)\right).$$

It is noted that for any $\rho>0$, the function G is convex. We aim to determine $\rho>0$ such that H is also convex. It has been seen in [53] that if r_1 is a smooth function with Lipschitz continuous gradient then ρ is nothing but the Lipschitz constant of ∇r_1.

The proposition below shows that the gradient of r_1 is Lipschitz continuous and its Lipschitz constant can be calculated.

Proposition 3.1. *The function r_1 is smooth and its gradient is Lipschitz continuous with constant $\rho_0=\frac{2M^2}{\sigma^4}\sqrt{1+2J+4J^2}$, where*

$$M = \max_{\substack{q=1,\ldots,Q \\ j=1,\ldots,J}} \left\{H_{qe}^{SE}, H_{qq}^{SD}, H_{jq}^{JD}, H_{je}^{JE}\right\}.$$

Proof. Obviously the function $r_1(\mathbf{x})$ is differentiable, and its gradient is given by

$$\nabla r_1(\mathbf{x}) = \left[\frac{\partial r_1}{\partial \mathbf{p}}(\mathbf{x}) = \left(\frac{\partial r_1}{\partial p_q}\right)_{q=1,\ldots,Q}, \frac{\partial r_1}{\partial \mathbf{p}^J}(\mathbf{x}) = \left(\frac{\partial r_1}{\partial p_{jq}^J}\right)_{\substack{j=1,\ldots,J \\ q=1,\ldots,Q}}\right]^T,$$

where

$$\begin{aligned}
\frac{\partial r_1}{\partial p_q} &= \frac{H_{qe}^{SE}}{\sigma^2 + H_{qe}^{SE} p_q + A} - \frac{H_{qq}^{SD}}{\sigma^2 + H_{qq}^{SD} p_q + B}, \\
\frac{\partial r_1}{\partial p_{jq}^J} &= \frac{H_{je}^{JE}}{\sigma^2 + H_{qe}^{SE} p_q + A} - \frac{H_{jq}^{JD}}{\sigma^2 + H_{qq}^{SD} p_q + B} - \frac{H_{je}^{JE}}{\sigma^2 + A} + \frac{H_{jq}^{JD}}{\sigma^2 + B},
\end{aligned}$$

with $A = \sum_{k=1}^{J} H_{ke}^{JE} p_{kq}^J$, $B = \sum_{k=1}^{J} H_{kq}^{JD} p_{kq}^J$.
We have

$$\|\nabla r_1(\mathbf{x}) - \nabla r_1(\widehat{\mathbf{x}})\|^2 = \|\frac{\partial r_1}{\partial \mathbf{p}}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}}(\widehat{\mathbf{x}})\|^2 + \|\frac{\partial r_1}{\partial \mathbf{p}^J}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}^J}(\widehat{\mathbf{x}})\|^2. \tag{3.6}$$

Firstly, we evaluate the first term of (3.6).

$$\begin{aligned}
&\|\frac{\partial r_1}{\partial \mathbf{p}}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}}(\widehat{\mathbf{x}})\|^2 \\
&= \sum_{q=1}^{Q} \left(\frac{H_{qe}^{SE}}{\sigma^2 + H_{qe}^{SE} p_q + A} - \frac{H_{qe}^{SE}}{\sigma^2 + H_{qe}^{SE} \widehat{p}_q + A_1} + \frac{H_{qq}^{SD}}{\sigma^2 + H_{qq}^{SD} \widehat{p}_q + B_1} \right. \\
&\qquad \left. - \frac{H_{qq}^{SD}}{\sigma^2 + H_{qq}^{SD} p_q + B} \right)^2 \\
&= \sum_{q=1}^{Q} \left(\frac{H_{qe}^{SE^2}(\widehat{p}_q - p_q) + \sum_{j=1}^{J} H_{qe}^{SE} H_{je}^{JE} (\widehat{p}_{jq}^J - p_{jq}^J)}{(\sigma^2 + H_{qe}^{SE} p_q + A)(\sigma^2 + H_{qe}^{SE} \widehat{p}_q + A_1)} \right. \\
&\qquad \left. + \frac{H_{qe}^{SD^2}(p_q - \widehat{p}_q) + \sum_{j=1}^{J} H_{qq}^{SD} H_{jq}^{JD} (p_{jq}^J - \widehat{p}_{jq}^J)}{(\sigma^2 + H_{qe}^{SD} p_q + B)(\sigma^2 + H_{qq}^{SD} \widehat{p}_q + B_1)} \right)^2,
\end{aligned}$$

Applying the Cauchy-Schwartz inequality and noting that all denominators of the fractions above are greater than σ^2 and $M = \max\limits_{\substack{q=1,\ldots,Q \\ j=1,\ldots,J}} \left\{ H_{qe}^{SE}, H_{qq}^{SD}, H_{jq}^{JD}, H_{je}^{JE} \right\}$, we obtain

$$\|\frac{\partial r_1}{\partial \mathbf{p}}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}}(\widehat{\mathbf{x}})\|^2 \leq \frac{4M^4}{\sigma^8}(1 + J)\|\mathbf{x} - \widehat{\mathbf{x}}\|^2. \tag{3.7}$$

In the following, we evaluate the second term of (3.6)

$$\begin{aligned}
&\|\frac{\partial r_1}{\partial \mathbf{p}^J}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}^J}(\widehat{\mathbf{x}})\|^2 \\
&= \sum_{j=1}^{J}\sum_{q=1}^{Q}\left[\frac{H_{je}^{JE}}{\sigma^2 + H_{qe}^{SE}p_q + A} - \frac{H_{je}^{JE}}{\sigma^2 + H_{qe}^{SE}p_q + A_1} + \frac{H_{je}^{JE}}{\sigma^2 + A_1} - \frac{H_{je}^{JE}}{\sigma^2 + A} + \right. \\
&\left. + \frac{H_{jq}^{JD}}{\sigma^2 + H_{qq}^{SD}\widehat{p}_q + B_1} - \frac{H_{jq}^{JD}}{\sigma^2 + H_{qq}^{SD}p_q + B} + \frac{H_{jq}^{JD}}{\sigma^2 + B} - \frac{H_{jq}^{JD}}{\sigma^2 + B_1}\right]^2 \\
&= \sum_{j=1}^{J}\sum_{q=1}^{Q}\left[\left(\frac{H_{je}^{JE}H_{qe}^{SE}}{M_1M_2} - \frac{H_{jq}^{JD}H_{qq}^{SD}}{M_3M_4}\right)(\widehat{p}_q - p_q)\right. \\
&\left. + \sum_{k=1}^{J}\left(\frac{H_{je}^{JE}H_{ke}^{JE}}{M_1M_2} - \frac{H_{je}^{JE}H_{ke}^{JE}}{M_3M_4} - \frac{H_{jq}^{JD}H_{kq}^{JD}}{M_5M_6} + \frac{H_{jq}^{JD}H_{kq}^{JD}}{M_7M_8}\right)(\widehat{p}_{kq}^J - p_{kq}^J)\right]^2.
\end{aligned}$$

where $M_i, (i = 1, ..., 8)$ are the denominators of the fractions in the above expression, respectively. Applying the Cauchy-Schwarz inequality we also gain

$$\|\frac{\partial r_1}{\partial \mathbf{p}^J}(\mathbf{x}) - \frac{\partial r_1}{\partial \mathbf{p}^J}(\widehat{\mathbf{x}})\|^2 \leq \frac{4M^4}{\sigma^8}J(1+4J)\|\mathbf{x} - \widehat{\mathbf{x}}\|^2. \tag{3.8}$$

Adding (3.7) to (3.8) and then taking the square root of both sides, we obtain

$$\|\nabla r_1(\mathbf{x}) - \nabla r_1(\widehat{\mathbf{x}})\| \leq \frac{2M^2}{\sigma^4}\sqrt{1+2J+4J^2}\|\mathbf{x} - \widehat{\mathbf{x}}\|.$$

This inequality implies that the gradient of r_1 is Lipschitz continuous with the constant $\rho_0 = \frac{2M^2}{\sigma^4}\sqrt{1+2J+4J^2}$. □

In summary, with $\rho = \rho_0 = \frac{2M^2}{\sigma^4}\sqrt{1+2J+4J^2}$ both $G(x)$ and $H(x)$ are convex functions and we obtain the following DC formulation of (3.5).

$$\begin{aligned}
\min_{(\mathbf{p},\mathbf{p}^J)\geq 0} \quad & r_1(\mathbf{p},\mathbf{p}^J) \triangleq G(\mathbf{p},\mathbf{p}^J) - H(\mathbf{p},\mathbf{p}^J) \\
\text{s.t} \quad & p_q \leq C_q,\ \forall q = 1,\ldots,Q, \\
& \sum_{r=1}^{Q} p_{jr}^J \leq C_j^J,\ \forall j = 1,\ldots,J, \\
& \sum_{j=1}^{J} a_{jq}p_{jq}^J \geq b_q,\ \forall q = 1,\ldots,Q
\end{aligned} \tag{3.9}$$

In the remainder of this section we will show how to solve this problem (with $\rho = \rho_0$) by DCA. According to the generic DCA scheme described above, DCA applied to (3.9) consists of computing the two sequences $\{\mathbf{y}^k\}$ and $\{\mathbf{x}^k\}$ such that

$$\mathbf{y}^k = (\bar{\mathbf{p}}^k, \bar{\mathbf{p}}^{J,k}) \in \partial H(\mathbf{x}^k),$$

and $\mathbf{x}^{k+1} = (\mathbf{p}^{k+1}, \mathbf{p}^{J,k+1})$ solves the convex problem

$$\min_{(\mathbf{p},\mathbf{p}^J)} \quad \frac{\rho}{2}\left(\|\mathbf{p}\|^2 + \sum_{q=1}^{Q}\|\mathbf{p}_q^J\|^2\right) - \langle \bar{\mathbf{p}}^k, \mathbf{p}\rangle - \sum_{q=1}^{Q}\langle \bar{\mathbf{p}}_q^{J,k}, \mathbf{p}_q^J\rangle \tag{3.10}$$

$$\text{s.t.} \quad (\mathbf{p}, \mathbf{p}^J) \geq 0, \tag{3.11}$$

$$p_q \leq C_q, \ \forall q = 1, \ldots, Q, \tag{3.12}$$

$$\sum_{j=1}^{J} a_{jq} p_{jq}^J \geq b_q, \ \forall q = 1, \ldots, Q, \tag{3.13}$$

$$\sum_{q=1}^{Q} p_{jq}^J \leq C_j^J, \ \forall j = 1, \ldots, J. \tag{3.14}$$

Clearly, the function H is differentiable and its gradient is computed as

$$\nabla H(\mathbf{x}) = \begin{bmatrix} \left(\rho p_q - \frac{H_{qe}^{SE}}{\sigma^2 + H_{qe}^{SE} p_q + A} + \frac{H_{qq}^{SD}}{\sigma^2 + H_{qq}^{SD} p_q + B}\right)_{q=1,\ldots,Q} \\ \left(\rho p_{jq}^J - \frac{H_{je}^{JE}}{\sigma^2 + H_{qe}^{SE} p_q + A} + \frac{H_{je}^{JE}}{\sigma^2 + A} + \frac{H_{jq}^{JD}}{\sigma^2 + H_{qq}^{SD} p_q + B} - \frac{H_{jq}^{JD}}{\sigma^2 + B}\right)_{\substack{j=1,\ldots,J \\ q=1,\ldots,Q}} \end{bmatrix} \tag{3.15}$$

with $A = \sum_{k=1}^{J} H_{ke}^{JE} p_{kq}^J$ and $B = \sum_{k=1}^{J} H_{kq}^{JD} p_{kq}^J$. We will discuss below the solution methods for solving the convex problem (3.10) - (3.14).

3.4.2 Solving the convex subproblem in the DCA scheme

The subproblem (3.10) - (3.14) is a linearly constrained quadratic program for which several standard softwares are available, for example the CPLEX [cpl]. Meanwhile, it is noted that the objective function of this problem is separate in its variables and the feasible set is a special polytope defined by the separate constraints (3.11) - (3.13) and the coupling constraints (3.14). Exploiting the special structure of this problem we propose a distributed dual decomposition method for solving it. The distributed method is considered as one of approaches for overcoming the large-scale setting, the most challenging issue of the optimization problems in communication systems. However, many problems cannot be solved in a distributed way due to either the inseparability of objective function or the presence of coupling constraints.

For facing out the coupling constraints in the problem (3.10) - (3.14), we solve its dual problem by the gradient projection method [58] in which the partial Lagrangian duality related to the coupling constraints is considered. Hence the inner problem in the dual algorithm can be decomposed into some problems with smaller size and so it can be solved in a distributed way.

More precisely, we first form the partial Lagrangian related to (3.14) as below

$$\begin{aligned} L(\mathbf{x},\lambda) &= \frac{\rho}{2}\|\mathbf{x}\|^2 - \langle \mathbf{y}^k, \mathbf{x}\rangle + \sum_{j=1}^{J}\lambda_j\Big(\sum_{q=1}^{Q} p_{jq}^J - C_j^J\Big) \\ &= \sum_{q=1}^{Q}\Big(\frac{\rho}{2}\|\mathbf{x}_q\|^2 + \langle \mathbf{z}_q^k, \mathbf{x}_q\rangle\Big) - \sum_{j=1}^{J}\lambda_j C_j^J \\ &= \sum_{q=1}^{Q} L_q(\mathbf{x}_q,\lambda) - \sum_{j=1}^{J}\lambda_j C_j^J, \end{aligned}$$

where

$$\mathbf{x}_q = (p_q, \mathbf{p}_q^J),\ \mathbf{z}_q^k = (-\overline{p}_q^k, \lambda - \overline{\mathbf{p}}_q^{J,k}),\ L_q(\mathbf{x}_q,\lambda) = \frac{\rho}{2}\|\mathbf{x}_q\|^2 + \langle \mathbf{z}_q^k, \mathbf{x}_q\rangle.$$

The dual problem associated with (3.10) - (3.14) is then

$$\max_{\lambda\geq 0}\left\{\gamma(\lambda) = \min_{\mathbf{x}\in S} L(\mathbf{x},\lambda)\right\} \tag{3.16}$$

with $S = \prod_{q=1}^{Q} S_q$, $S_q = \left\{ \mathbf{x}_q = (p_q, \mathbf{p}_q^J) : \begin{array}{l} 0 \leq p_q \leq C_q, \\ 0 \leq p_{jq}^J \leq C_j^J,\ \forall j = 1,\ldots,J \\ \sum_{j=1}^{J} a_{jq} p_{jq}^J \geq b_q \end{array} \right\}$. As the function L is affine w.r.t λ, it is obviously that the function γ is concave on $\mathbb{R}_+^J$, or equivalently $-\gamma$ is convex, therefore (3.16) is a convex program. Moreover, we show below that $-\gamma$ is differentiable. Indeed, from the definition of γ we have

$$-\gamma(\lambda) = \max_{\mathbf{x}\in S} -L(\mathbf{x},\lambda). \tag{3.17}$$

A subgradient of the convex function $-\gamma$ is computed following Theorem 4.4.2 in [33] (note that S is a compact set)

$$\partial(-\gamma(\lambda)) = \mathrm{co}\{\cup\partial_\lambda(-L(\mathbf{x},\lambda,)) : \mathbf{x}\in I(\lambda)\}, \tag{3.18}$$

where $\mathrm{co}(\cdot)$ denotes the convex hull of $(\cdot)$ and

$$I(\lambda) = \{\mathbf{x}(\lambda)\in S : -\gamma(\lambda) = -L(\mathbf{x},\lambda)\} = \{\mathbf{x}(\lambda)\in S : \mathbf{x}(\lambda) \text{ solves (3.17)}\}.$$

It is easy to see that the problem (3.17) is equivalent to

$$\min\left\{\sum_{q=1}^{Q} L_q(\mathbf{x}_q,\lambda) : \mathbf{x}\in\prod_{q=1}^{Q} S_q\right\}. \tag{3.19}$$

This problem has the unique solution $\widehat{\mathbf{x}}(\lambda) \triangleq (\widehat{\mathbf{x}}_q(\lambda))_{q=1,\ldots,Q}$, where each $\widehat{\mathbf{x}}_q(\lambda)$ $(q = 1,\ldots,Q)$ is the unique solution of the problem

$$\min\left\{L_q(\mathbf{x}_q,\lambda) = \frac{\rho}{2}\|\mathbf{x}_q\|^2 + \langle \mathbf{z}_q^k, \mathbf{x}_q\rangle : \mathbf{x}_q\in S_q\right\}. \tag{3.20}$$

Since (3.19) has the unique solution, according to (3.18) the function $-\gamma$ is differentiable on $\mathbb{R}_+^J$ and its gradient is calculated by (Theorem 4.4.2 in ([33])

$$\nabla_\lambda \left(-\gamma(\lambda)\right) = -\left(\sum_{q=1}^{Q} \widehat{\mathbf{p}}_q^J - \mathbf{C}^J\right), \text{ with } \mathbf{C}^J = (C_j^J)_{j=1,\ldots,J}.$$

We are now in a position to apply the gradient projection algorithm [58] on the dual problem (3.16) with the aim of solving the convex subproblem (3.10). As indicated above, the problem (3.17) can be decomposed into Q problems of the form (3.19), hence the distributed method can be used. According to the general scheme described in Appendix A.1, the distributed dual-decomposition based algorithm for solving the convex subproblem (3.10) computes iteratively, for $t = 0, 1, 2, ...$

$$\widehat{\mathbf{x}}_q(\lambda^t) = (\widehat{p}_q, \widehat{\mathbf{p}}_q^{J,t}) \triangleq \arg\min_{\mathbf{x}_q \in S_q} L_q(\mathbf{x}_q, \lambda^t) \ \ \forall q = 1, \ldots, Q, \tag{3.21}$$

$$\lambda^{t+1} \triangleq \left[\lambda^t + \alpha^t \left(\sum_{q=1}^{Q} \widehat{\mathbf{p}}_q^{J,t} - \mathbf{C}^J\right)\right]_+. \tag{3.22}$$

The major step of the distributed dual-decomposition based algorithm consists of solving Q convex problems of the form (3.20) which is nothing else the computation of the projection of $-\frac{\mathbf{z}_q^k}{\rho}$ onto S_q. Exploiting the quite simple structure of S_q, say

$$S_q = \Omega_q \cap [\mathbf{0}, \mathbf{C}] \text{ with } \Omega_q \triangleq \{(p_q, \mathbf{p}_q^J) : \sum_{j=1}^{J} a_{jq} p_{jq}^J \geq b_q\}, \ \mathbf{C} \triangleq (C_q, \mathbf{C}^J),$$

we investigate a very inexpensive procedure for (3.20) as shown below.
Denote by $\text{Proj}_{(\cdot)}(x)$ the projection of the point x onto the set $(\cdot)$. First of all, find $\bar{\mathbf{x}}$=$\text{Proj}_{[0,\mathbf{C}]}(-\frac{\mathbf{z}_q^k}{\rho})$ which is explicitly computed via the simple formula

$$\bar{\mathbf{x}}_j = 0 \text{ if } \mathbf{z}_{q,j}^k > 0, \ \mathbf{C}_j \text{ if } \mathbf{z}_{q,j}^k < -\rho \mathbf{C}_j, \ -\mathbf{z}_{q,j}^k/\rho \text{ otherwise, for } j = 1, \ldots, J+1.$$

If $\bar{\mathbf{x}} \in \Omega_q$, then we have immediately $\bar{\mathbf{x}} = \text{Proj}_{S_q}\left(-\mathbf{z}_q^k/\rho\right) \triangleq \widehat{\mathbf{x}}_q(\lambda)$. Otherwise, the lemma below shows that $\widehat{\mathbf{x}}_q(\lambda)$ satisfies the equation $\sum_{j=1}^{J} a_{jq} p_{jq}^J = b_q$, i.e., $\widehat{\mathbf{x}}_q(\lambda)$ belongs to the hyperplane $\overline{\Omega}_q \triangleq \{(p_q, \mathbf{p}_q^J) : \sum_{j=1}^{J} a_{jq} p_{jq}^J = b_q\}$ that defines the half space Ω_q. In such a case we have

$$\widehat{\mathbf{x}}_q(\lambda) \triangleq \text{Proj}_{S_q}\left(-\mathbf{z}_q^k/\rho\right) = \text{Proj}_{\overline{\Omega}_q \cap [\mathbf{0},\mathbf{C}]}\left(-\mathbf{z}_q^k/\rho\right).$$

Hence, in the case where $\bar{\mathbf{x}} \notin \Omega_q$, to compute $\widehat{\mathbf{x}}_q(\lambda)$ we can adapt an inexpensive procedure proposed in [69], called **BoxProjection**, for projecting a vector on the intersection of a hyperplane and a box in $\mathbb{R}^n$ (see Appendix A.3).

Lemma 3.1. *Given $x_0 \in \mathbb{R}^n$ and $\mathbb{P} = \{x \in [a, b] \subset \mathbb{R}^n : c^T x + d \leq 0\}$. Let y_0 and y_1 be respectively the projection of x_0 onto $\mathbb{P}$ and $[a, b]$, i.e., $y_0 = \arg\min_{x \in P} \|x - x_0\|^2$ and $y_1 = \arg\min_{x \in [a,b]} \|x - x_0\|^2$. Assume that $c^T y_1 + d > 0$. Then we have $c^T y_0 + d = 0$.*

Proof. It is obvious that $c^T y_0 + d \leq 0$. Suppose that $c^T y_0 + d < 0$, then there exists $\delta > 0$ small enough such that

$$c^T x + d < 0 \quad \forall x \in B(y_0, \delta) := \{x : \|x - y_0\| < \delta\}.$$

Since $y_0 = \arg\min_{x \in P} \|x - x_0\|^2$, we have $\|y_0 - x_0\|^2 \leq \|x - x_0\|^2$ for all $x \in [a, b] \cap B(y_0, \delta)$. Thus, y_0 is a local minimizer of the strongly convex problem $\min_{x \in [a,b]} \|x - x_0\|^2$. As this problem admits the unique (global and local) minimizer which is nothing that y_1, we have $y_0 = y_1$. Hence $c^T y_0 + d = c^T y_1 + d > 0$. This contradicts the assumption $c^T y_0 + d < 0$. The proof is then complete. □

To terminate this subsection 3.4.2, let us mention the convergence theorem of the distributed dual-decomposition based algorithm.

Theorem 3.1. *The sequence $\{\lambda^t\}$ generated by the distributed dual-decomposition algorithm converges to a solution of (3.16) and the sequence $\{\widehat{\boldsymbol{x}}(\lambda^t)\}$ converges to the unique solution of (3.10).*

Proof. Because the objective function of problem (3.10) is strongly convex, its gradient satisfies a Lipshitz condition and its feasible set is a compact convex one, the convergence of these two sequences are deduced from Theorem A.1 in Appendix A.1. □

Now, we are able to describe all the steps of the DCA schemes for solving the DC program (3.9).

3.4.3 DCA scheme for solving the DC program (3.9)

In what below DCAC (resp. DCAD) stands for the DCA scheme in which the subproblem (3.10) is solved directly in a centralized way by a standard software (resp. by the distributed dual-decomposition method). The DCAD scheme is described in the algorithm below.

Algorithm DCAD: The distributed DCA scheme for solving (3.9)

1: **initialization**: Let $\epsilon_1, \epsilon_2 > 0$ be tolerances. Choose an initial point $\mathbf{x}^0 = (\mathbf{p}^0, \mathbf{p}^{J,0}), k \leftarrow 0$.
2: **repeat**
3: **Step 1**: For each k, compute $\mathbf{y}^k = (\bar{\mathbf{p}}^k, \bar{\mathbf{p}}^{J,k}) = \nabla H(\mathbf{x}^k)$ via (3.15).
4: **Step 2**: {Compute $\mathbf{x}^{k+1} = (\mathbf{p}^{k+1}, \mathbf{p}^{J,k+1})$, an optimal solution of (3.10) using the distributed dual-decomposition method }
5: **initialization:** Choose an initial point $\lambda^0 \geq 0$ and a positive sequence $\{\alpha^t\}$ satisfying $\sum_t \alpha^t = \infty, \sum_t (\alpha^t)^2 < \infty$, $\alpha^t \longrightarrow 0$; $t \leftarrow 0$.
6: **repeat**

Step 2.1: {Successively solve (3.20) for $q = 1, ..., Q$ to obtain $\widehat{\mathbf{x}} = (\widehat{p}_q^t, \widehat{\mathbf{p}}_q^{J,t})_{q=1,...,Q}$ }
for $q = 1, \ldots, Q$ **do**
Compute $\mathbf{z}_q^k = \left(-\overline{p}_q^k, \lambda^t - \overline{\mathbf{p}}_q^{J,k}\right)$
Find $\bar{\mathbf{x}} = \text{Proj}_{[0,\mathbf{C}]}\left(-\mathbf{z}_q^k/\rho\right) = \max\left\{0, \min\left\{-\mathbf{z}_q^k/\rho, \mathbf{C}\right\}\right\}$
if $\bar{\mathbf{x}} \in (\Omega_q)$ **then**
Set $(\widehat{p}_q^t, \widehat{\mathbf{p}}_q^{J,t}) = \bar{\mathbf{x}}$
else
Compute $(\widehat{p}_q^t, \widehat{\mathbf{p}}_q^{J,t})$ by applying the **BoxProjection** algorithm
end if
end for
Step 2.2: Update λ by the formula $\lambda^{t+1} \triangleq \left[\lambda^t + \alpha^t\left(\sum_{q=1}^{Q} \widehat{\mathbf{p}}_q^{J,t} - \mathbf{C}^J\right)\right]_+$
Step 2.3: $t \leftarrow t + 1$
until $\|\lambda^{t+1} - \lambda^t\| < \epsilon_2$.
Set $\mathbf{x}^{k+1} = \widehat{\mathbf{x}} = (\widehat{p}_q^t, \widehat{\mathbf{p}}_q^{J,t})_{q=1,...,Q}$
Step 3: $k \leftarrow k + 1$
until either $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| < \epsilon_1\|\mathbf{x}^k\|$ or $|r_1(\mathbf{x}^{k+1}) - r_1(\mathbf{x}^k)| < \epsilon_1|r_1(\mathbf{x}^k)|$

DCAC differs from DCAD by the step 2: instead of applying the distributed dual-decomposition method we use a standard software for solving the convex quadratic program (3.10). In our experiment we use the CPLEX software [cpl].

Theorem 3.2. *(convergence properties of DCAC and DCAD)*

(i) The sequence $\{G(\mathbf{x}^k) - H(\mathbf{x}^k)\}$ *is monotonously decreasing.*

(ii) Every limit points $\mathbf{x}^*$ *of the sequence* $\{\mathbf{x}^k\}$ *are critical points of the problem (3.9), and more strongly, they verify the necessary local optimality condition* $\partial H(\mathbf{x}^*) \subset \partial G(\mathbf{x}^*)$.

(iii) The series $\{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2\}$ *converges.*

Proof. Immediate consequences of the convergence properties of the generic DCA and the facts that H is differentiable and G is strongly convex. □

3.5 Computational Experiments

3.5.1 Datasets and experimental setups

All algorithms were implemented in the Visual Studio 2012, and performed on a PC Intel Core i5-2500S CPU 2.70GHz of 4GB RAM. The channel coefficients $H_{qq}^{SD}, H_{qe}^{SE}, H_{jq}^{JD}, H_{je}^{JE}$ are randomly generated in the same way as the one described in [5]. Only the channel coefficients satisfying the condition (3.13) are used to perform the algorithms (to ensure that the feasible set of test problems are nonempty).
The power budgets of all the users and jammer are equal, i.e., $C_q = C_j^J = P\ \forall j =$

$1, ..., J,\ q = 1, ..., Q$. The number of jammers J, the number of users Q and the value of $\texttt{snr}$ are set differently in the different experiments. More specifically, in the experiment 1 and the experiment 4, $\texttt{snr} = 10$ while Q varies in the set $\{10, 20, 30, 40, 50\}$, and $J = \lfloor Q/2 \rfloor$. In the experiment 2, we set $Q = 10$, $J = 5$ and $\texttt{snr}$ varies in the set $\{5, 10, 15\}$. In the experiment 3, we set $Q = 10, \texttt{snr} = 10$ while J is chosen from the set $\{2, 4, 6, 8\}$.

In all the algorithms, the initial points are set to zeros. The initial point $\{\lambda^0\}$ of the dual problem is also set to zero. To update λ after each inner iteration, we choose the sequence $\{\alpha^t = \frac{1}{t}\}$. The inner loop is terminated with the tolerance $\epsilon_2 = 10^{-2}$ while the DCA scheme is stopped with the tolerance $\epsilon_1 = 10^{-5}$.

The purpose of our experiments are threefold. The first is the efficiency of the proposed DCA compared with existing methods. In [5] several test problems have been performed by the SCA algorithm [5] implemented in the centralized and distributed way and some existing centralized solvers including the NEOS server [14] based on MINOS solver and PSwarm. Numerical results in [5] have showed that the distributed SCA scheme outperform the MINOS solver while it has the same performance of the two SCA schemes and PSwarm. Hence, we chose the distributed SCA as the comparative algorithm in our experiments. Note that SCA is also a DCA based algorithm corresponding to a natural DC decomposition of the objective function. Hence the comparison between our DCA schemes and the SCA is a meaning to study the effect of DC decomposition in DCA based approaches.
Our second purpose is to evaluate the influence of solution methods for solving the resulting convex subproblem in DCA (this is also a consequence of the choice of DC decomposition). For this purpose we compare the two DCA schemes: DCAC and DCAD.
The three first experiments are performed for the two mentioned purposes. For each case of the considered parameters we performed three algorithms DCAC, DCAD and SCA on 10 independent channels. The average value (SSR-AVER) as well as the best value (SSR-Best) of SSR and the average of CPU time (in seconds) of these algorithms are reported in Table 3.1 (experiment 1), Table 3.2 (experiment 2), and Table 3.3 (experiment 3).
The third purpose is to study the efficiency of the algorithm developed for the projection problem (the inner problem in the dual based projection algorithm). We compare DCAD (using the projection algorithm) with the *modified* DCAD in which the CPLEX software is applied on the projection problem. The numerical of this experiment (experiment 4) is reported in Table 3.4.

3.5.2 Numerical results and comments

Comments on computational results.
Generally speaking, the SSR given by all the algorithms tends to increase when the number of users or $\texttt{snr}$ or J goes up. Moreover, the average and the best SSR achieved by all the algorithms are quite comparable while the difference of CPU time among

Table 3.1: Comparison of System secrecy rate (SSR) obtained by all the algorithms versus number Q of legitimate users (`snr`=10)

Q		DCAD	DCAC	SCA
10	SSR-AVER	18.700	**18.836**	18.346
	SSR-Best	25.402	25.402	**25.807**
	CPU(s)	**0.013**	1.703	59.585
20	SSR-AVER	47.554	**47.848**	46.045
	SSR-Best	59.368	59.368	**59.436**
	CPU(s)	**0.118**	4.187	114.431
30	SSR-AVER	68.021	**68.885**	67.569
	SSR-Best	81.096	**82.047**	80.735
	CPU(s)	**0.843**	7.175	501.005
40	SSR-AVER	90.837	**91.174**	90.061
	SSR-Best	105.119	**105.800**	104.530
	CPU(s)	**2.421**	11.215	1369.514
50	SSR-AVER	109.983	**116.493**	109.831
	SSR-Best	151.231	**163.866**	147.901
	CPU(s)	**9.175**	18.312	2041.382

Table 3.2: System secrecy rate (SSR) versus various values of `snr` in the case of 10 users

`snr`		DCAD	DCAC	SCA
5	SSR-AVER	16.468	**16.491**	16.300
	SSR-Best	21.849	21.849	**22.090**
	CPU(s)	**0.01**	1.07	51.06
10	SSR-AVER	18.700	**18.836**	18.346
	SSR-Best	25.402	25.402	**25.807**
	CPU(s)	**0.013**	1.702	59.585
15	SSR-AVER	20.590	20.596	20.576
	SSR-Best	**27.758**	**27.758**	**27.758**
	CPU(s)	**0.012**	1.492	29.596

Table 3.3: System secrecy rate (SSR) versus various number of jammers (J) in the case of 10 users

J		DCAD	DCAC	SCA
2	SSR-AVER	20.413	**20.650**	20.267
	SSR-Best	**29.384**	**29.384**	**29.384**
	CPU(s)	**0.01**	1.803	28.033
4	SSR-AVER	**20.652**	20.650	20.645
	SSR-Best	**30.555**	**30.555**	**30.555**
	CPU(s)	**0.02**	1.809	40.848
6	SSR-AVER	21.061	21.047	20.818
	SSR-Best	**27.124**	27.123	26.176
	CPU(s)	**0.02**	1.825	35.007
8	SSR-AVER	**22.749**	22.74	22.614
	SSR-Best	**27.836**	**27.836**	**27.836**
	CPU(s)	**0.02**	1.908	56.337

Table 3.4: The runtime of DCAD when the subproblem is solved by CPLEX and Projection Algorithm, respectively (`snr = 10`).

Q	10	20	30	40	50
DCAD-CPLEX	7.958	93.371	499.211	1025.917	1039.171
DCAD-Proj	0.013	0.118	0.843	2.421	9.175

them is vast. More specifically, to fulfil the purpose of our experiments it is worth to mention the following observations.

• *DCAs versus SCA.*

In terms of secrecy rate, it can be observed from Table 3.1, Table 3.2 and Table 3.3 that, in general, both DCAD and DCAC yield the average system secrecy rates better than those of SCA. DCAC is the best in the sense that it furnishes the most superior SSR-AVER and SSR-Best. SCA although gains the superior SSR-best in some cases, it always gives the worst SSR-AVER.

In terms of CPU time, both DCAD and DCAC are much faster than SCA. The ratios of gain of both DCA schemes versus SCA are significant, especially when the number of users is large. The gain ratio is up to 970 times between DCAD and SCA and 124 times between DCAC and SCA. More precisely, when the number of users is less than 30, the CPU time of DCAD (resp. DCAC) is nearly zero (resp. less than 10 seconds) while SCA consumes more than 100 seconds. When the number of users are greater than 30, the DCAD (resp. DCAC) consumes less than 10 seconds (resp. less than 20seconds) whereas the CPU time of SCA is more than 1000 seconds.

The efficiency of two proposed DCA schemes demonstrates that the various DC decompositions bring very different effects on the quality of the obtained solutions as well as the rapidity of the corresponding DCA.

• *DCAD versus DCAC.*

It can be observed from three first tables that, in terms of the secrecy rate, DCAC provides a bit better result than DCAD in most of cases. This can be explained by the fact that, the solution of the convex subproblems given by the dual based projection algorithm used in DCAD may not be exact (the tolerance $\epsilon_2 = 10^{-2}$ may be not small enough) while the CPLEX software employed in DCAC for solving this strongly convex quadratic program gives an exact solution. By contrast (and unsurprisingly), DCAC is more expensive than DCAD. The gain ratio of the CPU time is up to 180 times. This experimental results show that the distributed DCA is an effective and scalable approach and thereby very recommended for large-scale problems, it realizes well the trade-off between the efficiency and the rapidity.

• *The proposed projection algorithm versus CPLEX.*

The results in Table 3.4 shows that the projection algorithm is much faster than the CPLEX software for the projection problem in the dual based gradient projection algorithm, especially when the number of users is large. The gap is up to the thousands of seconds. This considerable distinction as well as the gain between DCAD and DCAC illustrated the necessity of the development of efficient convex optimization methods for large-scale problems having special structure, even if efficient standard softwares are available.

3.6 Conclusion

In this chapter, we have investigated DC programming and DCA for tackling a resource allocation problem which aims to maximize the system secrecy rate in the physical layer. We have carefully studied the two main challenges in DC programming and DCA that are the effect of DC decomposition and the efficiency of solution methods to convex subproblems. The double advantages of the new proposed DC formulation have been explored and exploited in the design of two efficient DCA schemes based on centralized and distributed approaches. Firstly, the resulting convex subproblem is a linearly constrained strongly quadratic program for which several standard softwares are available. Secondly, the very special structure of the feasible set of this problem has been exploited in an elegant and deeper way to develop the distributed dual decomposition algorithm based on the gradient projection algorithm which requires computing iteratively the projection of points onto the intersection of a box and a half space. We have proposed a very inexpensive algorithm for this projection problem by adapting an existing projection method to a similar structural set. The computational results on several datasets have shown the robustness as well as the efficiency of the proposed DCA schemes in terms of both quality and rapidity, and their superiority compared with the related existing approach SCA. These results confirm, once again, the nice effect of DC decomposition as well as the crucial role of solution methods investigated to resulting convex subproblems, in particular in the large-scale setting.

The techniques proposed in this chapter can be extended to the more general classes of problems in DC programming framework. For instance, they can be directly applied for minimizing a smooth function with Lipschitz continuous gradient on a bounded polyhedral convex set defined by some separate constraints and some coupling constraints. This chapter provides more evidences to show that DC programming and DCA is an efficient and robust approach for solving the nonconvex optimization problems in a wide range of areas.

Chapter 4

DC Programming and DCA for Enhancing Physical Layer Security via Relay Beamforming Strategies[1]

Abstract: *Apart from cryptography which is the primary traditional method for ensuring information security and confidentiality, the appearance of the physical layer security approach plays an important role for not only enabling the data transmission confidentially without relying on higher-layer encryption, but also enhancing confidentiality of the secret key distribution in cryptography. Many techniques are employed in physical layers to improve secure transmission including cooperative relaying and beamforming technique. In this chapter, we consider the secrecy rate maximization problems using two techniques mentioned above with two different relaying protocols: Amplify-and-Forward (AF) and Decode-and-Forward (DF). The optimization problems with the aim of maximizing secrecy rate subject to total and individual relay power constraints were formulated as nonconvex optimization problems, which can be reformulated as DC (difference of two convex functions) programs and thus can be solved by DCA (DC Algorithms). The special structure of the feasible set is exploited to propose an efficient DC decomposition in the sense that it leads to convex optimization subproblems that can be explicitly solved. The numerical results show that the proposed DCA schemes are better than the existing methods in terms of both runtime and secrecy rate.*

4.1 Introduction and Related Works

In this chapter, we consider a wireless relay network comprising one source, one destination, multiple relays and one eavesdropper. In this network, the source tries to transmit signal to the destination with the help of relays employing beamforming tech-

1. The material of this chapter is developed from the following work:
[1]. Tran Thi Thuy, Nguyen Nhu Tuan, Le Thi Hoai An and Alain Gély. DC programming and DCA for Enhancing Physical Layer Security via Relay Beamforming Strategies. In "Intelligent Information and Database Systems", ACIIDS 2016, Lecture Note in Computer Science LNCS, pp 640-650, Springer 2016.

nique so that the transmitted information is kept secret as much as possible from the eavesdropper. Beamforming is a signal processing technique used for directional signal transmission or reception. It aims to direct the signal to the given direction while having attenuation in others. The perfect channel state information (CSI) is assumed to be available. The two most well-known relaying protocols AF and DF are considered. In both these protocols, the messages from the source are transmitted to the destination in two stages. The first stage is the same for both the AF and DF protocols. In this stage, the source broadcasts the encoded signal to the relays and the relays receive a noisy version of this encoded signal. In the second stage, for the AF protocol, the relays transmit a weighted version of the noisy signal that they received from the first stage. Meanwhile, for the DF protocol, the relays first decode the noisy signal, re-encode it and then forward a weighted version of the re-encoded signal to the destination. Our purpose is to determine the weights at relays for both the AF and DF scenarios, called beamforming coefficients, in order to maximize the secrecy rate of this system subject to total or individual relay power constraints. It should be noted that the noise at the relays is dropped out in the DF scenario whereas it is forwarded to the destination in the AF scenario. Therefore, the mathematical expression of the secrecy rate is simpler in the DF scenario than in the AF scenario. As a result, the secrecy rate maximization (SRM) problem in the DF scenario is easier to deal with than that in the AF scenario.

The SRM problem in the DF scenario were established in [123]. It was shown that, under the total power constraint the optimal solution was found. For the individual power constraints, three state-of-the-art approaches were proposed to deal with the SRM problem in this case, namely semidefinite relaxation, second-order cone programming and suboptimal. The experiments in [123] indicated that the semidefinite relaxation method outperformed the other ones. The SRM problem in the AF scenario was mentioned in some works [17, 124, 85]. In [17], the authors only considered the case of total power constraint and proposed a suboptimal approach by maximizing the upper and lower bounds of the objective function. The iterative algorithm based on semidefinite relaxation was provided in [124] to address this problem in both cases of total and individual power constraints. However, because of relaxation technique, the obtained solution may not return a feasible solution to the original problem. To attain such a solution, one had to apply some randomization techniques on the relaxed solution, which is only a heuristic search. Therefore, the convergence of this algorithm and the solution property have been still unknown. The latest work [85] proposed a polynomial-time algorithm to solve this problem but only in two special cases of channel condition: degraded eavesdropper channel with complex channel gain and scaled eavesdropper channel with real-valued channel gains. To the best of our knowledge, there is not any existing work dealing with these problems comprehensively, with convergence guarantee for the algorithms. This motivates us to develop a new approach to handle these problems in a more efficient and general way. DC programming and DCA are well-known as powerful tools for nonconvex optimization problems. They are widely and successfully applied for tackling the hard and large-scale nonconvex programs in various areas such as communication systems [4, 36, 50, 52, 53, 93, 94, 95, 104, 106, 128] and other fields [48, 49, 55, 56, 78, 79, 80] and references therein as well as the list

of references in [Le Thi]. The experiments in many works show that DCA based approach outperforms other standard methods. In addition, the convergence of DCA is guaranteed by the rigorous and complete theory of DC programming and DCA. All of these reasons make us choose and develop these tools for solving the aforementioned SRM problems.

Our contributions reside in developing a new approach based on DC programming and DCA to efficiently handle both SRM problems in the AF and DF scenario, respectively. First, we reformulate these SRM problems as standard DC programs and then develop the efficient DCA schemes for solving them. We exploit the special structure of the feasible set to propose an efficient DC decomposition in the sense that it arises convex subproblems which can be explicitly solved. It turns out that, the corresponding DCA scheme requires computing iteratively the projection of points onto an Euclidean ball or an intersection of Euclidean balls, which can be explicitly determined. Searching a good DC decomposition that results in easy-to-solve convex subproblems is highly recommended in DC programming and DCA because it brings good effects on the convergence speed of DCA as well as the properties of the found solution. In fact, the numerical results show that our approach is better than the existing methods in both computation time and security aspects.

The rest of this chapter is organized as follows. In Section 4.2, we describe the considered secrecy rate maximization problems in both AF and DF scenarios. Section 4.3 illustrates how to apply DCA to solve the considered problems. Experimental results are reported in Section 4.4. Finally, Section 4.5 concludes the chapter.

4.2 Secrecy Rate Maximization via Relay Beamforming

In this section, we present the models, which were introduced in [123, 124]. Consider a communication system comprised of a source (S), a destination (D), an eavesdropper (E) and M relays ($R_1, R_2, ..., R_M$). It is assumed that there is no direct link between S and D as well as S and E. Let $[f_1, ..., f_M] \in \mathbb{C}^M$, $[h_1, ..., h_M] \in \mathbb{C}^M$, $[z_1, ..., z_M] \in \mathbb{C}^M$ denote the channel coefficients between the source and the relays, the relays and the destination, the relays and the eavesdropper, respectively. The common purpose of these models is to determine beamforming coefficients at the relays in order to maximize the secrecy rate subject to the total or individual power constraints.

4.2.1 Amplify-and-Forward (AF) relay beamforming design

4.2.1.1 Problem formulation

In the AF scenario, the signal is transmitted through two hops. In the first hop, the signal x_s with power $E(|x_s|^2) = P_s$ is transmitted to all the relays by the source.

The relay R_m receives the signal given by $y_{R,m} = f_m x_s + n_{R,m}$, where $n_{R,m}$ is the background noise that has a Gaussian distribution with zero mean and variance σ_m^2. In the second hop, this signal is forwarded to the destination D after being multiplied by $k_m w_m$ without decoding, where w_m is a beamforming coefficient and k_m is a scaling factor. The relay output signal can be written as

$$x_{R,m} = w_m k_m (f_m x_s + n_{R,m}).$$

The scaling factor is chosen such that $E[|x_{R,m}|^2] = |w_m|^2$ therefore it is computed by $k_m = \frac{1}{\sqrt{|f_m|^2 P_s + \sigma_m^2}}$. The received signals at the destination D and the eavesdropper E are given by

$$y_D = \sum_{m=1}^{M} h_m w_m k_m (f_m x_s + n_{R,m}) + n_D$$

and

$$y_E = \sum_{m=1}^{M} z_m w_m k_m (f_m x_s + n_{R,m}) + n_E,$$

where n_D, n_E are the Gaussian background noise components at D and E, respectively, with zero mean and variance σ_0^2. The received SNR at D and E are computed as ([124])

$$\Gamma_D = \frac{|\sum_{m=1}^{M} h_m w_m k_m f_m|^2 P_s}{\sum_{m=1}^{M} |h_m|^2 k_m^2 |w_m|^2 \sigma_m^2 + \sigma_0^2}$$

and

$$\Gamma_E = \frac{|\sum_{m=1}^{M} z_m w_m k_m f_m|^2 P_s}{\sum_{m=1}^{M} |z_m|^2 k_m^2 |w_m|^2 \sigma_m^2 + \sigma_0^2}.$$

Denote

$$\begin{aligned} \mathbf{w} &= [w_1, ..., w_M]^T, \mathbf{h} = [h_1^* k_1 f_1^*, ..., h_M^* k_M f_M^*]^T, \mathbf{z} = [z_1^* k_1 f_1^*, ..., z_M^* k_M f_M^*]^T, \\ \mathbf{D}_h &= \text{diag}\left[|h_1|^2 k_1^2 \sigma_1^2, ..., |h_M|^2 k_M^2 \sigma_M^2\right], \mathbf{D}_z = \text{diag}\left[|z_1|^2 k_1^2 \sigma_1^2, ..., |z_M|^2 k_M^2 \sigma_M^2\right]. \end{aligned}$$

The received SNR at D and E can be rewritten as follows

$$\begin{aligned} \Gamma_D &= \frac{|\mathbf{h}^\dagger \mathbf{w}|^2 P_s}{\mathbf{w}^\dagger \mathbf{D}_h \mathbf{w} + \sigma_0^2} = \frac{\mathbf{w}^\dagger \mathbf{h}\mathbf{h}^\dagger \mathbf{w} P_s}{\mathbf{w}^\dagger \mathbf{D}_h \mathbf{w} + \sigma_0^2} = \frac{\mathbf{w}^\dagger P_s \mathbf{H} \mathbf{w}}{\mathbf{w}^\dagger \mathbf{D}_h \mathbf{w} + \sigma_0^2}, \\ \Gamma_E &= \frac{|\mathbf{z}^\dagger \mathbf{w}|^2 P_s}{\mathbf{w}^\dagger \mathbf{D}_z \mathbf{w} + \sigma_0^2} = \frac{\mathbf{w}^\dagger \mathbf{z}\mathbf{z}^\dagger \mathbf{w} P_s}{\mathbf{w}^\dagger \mathbf{D}_z \mathbf{w} + \sigma_0^2} = \frac{\mathbf{w}^\dagger P_s \mathbf{Z} \mathbf{w}}{\mathbf{w}^\dagger \mathbf{D}_z \mathbf{w} + \sigma_0^2}, \end{aligned}$$

where $\mathbf{H} = \mathbf{h}\mathbf{h}^\dagger$ and $\mathbf{Z} = \mathbf{z}\mathbf{z}^\dagger$.
The secrecy rate of this system is given by

$$\begin{aligned} R_s &= \log_2(1 + \Gamma_D) - \log_2(1 + \Gamma_E) \\ &= \log_2 \left(\frac{\mathbf{w}^\dagger \mathbf{A}_h \mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger \mathbf{D}_h \mathbf{w} + \sigma_0^2} . \frac{\mathbf{w}^\dagger \mathbf{D}_z \mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger \mathbf{A}_z w + \sigma_0^2} \right), \end{aligned}$$

where $\mathbf{A}_h = \mathbf{D}_h + P_s\mathbf{H}$ and $\mathbf{A}_z = \mathbf{D}_z + P_s\mathbf{Z}$.
The relays can be imposed by either the total power constraint given by

$$\|\mathbf{w}\|^2 = \mathbf{w}^\dagger\mathbf{w} \leq P_T,$$

or the individual power constraints as follows

$$|w_m|^2 \leq p_m, \quad \forall m = 1, ..., M.$$

The problem of secrecy rate maximization with total (individual) relay power constraints takes the form

$$\begin{aligned} \max_{\mathbf{w}} \quad & \log_2\left(\frac{\mathbf{w}^\dagger\mathbf{A}_h\mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger\mathbf{D}_h\mathbf{w} + \sigma_0^2}.\frac{\mathbf{w}^\dagger\mathbf{D}_z\mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger\mathbf{A}_z\mathbf{w} + \sigma_0^2}\right) \\ \text{s.t.} \quad & \|\mathbf{w}\|^2 \leq P_T, \\ & (\text{or } |w_m|^2 \leq p_m, \ \forall m = 1, ..., M). \end{aligned} \tag{4.1}$$

The problem (4.1) is nonconvex and thus hard to deal with.

4.2.1.2 Existing methods

There are two state-of-the-art methods to solve this nonconvex program. The widely used method to solve the problem (4.1) is based on a semidefinite relaxation ([124]). Another method to solve that problem is a suboptimal approach ([17]). In what follows, we will briefly describe these methods.

The semidefinite relaxation based approach (SDR)
By denoting $\mathbf{X} = \mathbf{w}\mathbf{w}^\dagger$ then $\mathbf{X} \succeq 0$ and $\text{rank}(\mathbf{X}) = 1$ and note that $\mathbf{w}^\dagger\mathbf{A}\mathbf{w} = \text{tr}(\mathbf{A}\mathbf{X})$ $\forall$ matrices $\mathbf{A}$, $\|\mathbf{w}\|^2 = \text{tr}(\mathbf{X})$, $|w_m|^2$ is actually the mth entry of $\text{diag}(\mathbf{X})$, the problem (4.1) was reformulated and then relaxed by ignoring the rank-one constraint as follows ([124])

$$\begin{aligned} \max_{\mathbf{X}, t_1, t_2} \quad & t_1 t_2 \\ \text{s.t.} \quad & \text{tr}(\mathbf{X}) \leq P_T, (\text{or } \text{diag}(\mathbf{X}) \leq \mathbf{p}), \\ & \text{tr}(\mathbf{X}(\mathbf{D}_z - t_2\mathbf{D}_h)) \geq \sigma_0^2(t_2 - 1), \\ & \text{tr}(\mathbf{X}(\mathbf{A}_h - t_1\mathbf{A}_z)) \geq \sigma_0^2(t_1 - 1), \\ & \mathbf{X} \succeq 0. \end{aligned}$$

First of all, the optimal values $t_{1,u}, t_{2,u}$ of the following problems were found based on a bisection method combined with a semidefinite programming.

$$\begin{aligned} \max_{\mathbf{X}, t_1} \quad & t_1 \\ \text{s.t.} \quad & \text{tr}(\mathbf{X}) \leq P_T, (\text{or } \text{diag}(\mathbf{X}) \leq \mathbf{p}), \\ & \text{tr}(\mathbf{X}\mathbf{A}_h - t_1\mathbf{A}_z) \geq \sigma_0^2(t - 1), \\ & \mathbf{X} \succeq 0. \end{aligned}$$

and

$$\begin{aligned} \max_{\mathbf{X},t_2} \quad & t_2 \\ \text{s.t.} \quad & \text{tr}(\mathbf{X}) \leq P_T, (\text{or } \text{diag}(\mathbf{X}) \leq \mathbf{p}) \\ & \text{tr}(\mathbf{X}(\mathbf{D}_z - t_2\mathbf{D}_h)) \geq \sigma_0^2(t_2 - 1), \\ & \mathbf{X} \succeq 0. \end{aligned}$$

Next, from those optimal values, the authors in [124] proposed an iterative algorithm to search for the optimal values $t_{1,0}, t_{2,0}$ that maximize the product $t_1 t_2$. Finally, the optimal solution was obtained by finding the solution $\mathbf{X}$ with the smallest trace among the solutions corresponding to $t_{1,0}$ and $t_{2,0}$.

Suboptimal approach (SubOpt)

To obtain a suboptimal solution to the problem (4.1), instead of maximizing the whole objective function, one only maximizes a part of it. More particularly, the suboptimal $\mathbf{w}$ is found by solving the following problem

$$\begin{aligned} \max_{\mathbf{w}} \quad & \log_2\left(\frac{\mathbf{w}^\dagger \mathbf{A}_h \mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger \ \mathbf{A}_z \mathbf{w} + \sigma_0^2}\right) \\ \text{s.t.} \quad & \|\mathbf{w}\|^2 \leq P_T, \\ & (\text{or } |w_m|^2 \leq p_m, \ \forall m = 1, ..., M). \end{aligned} \tag{4.2}$$

By introducing a slack variable t_1 and using a variable transformation $X = ww^\dagger$ (thus $X \succeq 0, \text{rank}(X) = 1$), this problem is relaxed to the following form by eliminating the constraint rank(X) = 1.

$$\begin{aligned} \max_{\mathbf{X},t_1} \quad & t_1 \\ \text{s.t.} \quad & \text{tr}(\mathbf{X}) \leq P_T, (\text{or } \text{diag}(\mathbf{X}) \leq \mathbf{p}), \\ & \text{tr}(\mathbf{X}\mathbf{A}_h - t_1\mathbf{A}_z) \geq \sigma_0^2(t - 1), \\ & \mathbf{X} \succeq 0. \end{aligned} \tag{4.3}$$

From the relaxed solution $\mathbf{X}$ obtained by solving (4.3), compute $t_2 = \frac{\sigma_0^2 + \text{tr}(\mathbf{D}_z\mathbf{X})}{\sigma_0^2 + \text{tr}(\mathbf{D}_h\mathbf{X})}$ and then $\log_2(t_1 t_2)$ is regarded as a suboptimal value of the secrecy rate.

4.2.2 Decode-and-Forward (DF) relay beamforming Design

4.2.2.1 Problem formulation

In this DF relay beamforming model, the source S transmits a signal x_s with power $E(|x_s|^2) = P_s$ to the relays. Afterwards, each relay R_m first decodes the message x_s and then normalized it as $x'_s = \frac{x_s}{\sqrt{P_s}}$. Subsequently, the normalized message is multiplied by the weight factor w_m to generate the transmitted signal $x_{r,m} = w_m x'_s$. The output power of each relay R_m is given by $E(\|x_{r,m}\|^2) = E(\|w_m x'_s\|^2) = |w_m|^2$. The received signals at the destination and the eavesdropper are respectively given by

$$y_D = \sum_{m=1}^{M} h_m w_m x'_s + n_D \quad \text{and} \quad y_E = \sum_{m=1}^{M} z_m w_m x'_s + n_E,$$

where n_D and n_E are the Gaussian background noise components at D and E, respectively, with zero mean and variance σ_0^2.

The received SNR levels at the destination and the eavesdropper are given by

$$\Gamma_D = \frac{|\sum_{m=1}^{M} h_m w_m|^2}{\sigma_0^2} \quad \text{and} \quad \Gamma_E = \frac{|\sum_{m=1}^{M} z_m w_m|^2}{\sigma_0^2}.$$

Denote $\hat{\mathbf{h}} = [h_1^*, ..., h_M^*]^T$ and $\hat{\mathbf{z}} = [z_1^*, ..., z_M^*]^T$. The secrecy rate of this system is given by

$$\begin{aligned} R_s &= \log_2(1+\Gamma_D) - \log_2(1+\Gamma_E) \\ &= \log_2\left(\frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{h}}\hat{\mathbf{h}}^\dagger \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{z}}\hat{\mathbf{z}}^\dagger \mathbf{w}}\right) = \log_2\left(\frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}}\right), \end{aligned}$$

where $\hat{\mathbf{H}} = \hat{\mathbf{h}}\hat{\mathbf{h}}^\dagger$ and $\hat{\mathbf{Z}} = \hat{\mathbf{z}}\hat{\mathbf{z}}^\dagger$. We consider the optimization problem of maximizing the secrecy rate with the relay power constraints as follows

$$\begin{aligned} \max_{\mathbf{w}} \quad & \log_2\left(\frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}}\right) \\ \text{s.t.} \quad & |w_m|^2 \le p_m, \ \forall m = 1, ..., M. \\ & (\text{or } \|\mathbf{w}\|^2 \le P_T). \end{aligned} \tag{4.4}$$

Because the logarithm base 2 function is monotonically increasing, thus rather than solve the problem (4.4), one can solve the following equivalent problem

$$\begin{aligned} \max_{\mathbf{w}} \quad & \frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}} \\ \text{s.t.} \quad & |w_m|^2 \le p_m \ , \forall m = 1, ..., M. \\ & (\text{or } \|\mathbf{w}\|^2 \le P_T). \end{aligned} \tag{4.5}$$

4.2.2.2 Existing methods

Recall that if the total relay power constraint equality is imposed, i.e., $\|\mathbf{w}\|^2 = P_T$ then the problem (4.5) is completely solved using the generalized eigenvalue. In more detail, we have

$$\begin{aligned} & \max_{\|\mathbf{w}\|^2 = P_T} \quad \frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}} \\ = & \max_{\|\mathbf{w}\|^2 = P_T} \quad \frac{\sigma_0^2 \mathbf{w}^\dagger \mathbf{I}_M \mathbf{w}/P_T + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 \mathbf{w}^\dagger \mathbf{I}_M \mathbf{w}/P_T + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}} \\ = & \max_{\|\mathbf{w}\|^2 = P_T} \quad \frac{\mathbf{w}^\dagger (\sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{H}}) \mathbf{w}}{\mathbf{w}^\dagger (\sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{Z}}) \mathbf{w}} \\ = & \ \lambda_{max} \quad \left(\sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{H}}, \sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{Z}}\right), \end{aligned}$$

where $\lambda_{max}(\mathbf{A}, \mathbf{B})$ is the largest generalized eigenvalue of the matrix pair $(\mathbf{A}, \mathbf{B})$. In case of total relay power constraint inequality, i.e., $\|\mathbf{w}\|^2 \leq P_T$, if there exist at least a point satisfying this constraint such that $R_s \geq 0$, which is equivalent to $\frac{\sigma_0^2+\mathbf{w}^\dagger\hat{\mathbf{H}}\mathbf{w}}{\sigma_0^2+\mathbf{w}^\dagger\hat{\mathbf{Z}}\mathbf{w}} \geq 1$, then it is not difficult to show that

$$\begin{aligned} & \max_{\|\mathbf{w}\|^2 \leq P_T} \frac{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{H}} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}} \\ = & \max_{\|\mathbf{w}\|^2 = P_T} \frac{\sigma_0^2 + w^\dagger \hat{H} \mathbf{w}}{\sigma_0^2 + \mathbf{w}^\dagger \hat{\mathbf{Z}} \mathbf{w}} \\ = & \lambda_{max}\left(\sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{H}}, \sigma_0^2/P_T.\mathbf{I}_M + \hat{\mathbf{Z}}\right). \end{aligned}$$

Actually, we should choose the channel coefficients such that the above condition is satisfied. Otherwise, the secrecy rate of the system will be zero, which is insignificant.

When the individual relay power constraints are used, the problem (4.5) becomes more difficult to solve. As mentioned before, there are three existing methods to solve (4.5) in this case and the SDR based method is the best among these methods. Thus, in the followings we only present shortly this method. By denoting the square matrix $\mathbf{X}$ as $\mathbf{w}.\mathbf{w}^\dagger$, the problem (4.5) can be reformulated as follows:

$$\begin{aligned} \max_{\mathbf{X},t} \quad & t \\ \text{s.t.} \quad & \text{diag}(\mathbf{X}) \leq \mathbf{p}, \\ & \mathbf{X} \geq 0, \\ & \text{rank}(\mathbf{X}) = 1, \\ & \text{tr}((\hat{\mathbf{H}} - t\hat{\mathbf{Z}})\mathbf{X}) \geq \sigma_0^2(t-1), \end{aligned} \tag{4.6}$$

where $\mathbf{p} = [p_1, ..., p_M]^T$. If the rank constraint is discarded, the resulting problem can be solved efficiently by the interior point method with bisection algorithm ([123]). However, the relaxed solution may not satisfy the rank-one constraint. To find the solution to the original problem, one has to apply some randomization techniques. This is only a heuristic approach and thus the properties of the sought solution are still unknown.

In the next section, we will study how to solve both problems (4.1) and (4.4) via DC programming and DCA.

4.3 Solution Methods Based on DC Programming and DCA

4.3.1 DC Programming and DCA for solving the secrecy rate maximization problem in the AF scenario

In this section, we investigate DC programming and DCA to solve the problem (4.1) in two cases. In the first case, rather than directly deal with the problem (4.1), we propose a null-space relay beamforming design in which the beamforming vector $\mathbf{w}$ is designed to completely eliminate the signal at the eavesdropper while maximizing the signal at the destination. It means that $\mathbf{w}$ has to satisfy the constraint $\mathbf{z}^\dagger\mathbf{w} = 0$, where $\mathbf{z}^\dagger$ is the vector of equivalent channel coefficients between the source and the eavesdropper. This constraint makes the mathematical form of the objective function to be reduced, and thus easier to deal with. However, the additional constraint makes the feasible set of the problem (4.1) smaller, hence the obtained solution $\mathbf{w}$ might be not the best beamforming design in terms of secrecy. Therefore, in the second case, we consider a more general beamforming design in which the signal might be not altogether eliminated at the eavesdropper. The beamforming vector $\mathbf{w}$ is found by directly solving the hard problem (4.1). In the sequel, we will illustrate how to apply DC programming and DCA to the two above cases.

4.3.1.1 Null-Space relay beamforming design

The null-space relay beamforming design is often adopted in order to completely eliminate the information leakage in the eavesdropper's channel ([18], [127]). It means that beside of the relay power constraint, the beamforming vector $\mathbf{w}$ has to satisfy the equation $\mathbf{z}^\dagger\mathbf{w} = 0$. As a result, the optimization problem (4.1) is reduced to the following problem

$$\begin{aligned} \max_{\mathbf{w}} \quad & \frac{\mathbf{w}^\dagger\mathbf{A}_h\mathbf{w} + \sigma_0^2}{\mathbf{w}^\dagger\mathbf{D}_h\mathbf{w} + \sigma_0^2} \\ \text{s.t.} \quad & \|\mathbf{w}\|^2 \le P_T, \\ & (\text{or } |w_m|^2 \le p_m,\ \forall m = 1, ..., M), \\ & \mathbf{z}^\dagger\mathbf{w} = 0. \end{aligned} \tag{4.7}$$

By denoting $\mathbf{M}_h = \begin{bmatrix} \text{Re}(\mathbf{A}_h) & -\text{Im}(\mathbf{A}_h) \\ \text{Im}(\mathbf{A}_h) & \text{Re}(\mathbf{A}_h) \end{bmatrix}$, $\mathbf{T}_h = \begin{bmatrix} \text{Re}(\mathbf{D}_h) & -\text{Im}(\mathbf{D}_h) \\ \text{Im}(\mathbf{D}_h) & \text{Re}(\mathbf{D}_h) \end{bmatrix}$, $\mathbf{x} = [\text{Re}(\mathbf{w}^T)\ \text{Im}(\mathbf{w}^T)]^T, \hat{\mathbf{z}} = \begin{bmatrix} \text{Re}(\mathbf{z}^\dagger) & -\text{Im}(\mathbf{z}^\dagger) \\ \text{Im}(\mathbf{z}^\dagger) & \text{Re}(\mathbf{z}^\dagger) \end{bmatrix}$ and introducing a slack variable t, the

problem (4.7) can be equivalently transfered to the problem below

$$\begin{aligned} \min_{\mathbf{x},t} \quad & -\frac{\mathbf{x}^T\mathbf{M}_h\mathbf{x}+\sigma_0^2}{t} \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \le P_T, \\ & (\text{or } |x_m|^2 + |x_{M+m}|^2 \le p_m,\ \forall m = 1, ..., M), \\ & \hat{\mathbf{z}}\mathbf{x} = 0, \\ & \mathbf{x}^T\mathbf{T}_h\mathbf{x} + \sigma_0^2 \le t. \end{aligned} \tag{4.8}$$

Since $t > 0$ and $\mathbf{M}_h \succeq 0$, $\frac{\mathbf{x}^T\mathbf{M}_h\mathbf{x}+\sigma_0^2}{t}$ is a convex function. Therefore, (4.8) is actually a special version of a DC program, in which the objective function, denoted as $F_1(x,t)$, is of the difference of two convex functions $G_1(\mathbf{x},t) = 0$ and $H_1(\mathbf{x},t) = \frac{\mathbf{x}^T\mathbf{M}_h\mathbf{x}+\sigma_0^2}{t}$. Note that, $H_1(\mathbf{x},t)$ is differentiable and its gradient at the point $(\mathbf{x}^k, t^k)$ is given by

$$\nabla H_1(\mathbf{x}^k, t^k) = \left[\left(\frac{2\mathbf{M}_h\mathbf{x}^k}{t^k}\right)^T \quad -\frac{(\mathbf{x}^k)^T\mathbf{M}_h\mathbf{x}^k + \sigma_0^2}{(t^k)^2}\right]^T$$

Applying DCA generic scheme for this DC program, we obtain the algorithm, called DCA-NS, as follows

---

Algorithm 1: The DCA-NS scheme.

Initialization: Choose $\mathbf{u}^0 = (\mathbf{x}^0, t^0) \in (\mathbb{R}^{2M}, \mathbb{R}^+)$ as an initial guess, set the value for the tolerance ϵ, $k \leftarrow 0$.

Repeat

• Calculate $\mathbf{u}^{k+1} = (\mathbf{x}^{k+1}, t^{k+1})$ by solving the following convex subproblem

$$\begin{aligned} \min_{\mathbf{x},t} \quad & \frac{(\mathbf{x}^k)^T\mathbf{M}_h\mathbf{x}^k + \sigma_0^2}{(t^k)^2}(t - t^k) - \left(\frac{2\mathbf{M}_h\mathbf{x}^k}{t^k}\right)^T(\mathbf{x} - \mathbf{x}^k) \qquad (4.9) \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \le P_T, \\ & (\text{or } |x_m|^2 + |x_{M+m}|^2 \le p_m,\ \forall m = 1, ..., M), \\ & \hat{\mathbf{z}}\mathbf{x} = 0, \\ & \mathbf{x}^T\mathbf{T}_h\mathbf{x} + \sigma_0^2 \le t. \qquad (4.10) \end{aligned}$$

• $k \leftarrow k + 1$.

Until $\left(\frac{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|}{1+\|\mathbf{u}^{k-1}\|} < \epsilon \text{ or } \frac{|F_1(\mathbf{u}^k) - F_1(\mathbf{u}^{k-1})|}{1+|F_1(\mathbf{u}^{k-1})|} < \epsilon\right)$.

Theorem 4.1. *(convergence properties of DCA-NS)*

(1) The Algorithm DCA-NS generates the sequence $\{(\boldsymbol{x}^k, t^k)\}$ *such that the sequence* $\{F_1(\boldsymbol{x}^k, t^k)\}$ *is monotonously decreasing.*

(2) Every limit point $(\boldsymbol{x}^*, t^*)$ *of the sequence* $\{(\boldsymbol{x}^k, t^k)\}$ *is a critical point of the problem (4.8), and more strongly, they verify the nesessary local optimality condition* $\partial H_1(\boldsymbol{x}^*, t^*) \subset \partial G_1(\boldsymbol{x}^*, t^*)$.

Proof. Obviously the sequence $\{\mathbf{x}^k\}$ is bounded because of the power constraints. Moreover, the objective function of (4.9) is monotonously increasing in t, thus its optimal value is obtained when the constraint (4.10) holds with equality, i.e. $t^k = (\mathbf{x}^k)^T \mathbf{T}_h \mathbf{x}^k + \sigma_0^2$. Therefore the sequence $\{(\mathbf{x}^k, t^k)\}$ is bounded. Furthermore, H_1 and G_1 are differentiable. As a consequence, the assertions in this theorem are straightforwardly deduced from the convergence properties of the generic DCA. □

As mentioned above, the null space relay beamforming design is considered for the purpose of simplifying the complexity of the original problem (4.1). This is performed by imposing an additional constraint on the beamforming vector $\mathbf{w}$. This makes the feasible set narrower, so the obtained solution might be only a suboptimal solution. To seek the best solution possible, we will directly address the hard problem (4.1).

4.3.1.2 General relay beamforming design

By denoting $\mathbf{M}_h = \begin{bmatrix} \mathrm{Re}(\mathbf{A}_h) & -\mathrm{Im}(\mathbf{A}_h) \\ \mathrm{Im}(\mathbf{A}_h) & \mathrm{Re}(\mathbf{A}_h) \end{bmatrix}$,$\mathbf{M}_z = \begin{bmatrix} \mathrm{Re}(\mathbf{D}_z) & -\mathrm{Im}(\mathbf{D}_z) \\ \mathrm{Im}(\mathbf{D}_z) & \mathrm{Re}(\mathbf{D}_z) \end{bmatrix}$, $\mathbf{T}_h = \begin{bmatrix} \mathrm{Re}(\mathbf{D}_h) & -\mathrm{Im}(\mathbf{D}_h) \\ \mathrm{Im}(\mathbf{D}_h) & \mathrm{Re}(\mathbf{D}_h) \end{bmatrix}$, $\mathbf{T}_z = \begin{bmatrix} \mathrm{Re}(\mathbf{A}_z) & -\mathrm{Im}(\mathbf{A}_z) \\ \mathrm{Im}(\mathbf{A}_z) & \mathrm{Re}(\mathbf{A}_z) \end{bmatrix}$, $\mathbf{x} = [\mathrm{Re}\mathbf{w}^T)\ \mathrm{Im}(\mathbf{w}^T)]^T$ the problem (4.1) can be rewritten in the following form

$$\begin{aligned} \max_{\mathbf{x}} \quad & \log_2 \left(\frac{\mathbf{x}^T \mathbf{M}_h \mathbf{x} + \sigma_0^2}{\mathbf{x}^T \mathbf{T}_h \mathbf{x} + \sigma_0^2} . \frac{\mathbf{x}^T \mathbf{M}_z \mathbf{x} + \sigma_0^2}{\mathbf{x}^T \mathbf{T}_z \mathbf{x} + \sigma_0^2} \right) \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \le P_T, \\ & (\text{or } |\mathbf{z}_m|^2 \le p_m, \text{ where } \mathbf{z}_m = [x_m \ x_{M+m}]^T, \ \forall m = 1, ..., M). \end{aligned} \tag{4.11}$$

The problem (4.11) is also equivalent to the problem below

$$\begin{aligned} \min_{\mathbf{x}} \quad & -\ln(\mathbf{x}^T \mathbf{M}_h \mathbf{x} + \sigma_0^2) + \ln(\mathbf{x}^T \mathbf{T}_h \mathbf{x} + \sigma_0^2) - \ln(\mathbf{x}^T \mathbf{M}_z \mathbf{x} + \sigma_0^2) \\ & + \ln(\mathbf{x}^T \mathbf{T}_z \mathbf{x} + \sigma_0^2) \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \le P_T, \\ & (\text{or } \|\mathbf{z}_m\|^2 \le p_m, \text{ where } \mathbf{z}_m = [x_m \ x_{M+m}]^T, \ \forall m = 1, ..., M). \end{aligned} \tag{4.12}$$

The objective function of (4.12), say $F(\mathbf{x})$, can be decomposed into a difference of two functions

$$G(\mathbf{x}) = \frac{1}{2} \rho \|\mathbf{x}\|^2$$

and

$$H(\mathbf{x}) = \frac{1}{2} \rho \|\mathbf{x}\|^2 + \ln(\mathbf{x}^T \mathbf{M}_h \mathbf{x} + \sigma_0^2) - \ln(\mathbf{x}^T \mathbf{T}_h \mathbf{x} + \sigma_0^2) + \ln(\mathbf{x}^T \mathbf{M}_z \mathbf{x} + \sigma_0^2) - \ln(\mathbf{x}^T \mathbf{T}_z \mathbf{x} + \sigma_0^2).$$

for some ρ. Note that if $\rho > 0$, the function $G(\mathbf{x})$ is convex. We aim to determine ρ such that the function $H(\mathbf{x})$ is also convex. The following theorem shows a sufficient condition of ρ to ensure the convexity of the function $H(\mathbf{x})$.

Theorem 4.2. *If ρ is greater than the largest eigenvalue of the matrix $\left(\frac{\boldsymbol{M}_h+\boldsymbol{M}_z+4(\boldsymbol{T}_h+\boldsymbol{T}_z)}{2\sigma_0^2}\right)$ then the function $H(\boldsymbol{x})$ is convex in $\boldsymbol{x}$.*

The proof of this theorem is straightforwardly deduced from the Proposition 2.1 .

As mentioned in the above theorem, when ρ is greater than the largest eigenvalue of matrix $\frac{\mathbf{M}_h+\mathbf{M}_z+4(\mathbf{T}_h+\mathbf{T}_z)}{2\sigma_0^2}$, both functions $G(\mathbf{x})$ and $H(\mathbf{x})$ are convex, hence $G(\mathbf{x})-H(\mathbf{x})$ is a DC decomposition of the objective function $F(\mathbf{x})$. As a result, we obtain a DC formulation of the problem (4.12) as below.

$$\begin{aligned} \min_{\mathbf{x}} \quad & G(\mathbf{x}) - H(\mathbf{x}) \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \leq P_T, \\ & (\text{or } \|\mathbf{z}_m\|^2 \leq p_m, \text{ where } \mathbf{z}_m = [x_m \; x_{M+m}]^T, \; \forall m = 1, ..., M). \end{aligned} \tag{4.13}$$

According to the generic DCA scheme, DCA applied to (4.13) consisting of computing, at each iteration k, a subgradient $\mathbf{y}^k = \partial H(\mathbf{x}^k)$ and solving the resulting convex program of the form

$$\min_{\mathbf{x}} \quad \frac{1}{2}\rho\|\mathbf{x}\|^2 - \langle \mathbf{y}^k, \mathbf{x}\rangle \tag{4.14}$$

$$\text{s.t.} \quad \|\mathbf{x}\|^2 \leq P_T, \tag{4.15}$$

$$(\text{or } \|\mathbf{z}_m\|^2 \leq p_m, \text{ where } \mathbf{z}_m = [x_m \; x_{M+m}]^T, \; \forall m = 1, ..., M). \tag{4.16}$$

Because the function $H(\mathbf{x})$ is differentiable, its subgradient at the point $\mathbf{x}^k$ is calculated by

$$\begin{aligned} &\mathbf{y}^k = \partial H(\mathbf{x}^k) = \nabla H(\mathbf{x}^k) = \rho\mathbf{x}^k + \\ &2\left(-\frac{\mathbf{T}_h}{(\mathbf{x}^k)^T\mathbf{T}_h\mathbf{x}^k + \sigma_0^2} - \frac{\mathbf{T}_z}{(\mathbf{x}^k)^T\mathbf{T}_z\mathbf{x}^k + \sigma_0^2} + \frac{\mathbf{M}_h}{(\mathbf{x}^k)^T\mathbf{M}_h\mathbf{x}^k + \sigma_0^2} + \frac{\mathbf{M}_z}{(\mathbf{x}^k)^T\mathbf{M}_z\mathbf{x}^k + \sigma_0^2}\right)\mathbf{x}^k. \end{aligned} \tag{4.17}$$

The convex subproblem (4.14) can be explicitly solved thanks to the nice structure of the proposed DC decomposition in combination with the spherical property of the feasible set. More specifically, solving the subproblem (4.14) is actually equivalent to finding a projection of vector $\mathbf{y}^k/\rho$ on the Euclidean ball $\|\mathbf{x}\|^2 \leq P_T$ with respect to the total constraint (4.15) and on the intersection of the Euclidean balls $\|\mathbf{z}_m\|^2 \leq p_m, \forall m = 1, ..., M$ for the individual constraint (4.16). As a consequence, we obtain explicitly the solution to the subproblem (4.14) as follows:

- if the total constraint (4.15) is imposed

$$\mathbf{x} = \begin{cases} \mathbf{y}^k/\rho, & \text{if } \|\mathbf{y}^k/\rho\|^2 \leq P_T \\ \frac{\sqrt{P_T}}{\|\mathbf{y}^k\|}\mathbf{y}^k & \text{otherwise} \end{cases} \tag{4.18}$$

- if the individual constraints (4.16) are imposed

$$\mathbf{x} = \begin{bmatrix} z_1^1 & ... & z_M^1 & z_1^2 & ... & z_M^2 \end{bmatrix}^T \tag{4.19}$$

where

$$[z_m^1 \ z_m^2]^T = \begin{cases} \mathbf{t}_m^k/\rho, & \text{if } \|\mathbf{t}_m^k/\rho\|^2 \leq p_m \\ \frac{\sqrt{p_m}}{\|\mathbf{t}_m^k\|}\mathbf{t}_m^k & \text{otherwise} \end{cases} \tag{4.20}$$

with $\mathbf{t}_m^k = [y_m^k \ y_{M+m}^k]^T$ for all $m = 1, ..., M$.

DCA applied to (4.13), namely DCA-AF, is described as follows.

Algorithm 2: The DCA-AF scheme for (4.13):

Initialization: choose $\mathbf{x}^0 \in \mathbb{R}^{2M}$ as an initial guess, set the value for the tolerance ϵ, $k \leftarrow 0$.
Repeat
• step 1. Compute $\mathbf{y}^k = \partial H(\mathbf{x}^k)$ via (4.17).
• step 2. Compute the solution $\mathbf{x}^{k+1}$ to the convex subproblem (4.14) via (4.18) for the total power constraint and via (4.19) for the individual power constraints.
• step 3. $k \leftarrow k + 1$.
Until $\left(\frac{\|\mathbf{x}^k - \mathbf{x}^{k-1}\|}{1+\|\mathbf{x}^{k-1}\|} < \epsilon \text{ or } \frac{|F(\mathbf{x}^k) - F(\mathbf{x}^{k-1})|}{1+|F(\mathbf{x}^{k-1})|} < \epsilon\right)$.

Since the objective function of (4.13) is continuous and its constraint set is compact, the optimal value of (4.13) is finite and the sequences $\{\mathbf{x}^k\}$ and $\{\mathbf{y}^k\}$ generated from Algorithm DCA-AF are bounded. According to the convergence properties of DCA presented in Section (1.2.1.2) and the fact that H is differentiable and G is strongly convex, it is straightforward to obtain the convergence theorem of the algorithm DCA-AF as follows.

Theorem 4.3. *(convergence properties of DCA-AF)*

(1) The Algorithm DCA-AF generates the sequence $\{\boldsymbol{x}^k\}$ *such that the sequence* $\{F(\boldsymbol{x}^k)\}$ *is monotonously decreasing.*

(2) Every limit point of the sequence $\{\boldsymbol{x}^k\}$ *is a critical point of the problem (4.13), and more strongly, they verify the nesessary local optimality condition* $\partial H(\mathbf{x}^*) \subset \partial G(\mathbf{x}^*)$.

(3) The series $\{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2\}$ *converges.*

4.3.2 DC Programming and DCA for solving the secrecy rate maximizations in the DF scenario

The problem (4.4) is equivalently transformed to the real form as below

$$\begin{aligned} \min_{\mathbf{x}} \quad & \log_2\left(\frac{\sigma_0^2+\mathbf{x}^T\mathbf{Z}_1\mathbf{x}}{\sigma_0^2+\mathbf{x}^T\mathbf{H}_1\mathbf{x}}\right) \\ \text{s.t.} \quad & |\mathbf{z}_m|^2 \le p_m, \text{ with } \mathbf{z}_m = \begin{bmatrix} x_m & x_{M+m} \end{bmatrix}^T \forall m = 1, ..., M, \end{aligned} \tag{4.21}$$

where $\mathbf{Z}_1 = \begin{bmatrix} \text{Re}(\hat{\mathbf{Z}}) & -\text{Im}(\hat{\mathbf{Z}}) \\ \text{Im}(\hat{\mathbf{Z}}) & \text{Re}(\hat{\mathbf{Z}}), \end{bmatrix}, \mathbf{H}_1 = \begin{bmatrix} \text{Re}(\hat{\mathbf{H}}) & -\text{Im}(\hat{\mathbf{H}}) \\ \text{Im}(\hat{\mathbf{H}}) & \text{Re}(\hat{\mathbf{H}}) \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \text{Re}(\mathbf{w}^T) & \text{Im}(\mathbf{w}^T) \end{bmatrix}^T$.
The above problem can be rewritten as a standard DC program of the form:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{\ln 2}(G_2(\mathbf{x}) - H_2(\mathbf{x})) \\ \text{s.t.} \quad & |\mathbf{z}_m|^2 \le p_m, \text{ with } \mathbf{z}_m = \begin{bmatrix} x_m & x_{M+m} \end{bmatrix}^T \forall m = 1, ..., M, \end{aligned} \tag{4.22}$$

where $G_2(\mathbf{x}) = \frac{1}{2}\tau\|\mathbf{x}\|^2, H_2(\mathbf{x}) = \frac{1}{2}\tau\|\mathbf{x}\|^2 - \ln(\sigma_0^2 + \mathbf{x}^T\mathbf{Z}_1\mathbf{x}) + \ln(\sigma_0^2 + \mathbf{x}^T\mathbf{H}_1\mathbf{x})$. The constant τ is chosen such that both G_2 and H_2 are convex. Following from Proposition 2.1, τ should equal to the largest eigenvalue of the matrix $\frac{2\mathbf{Z}_1}{\sigma_0^2} + \frac{\mathbf{H}_1}{2\sigma_0^2}$.

Assume that $\mathbf{x}^l$ is the current solution at the iteration l. DCA applied to (4.22) updates $\mathbf{x}^{l+1}$ via two steps:

- Step 1: compute $\mathbf{y}^l = \partial H_2(\mathbf{x}^l)$.
- Step 2: compute $\mathbf{x}^{l+1}$ as the optimal solution to the convex subproblem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\tau\|\mathbf{x}\|^2 - \langle \mathbf{y}^l, \mathbf{x}\rangle \\ \text{s.t.} \quad & |\mathbf{z}_m|^2 \le p_m, \text{ with } \mathbf{z}_m = \begin{bmatrix} x_m & x_{M+m} \end{bmatrix}^T \forall m = 1, ..., M. \end{aligned} \tag{4.23}$$

Obviously, the function $H_2(\mathbf{x})$ is smooth and its subgradient at a point $\mathbf{x}^l$ is given by

$$\mathbf{y}^l = \partial H_2(\mathbf{x}^l) = \tau\mathbf{x}^l - 2\left(\frac{\mathbf{Z}_1}{\sigma_0^2 + (\mathbf{x}^l)^T\mathbf{Z}_1\mathbf{x}^l} - \frac{\mathbf{H}_1}{\sigma_0^2 + (\mathbf{x}^l)^T\mathbf{H}_1\mathbf{x}^l}\right)\mathbf{x}^l. \tag{4.24}$$

The solution $\mathbf{x}^{l+1}$ to the subproblem (4.23) is nothing but the projection of $\frac{\mathbf{y}^l}{\tau}$ onto the intersection of the Euclidean balls $|\mathbf{z}_m|^2 \le p_m \ \ \forall m = 1, \ldots, M$. Therefore $\mathbf{x}^{l+1}$ can be explicitly computed as below

$$\mathbf{x}^{l+1} = \begin{bmatrix} u_1^1 & ... & u_M^1 & u_1^2 & ... & u_M^2 \end{bmatrix}^T \text{ where } \begin{bmatrix} u_m^1 & u_m^2 \end{bmatrix}^T = \begin{cases} \mathbf{t}_m^l/\tau, & \text{if } \|\mathbf{t}_m^l/\tau\|^2 \le p_m \\ \frac{\sqrt{p_m}}{\|\mathbf{t}_m^l\|}\mathbf{t}_m^l & \text{otherwise} \end{cases} \tag{4.25}$$

$$\text{with } \mathbf{t}_m^l = \begin{bmatrix} y_m^l & y_{M+m}^l \end{bmatrix}^T, \forall m = 1, ..., M.$$

Following the DCA generic scheme described in Section 1.2.1.2, DCA applied to the DC program (4.22) is given by the algorithm below.

Algorithm 3: The DCA-DF scheme for the problem (4.22).

Initialization: choose $\mathbf{x}^0 \in \mathbb{R}^{2M}$ as an initial guess, set the value for the tolerance ϵ, $l \leftarrow 0$.
Repeat
• Compute $\mathbf{y}^l = \partial H_2(\mathbf{x}^l)$ based on (4.24)
• Compute $\mathbf{x}^{l+1}$ based on (4.25)
• $l \leftarrow l + 1$,
Until $\left(\frac{\|\mathbf{x}^l - \mathbf{x}^{l-1}\|}{1+\|\mathbf{x}^{l-1}\|} < \epsilon \text{ or } \frac{|F_2(\mathbf{x}^l) - F_2(\mathbf{x}^{l-1})|}{1+|F_2(\mathbf{x}^{l-1})|} < \epsilon\right)$, where $F_2(\mathbf{x}) = \log_2\left(\frac{\sigma_0^2 + \mathbf{x}^T\mathbf{Z}_1\mathbf{x}}{\sigma_0^2 + \mathbf{x}^T\mathbf{H}_1\mathbf{x}}\right)$

Theorem 4.4. *(The convergent properties of DCA-DF.)*

(1) The Algorithm DCA-AF generates the sequence $\{x^l\}$ *such that the sequence* $\{G_2(\boldsymbol{x}^l) - H_2(\boldsymbol{x}^l)\}$ *is monotonously decreasing.*

(2) Every limit point of the sequence $\{\boldsymbol{x}^l\}$ *is a critical point of the problem (4.22), and more strongly, they verify the nesessary local optimality condition* $\partial H_2(\mathbf{x}^*) \subset \partial G_2(\mathbf{x}^*)$.

(3) The series $\{\|\mathbf{x}^{l+1} - \mathbf{x}^l\|^2\}$ *converges.*

Proof. Immediate consequences of the convergence properties of the generic DCA and the facts that H_2 is differentiable and G_2 is strongly convex. □

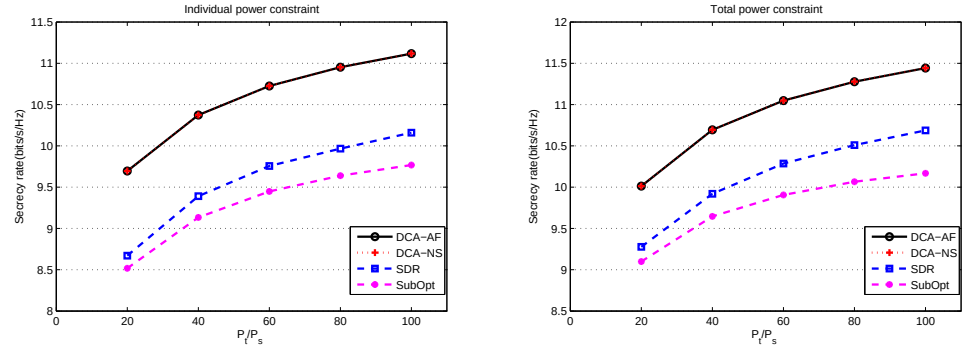
4.4 Experimental Results

In our experiments, all algorithms were implemented in the Matlab 2013b, and performed on a PC Intel Core i5-2500S CPU 2.70GHz of 4GB RAM. We stopped all the DCA schemes with the tolerance $\epsilon = 10^{-5}$.

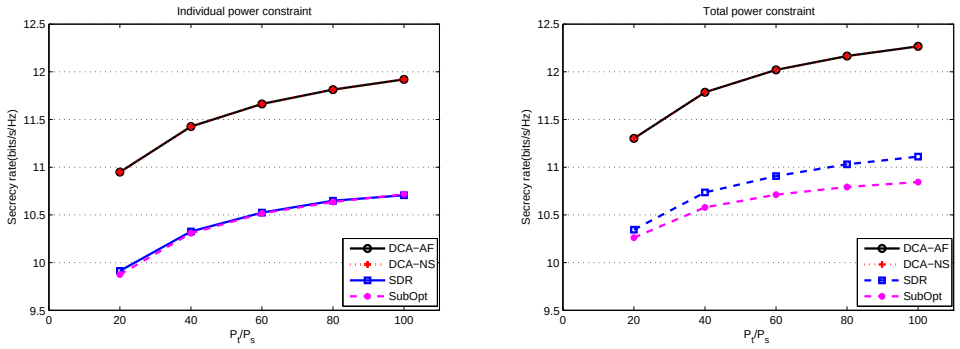
4.4.1 AF Scenario

4.4.1.1 Comparative algorithms

In the AF context, we tested both the DCA based algorithms (DCA-NS and DCA-AF) on some generated datasets and compared them with the semidefinite relaxation technique based approach proposed in [124] (SDR) and the suboptimal method presented in [17] (SubOpt).



(a) $\sigma_h^2 = 1, \sigma_z^2 = 2$.



(b) $\sigma_h^2 = 2, \sigma_z^2 = 2$.

Figure 4.1: Secrecy Rate vs. Pt/Ps in the AF scenario

Table 4.1: The computing time (in seconds) in the AF scenario when $\sigma_h^2 = 1, \sigma_z^2 = 2$.

Pt/Ps	Individual power constraint				Total power constraint			
	DCA-AF	DCA-NS	SDR	SubOpt	DCA-AF	DCA-NS	SDR	SubOpt
20	3.992	**2.761**	11.651	4.294	**0.849**	0.886	10.293	3.444
40	6.435	**3.324**	13.773	4.374	1.561	**1.036**	11.381	3.509
60	9.045	**3.704**	14.746	4.397	2.263	**1.108**	12.612	3.528
80	11.183	**3.926**	15.024	4.456	3.000	**1.220**	12.614	3.554
100	12.868	**4.234**	15.960	4.510	3.712	**1.270**	13.477	3.599

Table 4.2: The computing time (in seconds) in the AF scenario when $\sigma_h^2 = 2, \sigma_z^2 = 2$.

Pt/Ps	Individual power constraint				Total power constraint			
	DCA-AF	DCA-NS	SDR	SubOpt	DCA-AF	DCA-NS	SDR	SubOpt
20	**3.854**	4.015	9.788	4.608	**0.999**	1.131	7.999	3.817
40	5.698	4.880	9.428	**4.678**	1.944	**1.341**	8.194	3.668
60	7.666	5.267	9.113	**4.680**	2.880	**1.451**	8.024	3.651
80	9.215	5.646	8.982	**4.611**	3.780	**1.503**	8.406	3.635
100	10.716	5.779	8.950	**4.650**	4.659	**1.575**	8.480	3.614

4.4.1.2 Experimental setups and numerical results

The channel coefficients $\{f_m\}$ $\{h_m\}$ and $\{z_m\}$ are assumed to be complex, circularly symmetric Gaussian random variables with zero mean and variances σ_f^2, σ_h^2 and σ_z^2, respectively. We tested on two cases of these parameters, those are $\sigma_f^2 = 10, \sigma_h^2 = 1, \sigma_z^2 = 2$ and $\sigma_f^2 = 10, \sigma_h^2 = 2, \sigma_z^2 = 2$. The inequality $\sigma_h^2 \leq \sigma_z^2$ means the quality of users' channel is worse than that of eavesdropper. The other fixed parameters are set to $M = 10, \sigma_m^2 = 1\ \forall m = 1, ..., M, \sigma_0^2 = 1$ for both total and individual power constraint. In the case of individual power constraint, we assume that the relays have equal power budget, i.e., $p_m = \frac{P_T}{M}\ \forall m$.

We tested all algorithms in the AF scenario on 100 independent channel realizations. The average of secrecy rates obtained from all algorithms as well as their average runtime were recorded. In Fig.4.1, we compared the value of secrecy rate achieved by SDR, SubOpt, DCA-AF and DCA-NS versus P_t/P_s, for both total and individual relay power constraint in two different situations of channel condition. The average computing time of all algorithms in all cases are reported in Table 4.1 and Table 4.2.

Some comments on results.
It can be observed from Fig. 4.1 that

- Concerning the secrecy rate, in all cases, the optimal values of secrecy rate obtained by all algorithms show an increasing trend when the ratio Pt/Ps rises. Furthermore, Fig.4.1 indicates that if the channel of user is not worse than that of the eavesdropper (corresponding to the case $\sigma_h^2 = \sigma_z^2 = 2$), the secrecy rates obtained by all algorithms are greater than those achieved by corresponding ones in the remaining case ($\sigma_h^2 = 1, \sigma_z^2 = 2$). Besides, the imposition of power on each relay causes a decrease in the value of secrecy rate in comparison with the total relay power constraint.

More specifically, in all cases the secrecy rates obtained by DCA-AF and DCA-NS are the best, respectively followed by those achieved by SDR and SubOpt. The gaps between secrecy rates gained by DCA-AF and DCA-NS compared with those obtained from the remaining algorithms are significant. DCA-NS though solves the problem on a narrower feasible set compared to the original one, its performances are quite the same as those of DCA-AF and superior to those of SDR and SubOpt, which solve the problem on the original feasible set.

- In terms of the computing time, Table 4.1 and Table 4.2 show that DCA-NS is the best when the total power constraint is imposed. For the case of individual power constraint, DCA-NS and SubOpt are the best and comparable. In all cases, DCA based algorithms are less expensive than SDR meanwhile they achieve the better secrecy rate.

All things considered, it seems reasonable to say that two proposed DCA schemes are the best in terms of both secrecy rate and running time. Between two DCA schemes, DCA-NS runs faster than the other whereas both of them provide quite the same secrecy rates.

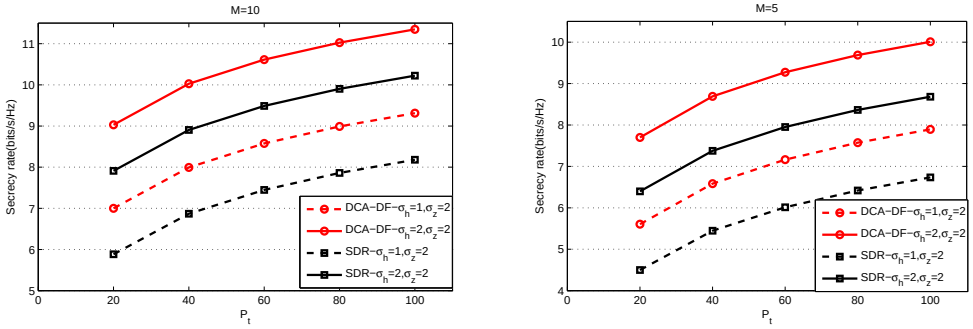


Figure 4.2: Secrecy Rate vs. Pt in the DF scenario

4.4.2 DF Scenario

4.4.2.1 Comparative algorithms

In the DF situation, we compared the proposed DCA scheme with the semidefinite relaxation based method (SDR), which can be solved by a bisection method that leads to solving the sequence of semidefinite programs.

4.4.2.2 Experimental setups and numerical results

Table 4.3: The computing time (in seconds) in the DF scenario.

	Pt	M=5		M=10	
		DCA-DF	SDR	DCA-DF	SDR
$\sigma_h^2 = 1, \sigma_z^2 = 2$	20	**0.236**	5.618	**1.507**	6.869
	40	**0.409**	5.812	**2.461**	6.890
	60	**0.572**	5.928	**3.302**	6.677
	80	**0.726**	5.957	**4.093**	6.788
	100	**0.872**	6.077	**4.751**	6.840
$\sigma_h^2 = 2, \sigma_z^2 = 2$	20	**0.241**	9.408	**0.669**	10.412
	40	**0.416**	9.768	**1.101**	10.637
	60	**0.585**	9.870	**1.503**	10.780
	80	**0.750**	9.974	**1.882**	10.879
	100	**0.909**	10.059	**2.257**	10.940

The number of relays M is set to 5 and 10. The channel coefficients $\{h_m\}$ and $\{z_m\}$ are assumed to be complex, circularly symmetric Gaussian random variables with zero mean and variances σ_h^2, and σ_z^2, respectively. We tested on two cases of these parameters, those are $\sigma_h^2 = 1, \sigma_z^2 = 2$ and $\sigma_h^2 = 2, \sigma_z^2 = 2$ for both cases of M. The remaining fixed parameter is set to $\sigma_0^2 = 1$. It is assumed that the relays have equal power budget, i.e., $p_m = \frac{P_T}{M}\ \forall m$.

Both DCA-DF and SDR were tested on 100 independent channel realizations. The average of secrecy rates obtained from both algorithms and its average runtime were recorded.

In Fig.4.2, we depict the value of secrecy rates achieved by SDR and DCA-DF in the respective cases that the eavesdropper has a stronger channel ($\sigma_z^2 > \sigma_h^2$) and the channel condition are the same for user and eavesdropper ($\sigma_z^2 = \sigma_h^2$) for two cases $M = 5$ and $M = 10$, respectively. Table 4.3 indicates the running time of both DCA-DF and SDR in all cases.

Some comments on results.

- Concerning the secrecy rate, it can be seen from Fig.4.2 that between two different channel conditions (w.r.t $\sigma_h^2 = 1, \sigma_z^2 = 2$ and $\sigma_h^2 = \sigma_z^2 = 2$), the higher secrecy rate obtained when the channel condition of the user is the same as that of the eavesdropper (w.r.t to $\sigma_h^2 = \sigma_z^2 = 2$). General speaking, for both algorithms, the larger number of relays is used, the better secrecy rate is obtained. Similarly, the value of secrecy rate achieved from both algorithms goes up when the power budget of each relay is increasing. DCA-DF always furnishes the better secrecy rates than SDR does in both cases of M and in both different conditions of channel. The gap between two gains of DCA and SDR is around 1 for all cases.

- In terms of the running time, it is indicated in Table 4.3 that DCA-DF runs faster than SDR. The ratio of runtime between DCA and SDR is up to 39 times. In addition, when $M = 5$, the runtime of DCA-DF is below 1 second while that of SDR is ranging from 5.618 to 10.059 seconds. When $M = 10$ and the channel conditions are the same for the user and the eavesdropper, DCA consumes less than 3 seconds whereas SDR spends more than 10 seconds.

To sum up, DCA-DF is better than SDR in both secrecy and runtime aspects.

4.5 Conclusion

In this chapter, we have investigated DC proramming and DCA for maximizing the secrecy rate of a relay beamforming network system subject to total and individual relay power constraints. Two different relaying schemes are considered: Amplify-and-Forward and Decode-and-Forward and three DCA schemes are proposed, the first two treat the model in AF scenario and the third one deals with the model in DF scenaro. The special structure of the feasible set is exploited to develop the efficient DCA schemes which require computing iteratively the projection of points onto an Euclidean ball or an intersection of Euclidean balls that can be explicitly determined. The computational results on several datasets have shown the robustness as well as the efficiency of the proposed DCA schemes in terms of both quality and rapidity. Based on DC programming and DCA, we have explored the more effective relay beamforming strategies to guarantee secrecy for the transmitted information compared with the existing algorithms.

Chapter 5

DC Programming and DCA for Physical Layer Security in a Wireless Relay Network with Multiple Eavesdroppers

Abstract: *In this chapter, we take account of a one-way wireless multi-relay network including one source, one destination and multiple eavesdroppers. To ensure the security for this system, the cooperative AF relay beamforming and CJ techniques are mentioned in which the common aim is to find the optimal beamforming vector to maximize the system secrecy rate under the total and/or individual relay power constraints. The existing methods in the literature for handling such problems are based on a semidefinite relaxation (SDR) technique or adding the appropriate constraint to simplify the original form so that a suboptimal can be found. We introduce herein a novel approach relied on DC (difference of convex functions) programming and DCA (DC algorithms). We first reformulate the considered problems as DC programs and then develop DCA schemes to solve it. Based on the proposed approach, a beamforming vector can be found directly without employing randomization mechanism as in the SDR based previous methods. In addition, the convergence of DCA is ensured while that of SDR-based algorithm have not been shown. Furthermore, to reduce the complexity of the considered problems, we design DCA schemes to address these problems in a special case (say, null-space beamforming) when a complete elimination of the information (noise) leakage to all the eavesdroppers (destination) is assumed, respectively with the AF and CJ scenario. The experimental performance shows that the secrecy rate obtained by the DCA based algorithms are better than those achieved by the existing ones.*

5.1 Introduction and Related Works

Physical layer security has recently attracted much attention of researchers in the field of wireless network security. Various techniques of coding designs and signal processing are exploited and developed in a wide range of communication systems in order to im-

prove their secrecy. Among them, node cooperation techniques are increasingly used in many works and their efficiency in enhancing secrecy is shown. In the node cooperation techniques, one installs external nodes in communication systems to increase network coverage and improve spatial diversity as well as enhance system secrecy. These nodes can play a role either as relays to forward the information transmitted by sources to destinations with two well-known relaying protocols amplify-and-forward (AF) and decode-and-forward (DF) or as friendly jammers to make artificial noise in order to confound eavesdroppers. The cooperative AF/DF relaying and cooperative jamming (CJ) are to refer to such node cooperation techniques, respectively. In addition, these node cooperation techniques are often combined with the beamforming technique at the relays to direct the received information to the intended destinations. An arising issue is how to design appropriate beamforming coefficients at the relays so as to maximize the secrecy rate subject to some power constraints. The various relaying protocols lead to the mathematically different forms of the secrecy rate maximization (SRM) problem. Overall, the SRM problems derived in the DF case are often simpler than those in the AF and CJ cases and the optimal/suboptimal solutions were found in several specific DF schemes ([18], [59]). For the AF scenario, rather than directly dealing with complex programs, one can choose a simpler approach using zero-forcing (ZF) or null-space in which the beamformer is designed to completely eliminate signal at eavesdroppers while maximizing signal at destinations ([18],[127]). For the CJ scenario, to reduce the complexity of the SRM problem, one proposes a null-space scheme in which a beamforming vector is designed to remove noise at destinations while making much confusion at eavesdroppers ([107]). Nevertheless, the solution obtained by this method is only a suboptimal. Besides, many works proposed a two-level algorithm to tackle the difficult model in both the AF and CJ scenarios. This method is a combination of semidefinite relaxation (SDR) technique in the inner level and one-dimension optimization technique in the outer level ([121],[123],[61]). However because the rank-one constraint is discarded, the obtained solution might be not a feasible one to the original problem. To attain a solution to the original problem, one had to employ the randomization techniques to find the rank-one solution and afterwards scaled this solution such that it satisfies constraints. This is only a heuristic search and thus the convergence of this algorithm has not been guaranteed. Recently, some works employ the sequential convex approximation method which is actually a special version of DCA to solve the models in the AF and CJ scenario ([5], [106], [68], [100]).

This chapter gives an extension of the models mentioned in the chapter 4, where multiple eavesdroppers are considered instead of single eavesdropper. More specifically, we consider the SRM problem in a wireless multi-relay system including one source, one destination and multiple eavesdroppers. The AF relay beamforming and CJ techniques are employed to enhance secrecy. These two different cooperative techniques result in two mathematically different SRM problems, which were established in [120] and [18], respectively. These problems are in essence nonconvex, nonsmooth optimization problems, thus hard to solve. In [120], the authors solved the nonconvex SRM problem with respect to the AF scenario by a two-level algorithm based on SDR. In [18], one suboptimal solution was found corresponding to the CJ case. However, these methods, as mentioned before, either have not convergence guarantee or give only a suboptimal

value of secrecy rate. Therefore, we want to explore more efficient methods to deal with these SRM problems. The approach based on DC programming and DCA is an appropriate choice due to the fact that it has been successfully applied to many intractable nonconvex programs in various areas such as communication systems ([106], [128], [104], [36], [4], [53], [50], [52], [93], [94], [95]) and other fields ([48], [49], [55], [56], [78], [79], [80] and references therein). Moreover, the convergence of a generic DCA were completely proved in [45], [78], [51], [47].

Our contributions are summarized as follows.

Firstly, we propose a new approach based on DC programming and DCA for dealing with the SRM problem in the AF and CJ scenario, in the presence of multiple eavesdroppers. We design DCAs for both general and null-space schemes. Not only the standard DCA but also the general DCA are developed in this chapter. It should be noted that the standard DCA has been exploited and successfully applied for solving nonconvex optimization problems in various areas of applied science for many years. Meanwhile the general DCA, which is generalized from the standard DCA, is only studied in recent years. General DCAs permit to solve a wider class of nonconvex optimization problems compared to standard DCAs, thus being a promising nonconvex optimization tool. The simulation results imply that the secrecy rate obtained by DCA based algorithms are considerably better than those gained by the existing ones. For the null-space scheme in the AF scenario, DCA may give the global solution although it is a local approach.

Secondly, the convergence property of the proposed general DCA schemes are thoroughly proved. We adapt the proof of convergence theorem in [47] for our situation that is quite different from the one mentioned in [47]. More particularly, [47] presented the convergence of a generic general DCA scheme in which a slack variable was introduced and penalized to the objective function when solving the resulting subproblems and the update of penalty coefficients was required. Meanwhile, in our DCA schemes, we are no need to use slack variables or update penalty coefficients. Therefore, the proof of our theorem is simplified compared with the one presented in [47].

The rest of this chapter is organized as follows. In Section 5.2, we describe the considered SRM problems in the CJ and AF relay beamforming design. The solution method is presented in Section 5.3, in which we show how to apply DC programming and DCA to solve the considered problems. Experimental results are reported in Section 5.4. Finally, Section 5.5 concludes the chapter.

5.2 The Secrecy Rate Maximization Problem in Physical Layers

5.2.1 Secrecy rate maximization via amplify-and-forward relay beamforming

In this section, the model mentioned in [120] is taken into account. In what follows, we restate this model. Consider a system consisting of a source, a destination, M relays and K eavesdroppers. Each node is equipped with a single antenna. It is supposed that there are neither direct link between the source and the destination nor direct links between the source and the eavesdroppers. White complex Gaussian noise at each node is assumed to has zero mean and variance σ^2 , i.e., $\mathcal{CN}(0, \sigma^2)$. In this system, the source transmits information to the destination in cooperation with M relays. In the first step, the source broadcasts its information and the relays get the signals given by:

$$\mathbf{y}_R = [y_{r,1}, ..., y_{r,M}]^T = \sqrt{P_s}\mathbf{f}s + \mathbf{n}_R,$$

where $y_{r,m}$ is the signal received at mth relay, P_s is the average transmit power at the source, s is the symbol sent by the source with power $E(|s|^2) = 1$; $\mathbf{f} = [f_1, ..., f_M]^T$ is the vector of channel coefficients between the source and the relays, and $\mathbf{n}_R \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_M)$ is the vector of noise at the relays.

In the second step, the relays forward the received signal by using the AF cooperative beamforming strategy. In more detail, the received signal at kth relay is first multiplied by a complex weight w_k before being sent to the destination. Denote $\mathbf{w} = [w_1, ..., w_M]^T$ as the cooperative beamforming vector. As a result, the signals that the relays transmit are $\mathbf{D}(\mathbf{y}_R)\mathbf{w}$, where $\mathbf{D}(\mathbf{y}_R) = \text{diag}(\mathbf{y}_R)$ is a diagonal matrix with the main diagonal $\mathbf{y}_R$. The beamforming vector is chosen such that these transmitted signals satisfy the total and individual power constraints.

$$E(\|\mathbf{D}(\mathbf{y}_R)\mathbf{w}\|^2) = \mathbf{w}^\dagger\mathbf{C}\mathbf{w} \leq P_{tot},$$

$$E(|y_{r,m}w_m|^2) = \mathbf{e}_m^T\mathbf{C}\mathbf{w}\mathbf{w}^\dagger\mathbf{e}_m \leq P_m \ \forall m = 1, ..., M,$$

where $\mathbf{C} = P_s\mathbf{D}(\mathbf{f})^\dagger\mathbf{D}(\mathbf{f}) + \sigma^2\mathbf{I}_M$ and $\mathbf{e}_m$ is the mth column of the identity matrix with size of M.

Finally, the destination and eavesdropper j get the following signals, respectively

$$y_d = \sqrt{P_s}\mathbf{g}^\dagger\mathbf{D}(\mathbf{f})\mathbf{w}s + \mathbf{n}_R^T\mathbf{D}(\mathbf{g})^\dagger\mathbf{w} + n_d,$$
$$y_{e,j} = \sqrt{P_s}\mathbf{h}_j^\dagger\mathbf{D}(\mathbf{f})\mathbf{w}s + \mathbf{n}_R^T\mathbf{D}(\mathbf{h}_j)^\dagger\mathbf{w} + n_{e,j},$$

where $\mathbf{g} = [g_1, ..., g_M]^\dagger, \mathbf{h}_j = [h_{1,j}, ..., h_{M,j}]^\dagger, j = 1, .., K$, are vectors of channel coefficients between relays and destination, and between relays and eavesdropers, respectively and $n_d, n_{e,j}$ are the noises at the destination and the jth eavesdropper with noise variance σ^2.

The SNR at the destination and the jth eavesdropper are given by

$$\begin{aligned}\gamma_d &= \frac{\mathbf{w}^\dagger \mathbf{A}\mathbf{w}}{1+\mathbf{w}^\dagger \mathbf{G}\mathbf{w}},\\ \gamma_{e,j} &= \frac{\mathbf{w}^\dagger \mathbf{B}_j\mathbf{w}}{1+\mathbf{w}^\dagger \mathbf{H}_j\mathbf{w}}\ \forall j=1,...,K,\end{aligned}$$

where $\mathbf{A} = \frac{P_s}{\sigma^2}\mathbf{D}(\mathbf{f})^\dagger \mathbf{g}\mathbf{g}^\dagger \mathbf{D}(\mathbf{f}), \mathbf{G} = \mathbf{D}(\mathbf{g})\mathbf{D}(\mathbf{g})^\dagger, \mathbf{B}_j = \frac{P_s}{\sigma^2}\mathbf{D}(\mathbf{f})^\dagger \mathbf{h}_j\mathbf{h}_j^\dagger \mathbf{D}(\mathbf{f}), \mathbf{H}_j = \mathbf{D}(\mathbf{h}_j)\mathbf{D}(\mathbf{h}_j)^\dagger$.
An achievable secrecy rate is given by ([120])

$$R_s = \frac{1}{2}\left(\min_{j=1,...,K}[\log_2(1+\gamma_d)-\log_2(1+\gamma_{e,j})]\right)$$

The aim of this model is to maximize the secrecy rate under both the total and individual power constraints at the relays, i.e.,

$$\begin{aligned}\max_{\mathbf{w}}\min_{j=1,...,K}\quad & [\log_2(1+\gamma_d)-\log_2(1+\gamma_{e,j})] \qquad (5.1)\\ \text{s.t}\quad & \mathbf{w}^\dagger\mathbf{C}\mathbf{w}\le P_{tot},\\ & \mathbf{e}_m^T\mathbf{C}\mathbf{w}\mathbf{w}^\dagger\mathbf{e}_m \le P_m\ \forall m=1,...,M.\end{aligned}$$

It is difficult to deal with this problem because of its nonconvex and nonsmooth property. The efforts to solve this problem were based on a semidefinite relaxation technique. However, the obtained solution was not ensured to be a feasible solution because the rank-one constraint was ignored. A randomization technique was proposed to find a feasible solution from the relaxed one, but it is only a heuristic search, thus the convergence of the algorithm is not guaranteed.

5.2.2 Secrecy Rate Maximization via Cooperative Jamming

In this section, we reconsider the model proposed in [18]. The system is comprised of a source, a destination, M relays and K eavesdroppers. Each node is equipped with a single antenna. In Cooperative Jamming (CJ), the relays play a role as jammers which transmit a weighted version of a jamming signal z to the channel with the aim of confusing the eavesdroppers, whereas the source sends the signal $\sqrt{P_s}x$ to the channel. Denote $h_{SD}^* \in \mathbb{C}$ as the channel coefficient between the source and the destination, $\mathbf{h}_{SE}^* \in \mathbb{C}^K$ as the vector of channel coefficients between the source and K eavesdroppers, $\mathbf{h}_{RD}^* \in \mathbb{C}^M$ as the channel vector between M relays and the destination, $\mathbf{H}_{RE}^*$ as the $M \times K$ matrix of channel coefficients between M relays and K eavesdroppers. Denote P_{tot} as the total transmit power budget of all relays and $\mathbf{w}$ as a vector of relay weights.

The received signal at the destination is

$$y_d = \sqrt{P_s}h_{SD}^* x + \mathbf{h}_{RD}^\dagger \mathbf{w} z + n_d$$

and the received signals at the eavesdroppers are given by

$$\mathbf{y}_e = \sqrt{P_s}\mathbf{h}_{SE}^* x + \mathbf{H}_{RE}^\dagger \mathbf{w} z + \mathbf{n}_e,$$

where n_d represents complex Gaussian noise at the destination with variance of σ^2, and $\mathbf{n}_e \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_K)$ is a noise vector at the K eavesdroppers. Therefore, the achievable rate at the destination is

$$R_d = \log_2 \left(1 + \frac{P_s |h_{SD}|^2}{\mathbf{w}^\dagger \mathbf{R}_{RD} \mathbf{w} + \sigma^2}\right)$$

and the achievable rate at the jth eavesdropper is

$$R_{e_j} = \log_2 \left(1 + \frac{P_s |\mathbf{h}_{SE}(j)|^2}{\mathbf{w}^\dagger \mathbf{R}_{RE_j} \mathbf{w} + \sigma^2}\right),$$

where $\mathbf{R}_{RD} = \mathbf{h}_{RD}\mathbf{h}_{RD}^\dagger, \mathbf{R}_{RE_j} = \mathbf{H}_{RE}(:, j)\mathbf{H}_{RE}(:, j)^\dagger$.

The problem of achievable secrecy rate maximization can be formulated as below ([18]).

$$\begin{aligned} \max_{\mathbf{w}} \min_{j=1,\ldots,K} \quad & \left[\log_2 \left(1 + \frac{P_s |h_{SD}|^2}{\mathbf{w}^\dagger \mathbf{R}_{RD} \mathbf{w} + \sigma^2}\right) - \log_2 \left(1 + \frac{P_s |\mathbf{h}_{SE}(j)|^2}{\mathbf{w}^\dagger \mathbf{R}_{RE_j} \mathbf{w} + \sigma^2}\right)\right] \quad (5.2) \\ \text{s.t} \quad & \mathbf{w}^\dagger \mathbf{w} \leq P_{tot}. \end{aligned}$$

This problem is nonsmooth and nonconvex and thus it is intractable. The existing method only solved this problem in a special situation when the relay weights were imposed such that the jamming signal was altogether eliminated at the destination. In such a case, this problem becomes simpler compared with the initial form. However, the closed-form solution has not been indicated yet and only a suboptimal solution was provided.

In this chapter, we will investigate DC programming and DCA for solving both nonconvex programs (5.1) and (5.2).

5.3 Solution Methods Based on DC programming and DCA

5.3.1 DC Programming and DCA for solving (5.1)

In this section, we propose two DCA schemes to address the problem (5.1). The first DCA is to solve the problem (5.1) in a special situation where the signal on the eavesdroppers' channel is assumed to be completely eliminated, called the null-space AF relay beamforming design. It means that $\mathbf{w}$ is designed to satisfy the constraints

$\mathbf{h}_k^\dagger \mathbf{D(f)w} = 0\ \forall k = 1, ..., K$. These constraints make the second term of the objective function in (5.1) become zero, thus the problem (5.1) is simplified and easier to deal with. This problem now is actually a standard DC program, therefore it can be addressed by a standard DCA. Nevertheless, the feasible set is narrowed due to the additional constraints, thus the obtained solution might be only a suboptimal. To seek the best solution possible, the second DCA is designed to solve the problem (5.1) without imposing any additional constraints. In this general case, the problem (5.1) can be recast as a general DC program where both the objective function and some constraints are DC. This is the most difficult form in DC programming and thus more problematic than the standard DC program in the null space AF relay beamforming design. The more details of these two DCA schemes are presented in the followings.

5.3.1.1 The null-space AF relay beamforming design

When there is an assumption of no information leakage to all the eavesdroppers, in other words the beamforming vector lies in the null space of the equivalent channel of the relay link from the source to the eavesdroppers, the problem (5.1) then is simplified to the following optimization one

$$\begin{aligned} \min_{\mathbf{w}} \quad & \left[-\log_2 \left(\frac{1 + \mathbf{w}^\dagger (\mathbf{A}+\mathbf{G}) \mathbf{w}}{1 + \mathbf{w}^\dagger \mathbf{G} \mathbf{w}} \right) \right] \\ \text{s.t} \quad & \mathbf{w}^\dagger \mathbf{C} \mathbf{w} \leq P_{tot}, \\ & \mathbf{e}_m^T \mathbf{C} \mathbf{w} \mathbf{w}^\dagger \mathbf{e}_m \leq P_m\ \forall m = 1, ..., M. \\ & \mathbf{h}_k^\dagger \mathbf{D(f)w} = 0\ \forall k = 1, ..., K. \end{aligned} \tag{5.3}$$

By denoting $\mathbf{x} = \left[\text{Re}(\mathbf{w}^T) \quad \text{Im}(\mathbf{w}^T) \right]^T$, $\mathbf{T}_1 = \begin{bmatrix} \text{Re}(\mathbf{A}+\mathbf{G}) & -\text{Im}(\mathbf{A}+\mathbf{G}) \\ \text{Im}(\mathbf{A}+\mathbf{G}) & \text{Re}(\mathbf{A}+\mathbf{G}) \end{bmatrix}$, $\mathbf{M}_1 = \begin{bmatrix} \text{Re}(\mathbf{G}) & -\text{Im}(\mathbf{G}) \\ \text{Im}(\mathbf{G})] & \text{Re}(\mathbf{G}) \end{bmatrix}$, $\mathbf{C}_1 = \begin{bmatrix} \text{Re}(\mathbf{C}) & -\text{Im}(\mathbf{C}) \\ \text{Im}(\mathbf{C}) & \text{Re}(\mathbf{C}) \end{bmatrix}$, $\mathbf{E}_k = \begin{bmatrix} \text{Re}(\mathbf{h}_k^\dagger \mathbf{D(f)}] & -\text{Im}(\mathbf{h}_k^\dagger \mathbf{D(f)}) \\ \text{Im}(\mathbf{h}_k^\dagger \mathbf{D(f)})] & \text{Re}(\mathbf{h}_k^\dagger \mathbf{D(f)}) \end{bmatrix}, k = 1, \ldots, K$, the above problem can be equivalently recast in the real form as below

$$\begin{aligned} \min_{\mathbf{x}} \quad & \left[-\ln \left(\frac{1 + \mathbf{x}^T \mathbf{T}_1 \mathbf{x}}{1 + \mathbf{x}^T \mathbf{M}_1 \mathbf{x}} \right) \right] \\ \text{s.t} \quad & \mathbf{x}^T \mathbf{C}_1 \mathbf{x} \leq P_{tot}, \\ & \mathbf{e}_m^T \mathbf{C}_1 \mathbf{x} \mathbf{x}^T \mathbf{e}_m + \mathbf{e}_{M+m}^T \mathbf{C}_1 \mathbf{x} \mathbf{x}^T \mathbf{e}_{M+m} \leq P_m\ \forall m = 1, ..., M, \\ & \mathbf{E}_k \mathbf{x} = 0\ \forall k = 1, ..., K. \end{aligned} \tag{5.4}$$

The feasible set of this problem is clearly convex while its objective function is nonconvex. However, this objective function can be expressed as a difference of two convex functions $g(\mathbf{x}) = \frac{1}{2}\tau \|\mathbf{x}\|^2$ and $h(\mathbf{x}) = \frac{1}{2}\tau \|\mathbf{x}\|^2 + \ln(1 + \mathbf{x}^T \mathbf{T}_1 \mathbf{x}) - \ln(1 + \mathbf{x}^T \mathbf{M}_1 \mathbf{x})$. It is straightforward to deduce from Proposition 2.1 that, the parameter τ should be chosen

as the greatest eigenvalue of the matrix $\frac{\mathbf{T}_1}{2} + 2\mathbf{M}_1$ to guarantee the convexity of h. Therefore, (5.4) is a standard DC program and its DC formulation is given by

$$\begin{aligned} \min_{\mathbf{x}} \quad & g(\mathbf{x}) - h(\mathbf{x}) \\ \text{s.t} \quad & \mathbf{x}^T\mathbf{C}_1\mathbf{x} \leq P_{tot}, \\ & \mathbf{e}_m^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_m + \mathbf{e}_{M+m}^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_{M+m} \leq P_m \ \forall m = 1, ..., M, \\ & \mathbf{E}_k\mathbf{x} = 0 \ \forall k = 1, ..., K. \end{aligned} \tag{5.5}$$

Following the generic DCA scheme, it requires computing at the kth iteration $\mathbf{u}^k \in \partial h(\mathbf{x}^k)$ and then solving the following convex subproblem to obtain $\mathbf{x}^{k+1}$.

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\tau\|\mathbf{x}\|^2 - \langle \mathbf{u}^k, \mathbf{x}\rangle \\ \text{s.t} \quad & \mathbf{x}^T\mathbf{C}_1\mathbf{x} \leq P_{tot}, \\ & \mathbf{e}_m^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_m + \mathbf{e}_{M+m}^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_{M+m} \leq P_m \ \forall m = 1, ..., M, \\ & \mathbf{E}_k\mathbf{x} = 0 \ \forall k = 1, ..., K. \end{aligned} \tag{5.6}$$

Because h is differentiable, its gradient at a point $\mathbf{x}^k$ is given by

$$\mathbf{u}^k = \partial h(\mathbf{x}^k) = \nabla h(\mathbf{x}^k) = \left(\tau\mathbf{x}^k + \frac{2\mathbf{T}_1\mathbf{x}^k}{1 + (\mathbf{x}^k)^T\mathbf{T}_1\mathbf{x}^k} - \frac{2\mathbf{M}_1\mathbf{x}^k}{1 + (\mathbf{x}^k)^T\mathbf{M}_1\mathbf{x}^k}\right).$$

The DCA scheme for solving the problem (5.5), namely DCA-AF-NS, is described as below.

DCA-AF-NS scheme

Initialization: choose $\mathbf{x}^0 \in \mathbb{R}^{2M}$ as an initial guess, set a tolerance ϵ for DCA-AF-NS, $k \leftarrow 0$.

Repeat

- Compute $\mathbf{x}^{k+1}$ by solving the subproblem (5.6).
- $k \leftarrow k + 1$.

Until $\left(\frac{\|\mathbf{x}^k - \mathbf{x}^{k-1}\|}{1+\|\mathbf{x}^{k-1}\|} < \epsilon \text{ or } \frac{|f(\mathbf{x}^k) - f(\mathbf{x}^{k-1})|}{1+|f(\mathbf{x}^{k-1})|} < \epsilon\right)$ where $f(\mathbf{x}^k) = -\ln\left(\frac{1+(\mathbf{x}^k)^T\mathbf{T}_1\mathbf{x}^k}{1+(\mathbf{x}^k)^T\mathbf{M}_1\mathbf{x}^k}\right)$.

Since the objective function of (5.5) is continuous and its feasible set is compact, the optimal value of (5.5) is finite and the sequences $\{\mathbf{x}^k\}$ and $\{\mathbf{u}^k\}$ generated from DCA-AF-NS are bounded. In addition, h is differentiable and g is strongly convex. Therefore, according to the convergence properties of DCA presented in Section 1.2.1.2, it is straightforward to obtain the convergence theorem of the algorithm DCA-AF-NS as follows.

Theorem 5.1.

(1) The sequence $\{g(\boldsymbol{x}^k) - h(\boldsymbol{x}^k)\}$ is monotonously decreasing.

(2) Every limit point $\boldsymbol{x}^*$ of the sequence $\{\boldsymbol{x}^k\}$ is a critical point of the problem (5.5) and more strongly, they verify the necessary local optimality condition $\partial h(\boldsymbol{x}^*) \subset \partial g(\boldsymbol{x}^*)$.

(3) The series $\{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2\}$ converges.

As mentioned before, the null-space scheme helps simplify the original problems, thus the computation of DCA corresponding to this scheme is diminished. However, it might only give the suboptimal of secrecy rate because of the complementary constraint. In the next section, we will design a DCA scheme to directly handle the original problem(5.1) with the expectation of finding a better solution. As a natural choice, we choose the solution obtained by DCA-AF-NS as an initial point for the DCA designed for the original problem.

5.3.1.2 The general AF relay beamforming design

The problem (5.1) can be recast as follows

$$\begin{aligned} \max_{\mathbf{w}} \quad & \left[\log_2\left(\frac{1+\mathbf{w}^\dagger(\mathbf{A}+\mathbf{G})\mathbf{w}}{1+\mathbf{w}^\dagger\mathbf{G}\mathbf{w}}\right) - \max_{j=1,\dots,K}\log_2\left(\frac{1+\mathbf{w}^\dagger(\mathbf{B}_j+\mathbf{H}_j)\mathbf{w}}{1+\mathbf{w}^\dagger\mathbf{H}_j\mathbf{w}}\right)\right] \\ \text{s.t} \quad & \mathbf{w}^\dagger\mathbf{C}\mathbf{w} \le P_{tot}, \\ & \mathbf{e}_m^T\mathbf{C}\mathbf{w}\mathbf{w}^\dagger\mathbf{e}_m \le P_m \ \forall m = 1,...,M. \end{aligned} \tag{5.7}$$

By introducing a slack variable t, the above problem is equivalent to the following one

$$\begin{aligned} \min_{\mathbf{w},t} \quad & \left[-\ln\left(\frac{1+\mathbf{w}^\dagger(\mathbf{A}+\mathbf{G})\mathbf{w}}{1+\mathbf{w}^\dagger\mathbf{G}\mathbf{w}}\right) + t\right] \\ \text{s.t} \quad & \mathbf{w}^\dagger\mathbf{C}\mathbf{w} \le P_{tot}, \\ & \mathbf{e}_m^T\mathbf{C}\mathbf{w}\mathbf{w}^\dagger\mathbf{e}_m \le P_m \ \forall m = 1,...,M, \\ & \ln\frac{1+\mathbf{w}^\dagger(\mathbf{B}_j+\mathbf{H}_j)\mathbf{w}}{1+\mathbf{w}^\dagger\mathbf{H}_j\mathbf{w}} \le t \ \forall j = 1,...,K. \end{aligned} \tag{5.8}$$

Denote $\mathbf{D}_0 = \begin{bmatrix} \mathrm{Re}(\mathbf{A}+\mathbf{G}) & -\mathrm{Im}(\mathbf{A}+\mathbf{G}) \\ \mathrm{Im}(\mathbf{A}+\mathbf{G}) & \mathrm{Re}(\mathbf{A}+\mathbf{G}) \end{bmatrix}, \mathbf{N}_0 = \begin{bmatrix} \mathrm{Re}(\mathbf{G}) & -\mathrm{Im}(\mathbf{G}) \\ \mathrm{Im}(\mathbf{G}) & \mathrm{Re}(\mathbf{G}) \end{bmatrix}$,
$\mathbf{C}_1 = \begin{bmatrix} \mathrm{Re}(\mathbf{C}) & -\mathrm{Im}(\mathbf{C}) \\ \mathrm{Im}(\mathbf{C}) & \mathrm{Re}(\mathbf{C}) \end{bmatrix}, \mathbf{N}_j = \begin{bmatrix} \mathrm{Re}(\mathbf{B}_j+\mathbf{H}_j) & -\mathrm{Im}(\mathbf{B}_j+\mathbf{H}_j) \\ \mathrm{Im}(\mathbf{B}_j+\mathbf{H}_j) & \mathrm{Re}(\mathbf{B}_j+\mathbf{H}_j) \end{bmatrix}$,
$\mathbf{D}_j = \begin{bmatrix} \mathrm{Re}(\mathbf{H}_j)] & -\mathrm{Im}(\mathbf{H}_j) \\ \mathrm{Im}(\mathbf{H}_j) & \mathrm{Re}(\mathbf{H}_j) \end{bmatrix}, j = 1,\dots,K, \ \mathbf{x} = \begin{bmatrix} \mathrm{Re}(\mathbf{w}^T) & \mathrm{Im}(\mathbf{w}^T) \end{bmatrix}^T$.
The problem (5.8) can be equivalently converted to a real variable form as follows

$$\min_{\mathbf{x},t} \quad \left[\ln\left(\frac{1+\mathbf{x}^T\mathbf{N}_0\mathbf{x}}{1+\mathbf{x}^T\mathbf{D}_0\mathbf{x}}\right) + t\right] \tag{5.9}$$

$$\text{s.t} \quad \mathbf{x}^T\mathbf{C}_1\mathbf{x} \le P_{tot}, \tag{5.10}$$

$$\mathbf{e}_m^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_m + \mathbf{e}_{M+m}^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_{M+m} \le P_m \ \forall m = 1,...,M, \tag{5.11}$$

$$\ln\frac{1+\mathbf{x}^T\mathbf{N}_j\mathbf{x}}{1+\mathbf{x}^T\mathbf{D}_j\mathbf{x}} \le t \ \ \forall j = 1,...,K. \tag{5.12}$$

Denote

$$S = \left\{ \begin{array}{ll} \mathbf{x}: & \mathbf{x}^T\mathbf{C}_1\mathbf{x} \leq P_{tot}, \\ & \mathbf{e}_m^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_m + \mathbf{e}_{M+m}^T\mathbf{C}_1\mathbf{x}\mathbf{x}^T\mathbf{e}_{M+m} \leq P_m \ \forall m = 1, ..., M \end{array} \right\}.$$

It is obvious that S is a convex set because the matrix $\mathbf{C}_1$ is positive semidefinite. The objective function of (5.9), namely $F(\mathbf{x}, t)$, can be written in the form $F(\mathbf{x}, t) = G_0(\mathbf{x}) + t - H_0(\mathbf{x})$, where $G_0(\mathbf{x}) = \frac{1}{2}\rho_0\|\mathbf{x}\|^2$ and $H_0(\mathbf{x}) = \frac{1}{2}\rho_0\|\mathbf{x}\|^2 - \ln(1 + \mathbf{x}^T\mathbf{N}_0\mathbf{x}) + \ln(1 + \mathbf{x}^\mathbf{T}\mathbf{D}_0\mathbf{x})$ for some $\rho_0 > 0$. Similarly, the left side of the constraint (5.12) can be expressed as a difference of two functions $G_j(\mathbf{x}) = \frac{1}{2}\rho_j\|\mathbf{x}\|^2$ and $H_j(\mathbf{x}) = \frac{1}{2}\rho_j\|\mathbf{x}\|^2 - \ln(1 + \mathbf{x}^T\mathbf{N}_j\mathbf{x}) + \ln(1 + \mathbf{x}^T\mathbf{D}_j\mathbf{x})$ for some $\rho_j > 0, j = 1, ..., K$. It is easy to verify that for any $\rho_j > 0$, $G_j(\mathbf{x})$ is convex in $\mathbf{x}$ $\forall j = 0, \ldots, K$. However, for each $j = 0, \ldots, K$ the functions $H_j(\mathbf{x})$ is only convex if ρ_j is large enough. The following proposition indicates a sufficient condition for ρ_j to ensure the convexity of $H_j(\mathbf{x}), j = 0, \ldots, K$.

Proposition 5.1. *For each $j = 0, \ldots, K$, if ρ_j is greater than the largest eigenvalue of the matrix $\frac{\boldsymbol{D}_j}{2} + 2\boldsymbol{N}_j$ then the function $H_j(\boldsymbol{x})$ is convex in $\boldsymbol{x}$.*

The proof of this proposition is straightforwardly deduced from Proposition 2.1.

As a consequence, when $\rho_j, j = 0, \ldots, K$ satisfying the condition indicated in the above proposition, we obtain the DC formulation of the problem (5.9) as below.

$$\begin{array}{ll} \min\limits_{\mathbf{x},t} & [G_0(\mathbf{x}) + t] - H_0(\mathbf{x}) \\ \text{s.t} & \mathbf{x} \in S, \\ & G_j(\mathbf{x}) - H_j(\mathbf{x}) \leq t \ \ \forall j = 1, ..., K. \end{array} \tag{5.13}$$

DCA applied to (5.13) involves computing at the kth iteration the sequences $\{(\mathbf{x}^k, t^k)\}$ and $\{\mathbf{z}_j^k\}$ such that

$$\mathbf{z}_j^k \in \partial H_j(\mathbf{x}^k) \ j = 0, \ldots, K,$$

and $(\mathbf{x}^{k+1}, t^{k+1})$ solves the convex subproblem below

$$\begin{array}{ll} \min\limits_{\mathbf{x},t} & \frac{1}{2}\rho_0\|\mathbf{x}\|^2 + t - \langle \mathbf{z}_0^k, \mathbf{x}\rangle \qquad (5.14) \\ \text{s.t} & \mathbf{x} \in S, \\ & \frac{1}{2}\rho_j\|\mathbf{x}\|^2 - H_j(\mathbf{x}^k) - \langle \mathbf{z}_j^k, \mathbf{x} - \mathbf{x}^k\rangle \leq t \ \ \forall j = 1, ..., K. \qquad (5.15) \end{array}$$

Obviously the functions $H_j(\mathbf{x}), j = 0, \ldots, K$ are differentiable and their gradient at $\mathbf{x}^k$ are computed as

$$\mathbf{z}_j^k = \left(\rho_j\mathbf{x}^k - \frac{2\mathbf{N}_j\mathbf{x}^k}{1 + (\mathbf{x}^k)^T\mathbf{N}_j\mathbf{x}^k} + \frac{2\mathbf{D}_j\mathbf{x}^k}{1 + (\mathbf{x}^k)^T\mathbf{D}_j\mathbf{x}^k} \right), j = 0, ..., K.$$

The general DCA scheme for solving (5.13), namely DCA-AF, can be summarized as below.

DCA-AF: The general DCA scheme for the problem (5.13)

Initialization: Denote $(\mathbf{x}^0, t^0)$ as the solution obtained by DCA-AF-NS. Choose $\mathbf{V}^0 = (\mathbf{x}^0, t^0)$ as an initial guess, set a tolerance ϵ for DCA-AF, $k \leftarrow 0$.
Repeat
• Compute $\mathbf{V}^{k+1} = (\mathbf{x}^{k+1}, t^{k+1})$ by solving the subproblem (5.14).
• $k \leftarrow k + 1$.
Until $\left(\frac{\|\mathbf{V}^k - \mathbf{V}^{k-1}\|}{1+\|\mathbf{V}^{k-1}\|} < \epsilon \text{ or } \frac{|F(\mathbf{V}^k) - F(\mathbf{V}^{k-1})|}{1+|F(\mathbf{V}^{k-1})|} < \epsilon \right)$ where $F(\mathbf{V}) = \ln\left(\frac{1+\mathbf{x}^T\mathbf{N}_0\mathbf{x}}{1+\mathbf{x}^T\mathbf{D}_0\mathbf{x}}\right) + t$.

The following theorem shows the convergence of the above DCA scheme.

Theorem 5.2.

(1) DCA-AF generates the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *such that the sequence of the corresponding objective function values* $\{F(\boldsymbol{V}^k)\}$ *is decreasing.*

(2) Every limit point of the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *generated by DCA-AF is a critical point to the problem (5.13).*

Proof.

(1) Because $\mathbf{V}^{k+1} = (\mathbf{x}^{k+1}, t^{k+1})$ is a minimizer of (5.14) and $\mathbf{V}^k = (\mathbf{x}^k, t^k)$ is a feasible point of (5.14), thus

$$\frac{1}{2}\rho_0\|\mathbf{x}^{k+1}\|^2 + t^{k+1} - H_0(\mathbf{x}^k) - \langle \mathbf{y}^k, \mathbf{x}^{k+1} - \mathbf{x}^k \rangle \leq \frac{1}{2}\rho_0\|\mathbf{x}^k\|^2 + t^k - H_0(\mathbf{x}^k) - \langle \mathbf{y}^k, \mathbf{x}^k - \mathbf{x}^k \rangle.$$

Due to the convexity of H_0, the left side of the above inequality is greater than or equal to $F(\mathbf{V}^{k+1})$ whereas the right side actually equals to $F(\mathbf{V}^k)$. Therefore $F(\mathbf{V}^{k+1}) \leq F(\mathbf{V}^k)$.

(2) It is obvious that S is a nonempty compact convex set. Because the left sides in (5.15) do not depend on t, it is easy to verify that the Slater's constraint qualification is satisfied. Furthermore $(\mathbf{x}^{k+1}, t^{k+1})$ is the optimal solution to the problem (5.14), therefore there exist some $\lambda_j^{k+1} \in \mathbb{R}, j = 1, ..., K$ such that

$$\bullet\ 0 \in \nabla G_0(\mathbf{x}^{k+1}) - \nabla H_0(\mathbf{x}^k) + \sum_{j=1}^{K} \lambda_j^{k+1}(\nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k)) + N(S, \mathbf{x}^{k+1}), \tag{5.16}$$

$$\bullet\ 1 - \sum_{j=1}^{K} \lambda_j^{k+1} = 0, \quad \mathbf{x}^{k+1} \in S, \tag{5.17}$$

$$\bullet\ G_j(\mathbf{x}^{k+1}) - H_j(\mathbf{x}^k) - \langle \nabla H_j(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle \leq t^{k+1},\ \lambda_j^{k+1} \geq 0\ \forall j = 1, ..., K, \tag{5.18}$$

$$\bullet\ \lambda_j^{k+1}\left[G_j(\mathbf{x}^{k+1}) - H_j(\mathbf{x}^k) - \langle \nabla H_j(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle - t^{k+1}\right] = 0\ \forall j = 1, ..., K. \tag{5.19}$$

Because $G_0(\mathbf{x}) = \frac{1}{2}\rho_0\|\mathbf{x}\|^2$ is strongly convex in $\mathbf{x}$ with module ρ_0 and $H_0(\mathbf{x})$ is convex in $\mathbf{x}$, we obtain

$$G_0(\mathbf{x}^k) \geq G_0(\mathbf{x}^{k+1}) + \langle \nabla G_0(\mathbf{x}^{k+1}), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\rho_0}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2,$$
$$H_0(\mathbf{x}^{k+1}) \geq H_0(\mathbf{x}^k) + \langle \nabla H_0(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k\rangle.$$

Adding these two inequalities together it leads to

$$\langle \nabla G_0(\mathbf{x}^{k+1}) - \nabla H_0(\mathbf{x}^k), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \leq F_0(\mathbf{x}^k) - F_0(\mathbf{x}^{k+1}) - \frac{\rho_0}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2, \quad (5.20)$$

where $F_0(\mathbf{x}^k) = G_0(\mathbf{x}^k) - H_0(\mathbf{x}^k)$.

Since $G_j(\mathbf{x}) = \frac{1}{2}\rho_j\|\mathbf{x}\|^2, \nabla^2 G_j(\mathbf{x}) = \rho_j I, \forall\ j = 1, ..., K$ which shows that G_j is strongly convex in $\mathbf{x}$ with modulus ρ_j. Because of this property we have

$$G_j(\mathbf{x}^k) \geq G_j(\mathbf{x}^{k+1}) + \langle \nabla G_j(\mathbf{x}^{k+1}), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\rho_j}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2\ \forall j = 1, ..., K$$
$$\Leftrightarrow G_j(\mathbf{x}^k) - H_j(\mathbf{x}^k) + \langle \nabla H_j(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k\rangle \geq$$
$$G_j(\mathbf{x}^{k+1}) - H_j(\mathbf{x}^k) + \langle \nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\rho_j}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2$$
$$\Leftrightarrow \langle \nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \leq$$
$$F_j(\mathbf{x}^k) - (G_j(\mathbf{x}^{k+1}) - H_j(\mathbf{x}^k) - \langle \nabla H_j(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k\rangle) - \frac{\rho_j}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2,$$

where $F_j(\mathbf{x}) = G_j(\mathbf{x}) - H_j(\mathbf{x}), j = 1, .., K$. Denote $P(x) = \max_{j=1,...,K}\{F_j(\mathbf{x})\}$. Multiply two sides of the above inequality by λ_j^{k+1}, taking the sum of K inequalities when $j = 1, ..., K$ and using (5.16)-(5.19), we obtain

$$\sum_{j=1}^{K} \lambda_j^{k+1}\langle \nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \leq P(\mathbf{x}^k) - t^{k+1} - \frac{\rho_{min}}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2,$$

where $\rho_{min} = \min_{j=1,..,K}\{\rho_j\}$. The following inequality is obtained by adding the above inequality to (5.20)

$$\langle G_0(\mathbf{x}^{k+1}) - \nabla H_0(\mathbf{x}^k) + \sum_{j=1}^{K} \lambda_j^{k+1}(\nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k)), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle$$
$$\leq F_0(\mathbf{x}^k) + P(\mathbf{x}^k) - F_0(\mathbf{x}^{k+1}) - t^{k+1} - \frac{\rho_0 + \rho_{min}}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \quad (5.21)$$

In addition, it is deduced from the first inclusion of (5.16) that

$$\langle G_0(\mathbf{x}^{k+1}) - \nabla H_0(\mathbf{x}^k) + \sum_{j=0}^{K} \lambda_j^{k+1}(\nabla G_j(\mathbf{x}^{k+1}) - \nabla H_j(\mathbf{x}^k)), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \geq 0,$$

thus

$$\frac{\rho_0 + \rho_{min}}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \leq F_0(\mathbf{x}^k) + P(\mathbf{x}^k) - F_0(\mathbf{x}^{k+1}) - t^{k+1}.$$

Since $t^{k+1} \geq G_j(\mathbf{x}^{k+1}) - H_j(\mathbf{x}^k) - \langle \nabla H_j(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle \geq F_j(\mathbf{x}^{k+1}) \ \forall j \Rightarrow t^{k+1} \geq P(\mathbf{x}^{k+1})$, this combined with the above inequality leads to

$$\frac{\rho_0 + \rho_{min}}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \leq F_0(\mathbf{x}^k) + P(\mathbf{x}^k) - F_0(\mathbf{x}^{k+1}) - P(\mathbf{x}^{k+1}). \tag{5.22}$$

This inequality shows that the sequence $\{P(\mathbf{x}^k) + F_0(\mathbf{x}^k)\}$ is decreasing. Moreover $F_j(\mathbf{x})$ is continuous on the compact set $S \ \forall j = 0, .., K$ so it is bounded for every j. Therefore the sequence $\{P(\mathbf{x}^k) + F_0(\mathbf{x}^k)\}$ is also bounded and thus convergent. This combined with (5.22) leads to $\lim_{k\to\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\| = 0$.

Assume $(\mathbf{x}^*, t^*)$ is a limit point of the sequence $\{(\mathbf{x}^k, t^k)\}$. Therefore there exists a subsequence $\{(\mathbf{x}^{k_i+1}, t^{k_i+1})\}$ such that

$$\lim_{i\to\infty} (\mathbf{x}^{k_i+1}, t^{k_i+1}) = (\mathbf{x}^*, t^*).$$

This combined with $\lim_{i\to\infty} \|\mathbf{x}^{k_i} - \mathbf{x}^{k_i+1}\| = 0$ results in $\lim_{i\to\infty} \mathbf{x}^{k_i} = \mathbf{x}^*$. In addition, the sequence $\{\lambda_j^{k+1}\}$ is bounded for every $j = 1, ..., K$ thus without loss of generality we can assume that $\lim_{i\to\infty} \lambda_j^{k_i+1} = \lambda_j^*, j = 1, .., K$.

Replace k in (5.16) by k_i and taking limits as $i \to \infty$, we obtain

- $0 \in \nabla G_0(\mathbf{x}^*) - \nabla H_0(\mathbf{x}^*) + \sum_{j=1}^{K} \lambda_j^* (\nabla G_j(\mathbf{x}^*) - \nabla H_j(\mathbf{x}^*)) + N(S, \mathbf{x}^*)$,
- $1 - \sum_{j=1}^{K} \lambda_j^* = 0, \quad \mathbf{x}^* \in S$,
- $G_j(\mathbf{x}^*) - H_j(\mathbf{x}^*) \leq t^*, \quad \lambda_j^* \geq 0 \ \forall j = 1, ..., K$,
- $\lambda_j^* \left[G_j(\mathbf{x}^*) - H_j(\mathbf{x}^*) - t^*\right] = 0 \ \forall j = 1, ..., K$.

It shows that $(\mathbf{x}^*, t^*)$ is a critical point of the problem (5.13). □

5.3.2 DC Programming and DCA for solving (5.2)

The problem (5.2) is nonsmooth and nonconvex, hence it is difficult to deal with. Therefore, one tries to simplify it and then find a suboptimal solution. Because the jamming signal, which is emitted by the friendly jammers to confuse the eavesdroppers, might also affect the destination, thus in a natural way one wants to design beamforming coefficients in order to completely eliminate this noise at the destination. It means that apart from the power constraint, $\mathbf{w}$ is imposed to satisfy an additional constraint $\mathbf{h}_{RD}^\dagger \mathbf{w} = 0$. This constraint makes the first term of the objective function in (5.2) become a constant, so the problem (5.2) is simplified and easier to tackle. This case is referred to as a null-space CJ beamforming design. In this case, the problem (5.2) can be recast as a general DC program in which the objective function is linear and some constrants are DC. In what follows, we will present how to address the problem (5.2) in the null-space CJ beamforming design via a general DCA that is a new tool in DC programming.

5.3.2.1 The null-space CJ beamforming design

In the null-space CJ beamforming design, the vector $\mathbf{w}$ has to satisfy the equality $\mathbf{h}_{RD}^{\dagger}\mathbf{w} = 0$. The problem (5.2) thus is reduced to the simpler form as below.

$$\begin{aligned} \min_{\mathbf{w}} \max_{j=1,\ldots,K} \quad & \log_2 \left\{ \frac{\sigma^2 + \mathbf{w}^{\dagger}\mathbf{H}_{RE}(:,j)\mathbf{H}_{RE}^{\dagger}(:,j)\mathbf{w} + P_s|\mathbf{h}_{SE}(j)|^2}{\sigma^2 + \mathbf{w}^{\dagger}\mathbf{H}_{RE}(:,j)\mathbf{H}_{RE}^{\dagger}(:,j)\mathbf{w}} \right\} \\ \text{s.t} \quad & \mathbf{w}^{\dagger}\mathbf{w} \leq P_{tot}, \\ & \mathbf{h}_{RD}^{\dagger}\mathbf{w} = 0. \end{aligned} \tag{5.23}$$

Denote $\mathbf{R}_{RE_j} = \mathbf{H}_{RE}(:,j)\mathbf{H}_{RE}^{\dagger}(:,j)$, $C_j = \sigma^2 + P_s|\mathbf{h}_{SE}(j)|^2, \mathbf{T}_{2j} = \begin{bmatrix} \mathrm{Re}(\mathbf{R}_{RE_j}) & -\mathrm{Im}(\mathbf{R}_{RE_j}) \\ \mathrm{Im}(\mathbf{R}_{RE_j}) & \mathrm{Re}(\mathbf{R}_{RE_j}) \end{bmatrix}$, $j = 1,\ldots,K$, $\mathbf{M}_2 = \begin{bmatrix} \mathrm{Re}(\mathbf{h}_{RD}^{\dagger}) & -\mathrm{Im}(\mathbf{h}_{RD}^{\dagger}) \\ \mathrm{Im}(\mathbf{h}_{RD}^{\dagger}) & \mathrm{Re}(\mathbf{h}_{RD}^{\dagger}) \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} \mathrm{Re}(\mathbf{w}^T) & \mathrm{Im}(\mathbf{w}^T) \end{bmatrix}^T$. The problem (5.23) is recast as follows.

$$\begin{aligned} \min_{\mathbf{x},t} \quad & t \\ \text{s.t.} \quad & \mathbf{x}^T\mathbf{x} \leq P_{tot}, \\ & \mathbf{M}_2\mathbf{x} = 0, \\ & \ln\left(\frac{C_j + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x}}{\sigma^2 + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x}}\right) \leq t, \quad \forall j = 1,\ldots,K. \end{aligned}$$

The DC formulation of the above problem is given by

$$\begin{aligned} \min_{\mathbf{x},t} \quad & t \\ \text{s.t.} \quad & \mathbf{x}^T\mathbf{x} \leq P_{tot}, \\ & \mathbf{M}_2\mathbf{x} = 0. \\ & G_{2j}(\mathbf{x}) - H_{2j}(\mathbf{x}) \leq t \quad \forall j = 1,\ldots,K, \end{aligned} \tag{5.24}$$

where $G_{2j}(\mathbf{x}) = \frac{\rho_{2j}}{2}\|\mathbf{x}\|^2, H_{2j}(\mathbf{x}) = \frac{\rho_{2j}}{2}\|\mathbf{x}\|^2 - \ln(C_j + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x}) + \ln(\sigma^2 + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x})$, in which ρ_{2j} is chosen such that both functions G_{2j} and H_{2j} are convex. It follows from Proposition 5.1 that ρ_{2j} is the maximal eigenvalue of the matrix $\left(\frac{2}{C_j} + \frac{1}{2\sigma^2}\right)\mathbf{T}_{2j}$. Following the idea of DCA, at the kth iteration with the iterate $\mathbf{x}^k$, we compute $\nabla H_{2j}(\mathbf{x}^k) = \rho_{2j}\mathbf{x}^k - \frac{2\mathbf{T}_{2j}\mathbf{x}^k}{C_j + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x}} + \frac{2\mathbf{T}_{2j}\mathbf{x}^k}{\sigma^2 + \mathbf{x}^T\mathbf{T}_{2j}\mathbf{x}}$ and then solve the derived convex subproblem below.

$$\begin{aligned} \min_{\mathbf{x},t} \quad & t \\ \text{s.t} \quad & \mathbf{x}^T\mathbf{x} \leq P_{tot}, \\ & \mathbf{M}_2\mathbf{x} = 0, \\ & \frac{\rho_{2j}}{2}\|\mathbf{x}\|^2 - H_{2j}(\mathbf{x}^k) - \langle \nabla H_{2j}(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \rangle \leq t, \quad \forall j = 1,\ldots,K. \end{aligned} \tag{5.25}$$

The description of the general DCA scheme applied to (5.24), namely DCA-CJ-NS, is given by.

DCA-CJ-NS

Initialization: choose randomly $\mathbf{V}^0 = (\mathbf{x}^0, t^0) \in (\mathbb{R}^{2M}, \mathbb{R}^+)$ as an initial guess, set a tolerance ϵ for DCA-CJ-NS, $k \leftarrow 0$.
Repeat
• Calculate $\mathbf{V}^{k+1} = (\mathbf{x}^{k+1}, t^{k+1})$ by solving the subproblem (5.25).
• $k \leftarrow k + 1$.
Until $\left(\frac{\|\mathbf{V}^k - \mathbf{V}^{k-1}\|}{1+\|\mathbf{V}^{k-1}\|} < \epsilon \text{ or } \frac{|f_2(\mathbf{V}^k) - f_2(\mathbf{V}^{k-1})|}{1+|f_2(\mathbf{V}^{k-1})|} < \epsilon \right)$ where $f_2(\mathbf{V}^k) = t^k$.

The below theorem shows the convergence of DCA-CJ-NS. The arguments to prove this theorem are similar to those in the proof of Theorem 5.2.

Theorem 5.3.

(1) DCA-CJ-NS generates the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *such that the sequence of the corresponding objective function values* $\{f_2(\boldsymbol{V}^k)\}$ *is decreasing.*

(2) Every limit point of the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *generated by DCA-CJ-NS is a critical point to the problem (5.24).*

Now we turn to deal with the problem (5.2) in a general case in which the jamming signal might be not altogether eliminated at the destination. By directly dealing with the intractable original problem, it is expected that the better solution can be found. This problem can be reformulated as a general DC program in which the objective function and some constraints are DC. This is the most general DC program, thus more difficult to solve compared to the general DC program in the null-space CJ beamforming design.

5.3.2.2 The general CJ beamforming design

The problem (5.2) can be recast as follows

$$\begin{aligned} \max_{\mathbf{w}} \quad & \left[\log_2 \left(1 + \frac{P_s |h_{SD}|^2}{\mathbf{w}^\dagger \mathbf{R}_{RD} \mathbf{w} + \sigma^2} \right) - \max_{j=1,\ldots,K} \log_2 \left(1 + \frac{P_s |\mathbf{h}_{SE}(j)|^2}{\mathbf{w}^\dagger \mathbf{R}_{RE_j} \mathbf{w} + \sigma^2} \right) \right] \\ \text{s.t} \quad & \mathbf{w}^\dagger \mathbf{w} \leq P_{tot}. \end{aligned} \tag{5.26}$$

By introducing a slack variable t, the above problem is equivalent to the following one

$$\begin{aligned} \min_{\mathbf{w},t} \quad & \left[-\ln \left(1 + \frac{P_s |h_{SD}|^2}{\mathbf{w}^\dagger \mathbf{R}_{RD} \mathbf{w} + \sigma^2} \right) + t \right] \\ \text{s.t} \quad & \mathbf{w}^\dagger \mathbf{w} \leq P_{tot}. \\ & \ln \left(1 + \frac{P_s |\mathbf{h}_{SE}(j)|^2}{\mathbf{w}^\dagger \mathbf{R}_{RE_j} \mathbf{w} + \sigma^2} \right) \leq t \ \forall j = 1, ..., K. \end{aligned} \tag{5.27}$$

Denote $\mathbf{M}_{30} = \begin{bmatrix} \mathrm{Re}(\mathbf{R}_{RD}) & -\mathrm{Im}(\mathbf{R}_{RD}) \\ \mathrm{Im}(\mathbf{R}_{RD}) & \mathrm{Re}(\mathbf{R}_{RD}) \end{bmatrix}, \mathbf{M}_{3j} = \begin{bmatrix} \mathrm{Re}(\mathbf{R}_{RE_j}) & -\mathrm{Im}(\mathbf{R}_{RE_j}) \\ \mathrm{Im}(\mathbf{R}_{RE_j}) & \mathrm{Re}(\mathbf{R}_{RE_j}) \end{bmatrix}, C_j = \sigma^2 + P_s|\mathbf{h}_{SE}(j)|^2$ $j = 1, \ldots, K, \mathbf{x} = \begin{bmatrix} \mathrm{Re}(\mathbf{w}^T) & \mathrm{Im}(\mathbf{w}^T) \end{bmatrix}^T, C_0 = \sigma^2 + P_s|h_{SD}|^2$.
The problem (5.27) can be equivalently converted to a real variable form as follows

$$\begin{aligned} \min_{\mathbf{x},t} \quad & \left[-\ln\left(\frac{\mathbf{x}^T\mathbf{M}_{30}\mathbf{x} + C_0}{\mathbf{x}^T\mathbf{M}_{30}\mathbf{x} + \sigma^2} \right) + t \right] && (5.28) \\ \text{s.t} \quad & \mathbf{x}^T\mathbf{x} \leq P_{tot}, \\ & \ln \frac{\mathbf{x}^T\mathbf{M}_{3j}\mathbf{x} + C_j}{\mathbf{x}^T\mathbf{M}_{3j}\mathbf{x} + \sigma^2} \leq t \ \ \forall j = 1, ..., K. && (5.29) \end{aligned}$$

The problem (5.28) is still nonconvex because both the objective function and constraints (5.29) are nonconvex. In what follows, we will reformulate this problem as a general DC program with DC constraints and then design a general DCA scheme to solve it. First of all, the objective function, namely $f_3(\mathbf{x}, t)$, can be rewritten as follows

$$f_3(\mathbf{x}, t) = G_{30}(\mathbf{x}) + t - H_{30}(\mathbf{x}),$$

where $G_{30}(\mathbf{x}) = \frac{1}{2}\rho_{30}\|\mathbf{x}\|^2$ and $H_{30}(\mathbf{x}) = \frac{1}{2}\rho_{30}\|\mathbf{x}\|^2 + \ln(\mathbf{x}^T\mathbf{M}_{30}\mathbf{x} + C_0) - \ln(\mathbf{x}^\mathbf{T}\mathbf{M}_{30}\mathbf{x} + \sigma^2)$ for some $\rho_{30} > 0$. Similarly, for each $j = 1, ..., K$, the constraint (5.29) is recast as

$$G_{3j}(\mathbf{x}) - H_{3j}(\mathbf{x}) \leq t,$$

where $G_{3j}(\mathbf{x}) = \frac{1}{2}\rho_{3j}\|\mathbf{x}\|^2$ and $H_{3j}(\mathbf{x}) = \frac{1}{2}\rho_{3j}\|\mathbf{x}\|^2 - \ln(C_j + \mathbf{x}^T\mathbf{M}_{3j}\mathbf{x}) + \ln(\sigma^2 + \mathbf{x}^T\mathbf{M}_{3j}\mathbf{x})$.

It is apparent that as long as $\rho_{3j} > 0$ then $G_{3j}(\mathbf{x})$ is convex for all $j = 0, ..., K$. Nevertheless, H_{3j} is convex if ρ_{3j} is chosen in a similar way as in Proposition 5.1. It follows that, if ρ_{3j} is the maximal eigenvalue of the matrix $(\frac{1}{2C_0} + \frac{2}{\sigma^2})\mathbf{M}_{30}$ if $j = 0$ and $\left(\frac{2}{C_j} + \frac{1}{2\sigma^2}\right)\mathbf{M}_{3j}$ if $j = 1, \ldots, K$ then H_{3j} is convex. With these conditions, the general DC formulation of the problem (5.28) is described as

$$\begin{aligned} \min_{\mathbf{x},t} \quad & G_{30}(\mathbf{x}) + t - H_{30}(\mathbf{x}) && (5.30) \\ \text{s.t} \quad & \mathbf{x}^T\mathbf{x} \leq P_{tot}, \\ & G_{3j}(\mathbf{x}) - H_{3j}(\mathbf{x}) \leq t \ \ \forall j = 1, ..., K. && (5.31) \end{aligned}$$

Apply the idea of DCA, the second DC components H_{3j}, $j = 0, ..., K$ are linearized at each iteration. In more detail, assume that $\mathbf{x}^k$ is the kth iterate, for each $j = 0, ..., K$ the component $H_{3j}(\mathbf{x})$ is replaced by its linear approximation given by

$$H_{3j}(\mathbf{x}^k) + \langle \mathbf{z}_j^k, \mathbf{x}\text{-}\mathbf{x}^k \rangle,$$

where $\mathbf{z}_j^k \in \partial H_{3j}(\mathbf{x}^k)$. Because the functions H_{3j} is differentiable for every $j = 0, \ldots, K$, $\mathbf{z}_j^k$ can be computed by

$$\mathbf{z}_0^k = \nabla H_{30}(\mathbf{x}^k) = \left(\rho_{30}\mathbf{x}^k + \frac{2\mathbf{M}_{30}\mathbf{x}^k}{C_0 + (\mathbf{x}^k)^T\mathbf{M}_{30}\mathbf{x}^k} - \frac{2\mathbf{M}_{30}\mathbf{x}^k}{\sigma^2 + (\mathbf{x}^k)^T\mathbf{M}_{30}\mathbf{x}^k} \right)$$

and

$$\mathbf{z}_j^k = \nabla H_{3j}(\mathbf{x}) = \left(\rho_{3j}\mathbf{x}^k - \frac{2\mathbf{M}_{3j}\mathbf{x}^k}{C_j + (\mathbf{x}^k)^T\mathbf{M}_{3j}x^k} + \frac{2\mathbf{M}_{3j}\mathbf{x}^k}{\sigma^2 + (\mathbf{x}^k)^T\mathbf{M}_{3j}\mathbf{x}^k}\right), j = 1, ..., K.$$

The general DCA scheme for solving (5.30), namely DCA-CJ, is depicted as follows.

DCA-CJ

Initialization: Choose randomly $t^0 \in \mathbb{R}^+$ and denote $\mathbf{x}^0$ as the solution obtained by DCA-CJ-NS. Choose $\mathbf{V}^0 = (\mathbf{x}^0, t^0)$ as an initial guess, set a tolerance ϵ for DCA-CJ, $k \leftarrow 0$.

Repeat

• Calculate $\mathbf{V}^{k+1} = (\mathbf{x}^{k+1}, t^{k+1})$ by solving the following subproblem

$$\begin{aligned} \min_{\mathbf{x},t} \quad & \frac{1}{2}\rho_{30}\|\mathbf{x}\|^2 + t - \langle \mathbf{z}_0^k, \mathbf{x}\rangle \\ \mathbf{s.t} \quad & \mathbf{x}^T\mathbf{x} \le P_{tot}, \\ & \frac{1}{2}\rho_{3j}\|\mathbf{x}\|^2 - H_{3j}(\mathbf{x}^k) - \langle \mathbf{z}_j^k, \mathbf{x} - \mathbf{x}^k\rangle \le t \quad \forall j = 1, ..., K. \end{aligned} \tag{5.32}$$

• $k \leftarrow k + 1$.

Until $\left(\frac{\|\mathbf{V}^k - \mathbf{V}^{k-1}\|}{1+\|\mathbf{V}^{k-1}\|} < \epsilon \text{ or } \frac{|f_3(\mathbf{V}^k) - f_3(\mathbf{V}^{k-1})|}{1+|f_3(\mathbf{V}^{k-1})|} < \epsilon\right)$

Theorem 5.4. *(The convergence property of DCA-CJ)*

(1) DCA-CJ generates the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *such that the sequence of the corresponding objective function values* $\{f_3(\boldsymbol{V}^k)\}$ *is decreasing.*

(2) Every limit point of the sequence $\{\boldsymbol{V}^k = (\boldsymbol{x}^k, t^k)\}$ *generated by DCA-CJ is a critical point to the problem (5.30).*

This theorem is proved in a similar way as Theorem 5.2.

5.4 Numerical Results

5.4.1 AF Scenario

5.4.1.1 Comparative algorithms

We tested all the DCA based algorithms on some generated datasets and compared them with the semidefinite relaxation technique based algorithms, namely SDR for the general beamforming scheme and SDR-NS for the null-space relay beamforming

design ([120]).
SDR scheme for the general AF relay beamforming design
By denoting $\mathbf{W} = \mathbf{w}\mathbf{w}^\dagger$ and note that $\mathbf{w}^\dagger \sum \mathbf{w} = \mathrm{Tr}(\sum \mathbf{W})$, the problem (5.1) can be equivalently recast as follows

$$\begin{aligned} \min_{\mathbf{W},\tau} \quad & \frac{\mathrm{Tr}(\mathbf{GW}) + 1}{(\mathrm{Tr}((\mathbf{A}+\mathbf{G})\mathbf{W}) + 1)\tau} \\ \text{s.t} \quad & \mathrm{Tr}(\mathbf{e}\mathbf{e}_m^T\mathbf{CW}) \leq P_m \ \forall m = 1, ..., M. \\ & \mathrm{Tr}(\mathbf{CW}) \leq P_{tot}, \\ & \frac{\mathrm{Tr}(\mathbf{H}_j\mathbf{W}) + 1}{\mathrm{Tr}((\mathbf{B}_j + \mathbf{G}_j)\mathbf{W}) + 1} \geq \tau \ \forall j = 1, ..., K. \end{aligned} \tag{5.33}$$

This problem was treated as two-level optimization problem in which the inner level is a quasi-convex problem when τ is fixed and the outer level is a single variable optimization problem in τ when $\mathbf{W}$ is known. To deal with the inner problem, one reformulated it as a convex semidefinite program, which can be efficiently solved by available solvers such as CVX ([28], [27]). The outer problem was handled by using one-dimensional optimization techniques.
SDR-NS scheme for the null-space relay beamforming design
To avoid the complexity of the SRM beamforming scheme, it is sometimes supposed that there is no information leaked to the eavesdroppers. The problem (5.1) then can be reduced to

$$\begin{aligned} \max_{\mathbf{w},\tau} \quad & \frac{\mathbf{w}^\dagger\mathbf{Aw}}{\mathbf{w}^\dagger\mathbf{Gw} + 1} \\ \text{s.t} \quad & \mathbf{w}^\dagger\mathbf{Cw} \leq P_{tot}, \\ & \mathbf{e}_m^T\mathbf{Cww}^\dagger\mathbf{e}_m \leq P_m \ \forall m = 1, ..., M. \\ & \mathbf{h}_k^\dagger\mathbf{D(f)w} = 0 \ \forall k = 1, ..., K. \end{aligned} \tag{5.34}$$

Firstly the variable $\mathbf{w}$ was transformed to the variable $\mathbf{v}$ through the transformation $\mathbf{w} = \mathbf{Uv}$ where $\mathbf{U}$ is a matrix including an orthonormal basis of the null space of matrix comprised of the row $\mathbf{h}_k^\dagger\mathbf{D(f)}, k = 1, .., K$. The resulting problem with variable $\mathbf{v}$ then was relaxed to a convex problem by using semidefinite relaxation technique in combination with Charnes-Cooper transformation. The solution of the final problem was proved to be of rank one, thus it allows to find the optimal solution to (5.34).

5.4.1.2 Experimental setups

In our experiments, all the algorithms were implemented in the Matlab 2013b, and performed on a PC Intel Core i5-2500S CPU 2.70GHz of 4GB RAM. We stopped the DCA schemes with the tolerance $\epsilon = 10^{-4}$. The channel coefficients **f, g** and $\mathbf{h}_j, j = 1, 2, ..., K$ are drawn from a circularly-symmetric and zero mean complex normal distribution with covariance matrix $\mathbf{I}_M$. The noise variance is set to $\sigma^2 = 1$. The relays are constrained by both the total and individual power budgets. Particularly, the individual budget of the mth relay P_m is set to $0.5P_{tot}/M$ if m is odd and $P_m = 2P_{tot}/M$ otherwise. The reported results were taken average over 100 independent trials.

Table 5.1: The computing time (in seconds) of the algorithms in Experiment 1

P_{tot}	$P_s = 20$				$P_s = 10$			
	SDR	DCA-AF	SDR-NS	DCA-AF-NS	SDR	DCA-AF	SDR-NS	DCA-AF-NS
4	9.257	**2.086**	0.344	**0.100**	9.004	**2.971**	0.342	**0.117**
8	9.409	**3.016**	0.343	**0.124**	9.097	**4.078**	0.340	**0.151**
12	9.482	**3.644**	0.344	**0.145**	9.240	**5.055**	0.340	**0.181**
16	9.440	**4.133**	0.344	**0.163**	9.246	**5.664**	0.339	**0.202**
20	9.449	**4.342**	0.345	**0.177**	9.237	**6.129**	0.338	**0.227**
24	9.441	**4.812**	0.345	**0.188**	9.231	**6.623**	0.341	**0.252**

5.4.1.3 Experiment 1

In the first experiment, we compare the secrecy rates obtained by all the algorithms corresponding to various values of the total relay power budget P_{tot} in two different cases $P_s = 10$ and $P_s = 20$. The number of relays and eavesdropper are set to 10 and 5, respectively. The running time of all algorithms are reported in Table 5.1.

Comment on the numerical result.

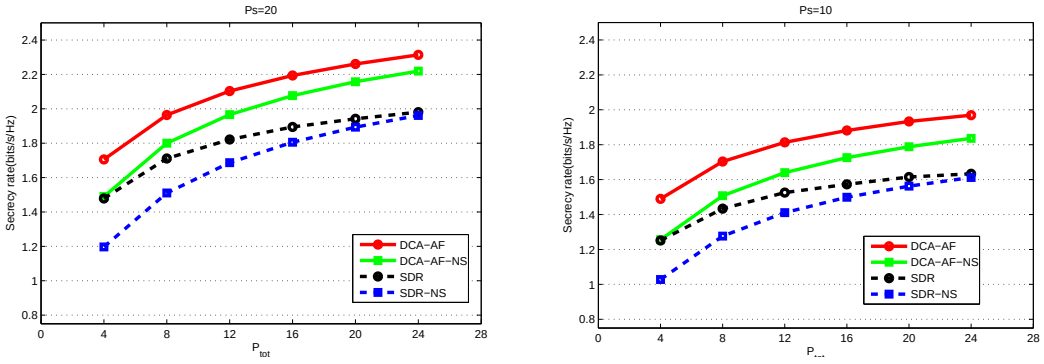


Figure 5.1: Secrecy Rate versus total relay power budget P_{tot}

Fig.5.1. illustrates the secrecy rate attained by DCA based algorithms and SDR based ones versus the total relay power budget with respect to two different values of P_s. In general, the secrecy rate gained by all the algorithms shows a rising trend when the total relay power budget goes up. Furthermore, the larger value of the source power P_s results in the higher secrecy rates. In both cases of P_s, DCA based schemes outperform SDR based ones. More specifically, DCA-AF returns the highest secrecy rate followed by DCA-AF-NS, SDR and SDR-NS, respectively.

Between two algorithms for the general beamforming design, DCA-AF gives the superior secrecy rate while it consumes less time compared with SDR. Table 5.1 shows that DCA-AF is at least twice as fast as SDR.

Between two DCA schemes, DCA-AF is better than DCA-AF-NS in terms of secrecy but DCA-AF-NS is more efficient in terms of runtime.

Table 5.2: The computing time (in seconds) of the algorithms in Experiment 2

K	SDR	DCA-AF	SDR-NS	DCA-AF-NS
11	17.458	**11.428**	0.537	**0.331**
13	17.841	**15.797**	0.522	**0.303**
15	**18.380**	22.038	0.486	**0.242**
17	**19.004**	32.992	0.458	**0.237**
19	**19.588**	55.886	0.288	**0.186**

Between two algorithms for the null-space AF relay beamforming design, DCA-AF-NS although is a local method, the experimental results show that the secrecy rate obtained by this algorithm is better than the global secrecy rate obtained by SDR-NS. This might be because the errors in computation make it difficult for solvers to get the actual global solution to the semidefinite program. In fact, although it is theoretically shown that the rank of the relaxed solution equals 1, which ensures that the original problem and the relaxed one are equivalent, the relaxed solution computed by the solver CVX actually does not satisfy this property. Concerning runtime, Table 5.1 shows that DCA-AF-NS runs faster than SDR-NS. The ratio between the computing time of SDR-NS and that of DCA-AF-NS is up to nearly 4 times.

5.4.1.4 Experiment 2

In this experiment, the secrecy rate gained by all the algorithms with respect to the different values of the number of eavesdroppers K are compared. The number of relays is set to $M = 20$, the source power and the total relay power budget are set to $P_s = 20$ and $P_{tot} = 20$, respectively.

Comment on the numerical results

Fig.5.2. depicts the variation of secrecy rate in the number of eavesdroppers. Overall,

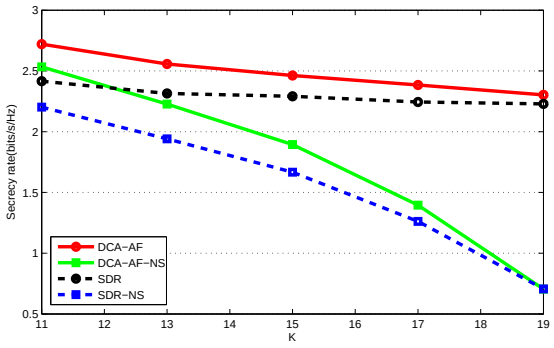


Figure 5.2: Secrecy Rate versus number of eavesdroppers K

it is observed from this figure that the secrecy rate suffers from a decrease when the number of eavesdroppers in the system augments. In addition, DCA-AF is still the best while SDR-NS is the worst in terms of secrecy rate.

Table 5.3: The computing time (in seconds) of the algorithms in Experiment 3

M	SDR	DCA-AF	SDR-NS	DCA-AF-NS
11	11.126	**8.949**	**0.174**	0.207
20	14.178	**8.971**	**0.403**	0.524
30	19.108	**4.781**	**0.545**	0.850
40	27.611	**4.041**	**0.880**	0.991
50	39.003	**2.907**	1.562	**1.321**

Between two algorithms for the general AF relay beamforming design, DCA-AF provides secrecy rates better than SDR does. The gap between secrecy rates respectively achieved by DCA-AF and SDR is considerable when the number of eavesdroppers K far smaller than that of relays M and gradually decreasing when they become close each other. In the computing aspect, DCA-AF is faster than SDR when $M - K$ is large but slower when $M - K$ is small.

Between two algorithms for the null-space AF relay beamforming design, when the number of eavesdroppers approaches that of relays, the secrecy rate obtained by both DCA-AF-NS and SDR-NS goes down steeply. DCA-AF-NS brings the better secrecy rate than SDR-NS does even though it has been theoretically proven in [120] that SDR-NS scheme obtained a global solution. This suggests that DCA-AF-NS might furnish a global optimal value of secrecy rate in the null-space beamforming design. It provides an example that DCA based algorithm is able to give a global solution though it is only a local approach. In the computing time aspect, DCA-AF-NS is approximately twice as fast as SDR-NS.

Between two DCA schemes, DCA-AF provides the better secrecy rates than DCA-AF-NS does, especially when the difference between the number of relays and eavesdroppers is small. The gap of secrecy rates achieved by these two algorithms is rising when the number of eavesdroppers is increasing and becomes large when it approaches the number of relays. However, DCA-AF-NS is much less expensive than DCA-AF.

In short, DCA-AF always gives the best secrecy rate in an acceptable time. DCA-AF-NS ensures the best trade-off between secrecy and runtime aspects when the number of eavesdroppers is much smaller than that of relays. SDR although has runtime less than DCA-AF in some cases, it obtains secrecy rate worse than DCA-AF. SDR-NS is inexpensive but ineffective in terms of secrecy.

5.4.1.5 Experiment 3

In this experiment, we illustrate how secrecy rates are affected by the number of relays M. The number of eavesdroppers is set to $K = 10$ and the total relay power budget is equal to $P_{tot} = 50$dB.

Comment on numerical results. Fig.5.3 indicates that the larger the number of relays is, the bigger the secrecy rates are attained by all the algorithms.

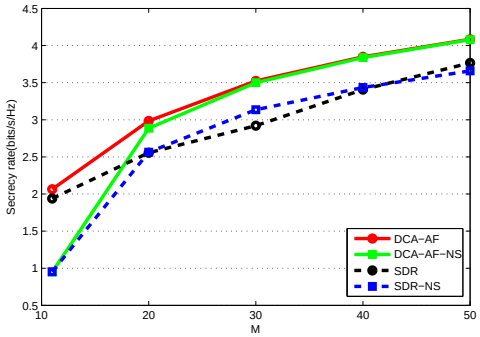


Figure 5.3: Secrecy Rate versus number of relays M

For the general AF relay beamforming design, DCA-AF is superior than SDR in the secrecy aspect and the gap between them is significant. On the contrary, DCA-AF is less expensive than SDR. The Table 5.3 shows that the ratio of computing times between them is up to 13 times.

For the null-space AF relay beamforming design, when the difference between the number of eavesdroppers and that of relays is small, both the algorithms DCA-AF-NS and SDR-NS are ineffective to ensure information confidentiality. Nevertheless, when this difference becomes larger DCA-AF-NS becomes the best while SDR-NS is still the worst. The runtime of these two algorithms are quite comparable.

Between two proposed DCA schemes, DCA-AF gives the better secrecy rate but spends more time than DCA-AF-NS does. When the number of relays is much bigger than that of eavesdroppers, the secrecy rate obtained by two DCA schemes are quite the same whereas DCA-AF-NS consumes much less time than DCA-AF. Therefore, it can be said that in that situation, DCA-AF-NS is the most efficient in terms of both secrecy and computing aspect.

5.4.2 CJ Scenario

5.4.2.1 The Comparative algorithm

One existing method, namely SubOpt, given in [18] provided a suboptimal solution to the secrecy rate maximization in the CJ scenario. The beamforming vector $\mathbf{w}$ was found to discard the jamming signal at the destination, i.e. it satisfies $\mathbf{h}_{RD}^{\dagger}\mathbf{w} = 0$. This leads to a simpler form of the problem (5.26) as below.

$$\begin{aligned} \max_{\mathbf{w}} \min_{j=1,\ldots,K} \quad & \frac{|\mathbf{w}^{\dagger}\mathbf{H}_{RE}(:,j)|^2 + \sigma^2}{|\mathbf{h}_{SE}(j)|^2} \\ \text{s.t.} \quad & \mathbf{w}^T\mathbf{w} \leq P_t, \\ & \mathbf{w}^{\dagger}\mathbf{h}_{RD} = 0, \end{aligned} \tag{5.35}$$

For each $j = 1, \ldots, K$, the problem

$$\begin{aligned} \max_{\mathbf{w}} \quad & \frac{|\mathbf{w}^\dagger \mathbf{H}_{RE}(:, j)|^2 + \sigma^2}{|\mathbf{h}_{SE}(j)|^2} \\ \text{s.t.} \quad & \mathbf{w}^T \mathbf{w} \leq P_t, \\ & \mathbf{w}^\dagger \mathbf{h}_{RD} = 0, \end{aligned} \tag{5.36}$$

can be explicitly solved and its closed-form solution was indicated in [18]. A suboptimal solution to (5.35) is the one that obtains the highest secrecy rate among K solutions attained from (5.36) when $j = 1, \ldots, K$.

5.4.2.2 Experimental setups

In this experiment, all the algorithms were implemented in the Matlab 2013b, and performed on a PC Intel Core i5-2500S CPU 2.70GHz of 4GB RAM. We stopped the DCA schemes with the tolerance $\epsilon = 10^{-4}$. The channel coefficients $\mathbf{h}_{RD}^*, \mathbf{h}_{RE}^*, \mathbf{H}_{RE}^*$ are drawn from a circularly-symmetric and zero mean complex normal distribution with covariance matrix $\mathbf{I}_M$, i.e. $\mathcal{CN}(0, \mathbf{I}_M)$ and h_{SD}^* is generated from the distribution $\mathcal{CN}(0, 1)$. The noise variance is set to $\sigma^2 = 1$. The number of relays is $M = 10$. The relays are constrained by the total power budget, which is chosen from the set $\{20, 40, 60, 80, 100\}$. The reported results were taken average over 100 independent trials.

5.4.2.3 Numerical results

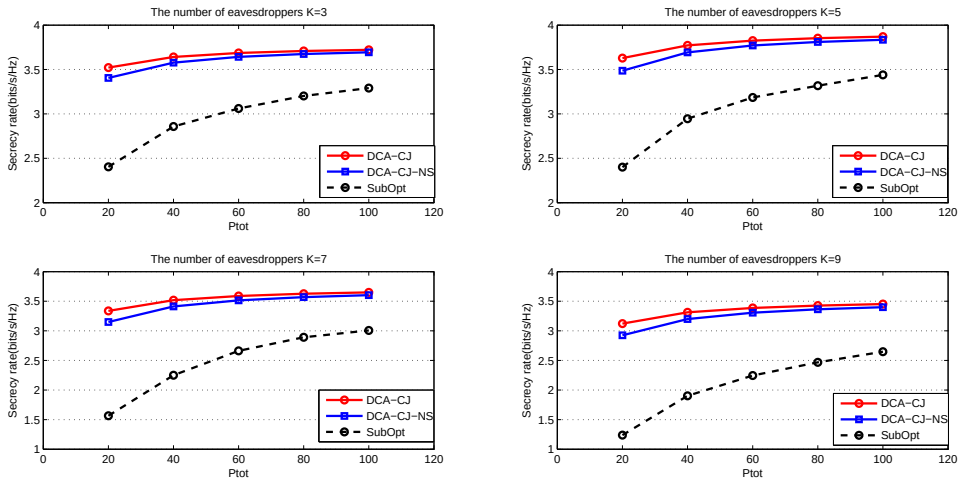


Figure 5.4: Secrecy Rate versus total relay power P_{tot}

The Figure 5.4 illustrates the secrecy rate obtained by three algorithms versus the total relay power in four cases of the number of eavesdroppers. Overall, it can be seen that the secrecy rate is increasing with increase in the total relay power and decreasing

Table 5.4: The computing time (in seconds) in the CJ scenario

P_{tot}	K=3			K=5		
	SubOpt	DCA-CJ	DCA-CJ-NS	SubOpt	DCA-CJ	DCA-CJ-NS
20	0.001	0.625	0.566	0.002	1.844	1.763
40	0.001	1.400	1.296	0.002	3.597	3.536
60	0.001	2.225	2.148	0.002	6.310	5.712
80	0.001	2.926	2.891	0.002	7.397	7.135
100	0.001	3.302	3.263	0.002	8.729	8.671
P_{tot}	K=7			K=9		
	SubOpt	DCA-CJ	DCA-CJ-NS	SubOpt	DCA-CJ	DCA-CJ-NS
20	0.003	3.332	3.118	0.005	4.849	4.261
40	0.003	6.173	6.100	0.005	10.245	8.241
60	0.003	9.725	8.894	0.005	15.323	12.541
80	0.003	12.408	12.329	0.005	17.081	16.991
100	0.003	15.964	15.887	0.005	22.235	22.142

with increase in the number of eavesdroppers. In all four cases of the number of eavesdroppers, DCA-CJ furnishes the best secrecy rates, followed by DCA-CJ-NS and SubOpt, respectively. The gaps of secrecy rate obtained by two DCA schemes and SubOpt are significant, especially when the total relay power is small. Meanwhile the gap of secrecy rate between DCA-CJ and DCA-CJ-NS is quite small. In terms of runtime, Table 5.4 shows that SubOpt runs fastest followed by DCA-CJ-NS and DCA-CJ, respectively. SubOpt consumes the least time because its computation is explicit, but it provides the worst secrecy rate. By contrast, DCA-CJ and DCA-CJ-NS though spend more time than SubOpt, they give much better secrecy rate than SubOpt. The efficiency of two proposed DCA schemes are quite the same in terms of both secrecy rate and runtime.

5.5 Conclusion

We have investigated into DC programming and DCA for solving the secrecy rate maximization problems in the cooperative beamforming relay networks. For both AF and CJ scenarios, two DCA based algorithms are designed to deal with the general and null-space beamforming design. Not only the standard DCA but also the general DCA, which is a generalization of the standard DCA, are proposed and their convergence are proven. The general DCA based approach enables to solve a wider class of nonconvex optimization problems, thus contributing to an expansion of applications of DC programming and DCA to more diverse fields of applied science. Compared to the existing methods, the proposed DCA schemes directly give the feasible solutions and achieve superior secrecy rates. Especially, for the null-space AF beamforming design, the experimental simulation reveals that DCA-AF-NS may provides a global optimal value of secrecy rate. The efficiency of the proposed DCAs suggests that the approach based on DC programming and DCA is worth considering when coping with

hard nonconvex optimization problems in physical layer security in particular as well as in communication systems in general, besides the common methods based on the semidefinite relaxation technique.

Chapter 6

Conclusion

In this dissertation, we study two important issues in communication systems: quality of service and physical layer security in which the majority of work is on the latter. The methodologies of the dissertation are DC (Difference of convex functions) programming and DCA (DC Algorithms), which are powerful tools in the nonconvex nondifferentiable optimization area. In addition to the standard DCAs, which have been completely studied and successfully applied in a lot of works in various areas of the applied science, the new tool based on a general DCA scheme is also developed in this dissertation. This tool is a generalization of the standard DCA based one and only studied and applied in some recent works. It permits to address the most general DC programs, thus being able to be applied to most optimization problems in practice.

With the aim of exploiting different effects of DC decompositions on the corresponding DCA scheme, we have proposed a novel DC decomposition and developed an efficient general DCA scheme to solve the max-min fairness optimization problem stemming from ensuring QoS for users. The proposed DCA scheme is a generalization of the standard one in which the convex approximation at each iteration of DCA is performed not only in the objective function but also in the DC constraints. The global convergence of the proposed general DCA scheme is rigorously shown.

The efficiency of DCA depends not only on the proposed DC decomposition but also on the way we treat the convex subproblems. This flexibility of DCA is verified when we address the power allocation problem derived when the cooperative jamming technique is deployed in a point-to-point network to maximize its secrecy rate. We propose a new DC decomposition for the objective function and develop two DCA schemes based on centralized and distributed methods for solving it. In the distributed DCA scheme, we design a highly efficient distributed dual based gradient projection algorithm to solve the convex subproblem by exploring and exploiting the special structure of this problem in a deep and efficient way. It turns out that our distributed DCA scheme requires computing iteratively the projection of points onto the intersection of a box and a half space which can be determined in a very inexpensive way. In comparison with the existing DCA, our proposed DCAs bring better secrecy rates and run much more rapidly.

DC programming and DCA are continued investigating to tackle the secrecy rate maximization problem in a relay network using the AF and DF relaying protocol or CJ technique. In the case when the relay network includes a single eavesdropper, our proposed DC decomposition leads to the convex subproblems whose solution can be found explicitly. Such a DC decomposition is highly recommended in DC programming and DCA because it brings good effects on the quality of DCA as well as the property of the sought soulution. In the case when multiple eavesdroppers are considered, the secrecy rate maximization problem is reformulated as a general DC program with the DC objective function and DC constraints. A general DCA scheme - a generalization of the standard one is designed to deal with that problem and its convergence is proved. The proposed algorithm is tested on the generated datasets and its results are compared with those of the existing methods based on a semidefinite relaxation technique. It implies that our proposed method outperforms previous approaches.

In summary, two issues of QoS and physical layer security are studied in this dissertation and DC programming and DCA are investigated to solve the derived optimization problems. For future research direction, several following issues should be continued developing from this research.

On the issue of QoS, there is no doubt that QoS will remain a fundamental requirement in the network design to meet higher and higher demands of customers. QoS can be regarded as an objective or a condition of the network design targets and assessed by other criteria such as error rate, bit rate, throughput besides SNR. The optimization problems related to QoS derived in future networks employing new techniques will be challenging and it requires to exploit different optimization tools for handling them. DC programming and DCA have been shown as the robust and efficient tools to cope with hard and large-scale programs in many applied sciences, thus we continue investigating these powerful tools to solve the models related to QoS.

On the issue of security, since physical layer security arises in multiusers systems of any kind, it is expected that new network scenarios using new techniques and corresponding security schemes will continue to be developed. DC programming and DCA remain promising tools to deal with such schemes. Furthermore, we intend to develop stochastic DCA schemes to address the secrecy models in which statistical channel knowledge is assumed rather than perfect channel state information that is a popular assumption in the literature but quite hard to obtain for practical communication systems. Stochastic DCA based approach also opens up a new perspective on tackling the secrecy models in wireless communication systems including a large number of users.

For a further goal, we will study DC programming and DCA for dealing with models arising from other requirements of communication systems such as internet congestion control, internet routing, etc, not only from QoS and physical layer security.

Appendix A

Appendix

A.1 The dual based gradient projection method ([8])

Let $X = \mathbb{R}^n$, A is a $m \times n$ matrix. Let us consider the problem

$$\begin{aligned} p^* = \min \quad & f_0(x) && (P) \\ \text{s.t} \quad & x \in C, \\ & Ax - b \leq 0, \end{aligned}$$

where f_0 is convex and ∇f_0 satisfies a Lipschitz condition, C is a nonempty compact convex set of X.
The Lagrangian of this problem is

$$L(x, \lambda) = f_0(x) + \lambda^T (Ax - b).$$

The dual problem of (P) is that

$$d^* = \max_{\lambda \geq 0} \{g(\lambda) = \min_{x \in C} L(x, \lambda)\}. \qquad (D)$$

Because the constraints are linear, Slater's condition holds if the feasible set of (P) is nonempty, thus the strong duality is ensured, i.e. $p^* = d^*$.
Denote $\widehat{x}(\lambda) \in \arg\min_{x \in \mathrm{C}} L(x, \lambda)$.

The dual based gradient projection algorithm for solving the problem (P)

- **Initialization**. Choose an initial point $\lambda^0 \geq 0$.
- **Repeat**
  Step 1 Compute $\widehat{x}^{k-1}(\lambda^{k-1}) \in \arg\min_{x \in \mathrm{C}} L(x, \lambda^{k-1})$.
  Step 2 Compute $\lambda^k = P_C(\lambda^{k-1} + \alpha^{k-1}(A\widehat{x}^{k-1} - b))$.
  Step 3 $k \leftarrow k + 1$.

- **Until** Stopping condition is satisfied.

The convergence of this method is shown in the following theorem.

Theorem A.1. *If the function f_0 is strongly convex with parameter σ then the sequence $\{\lambda^k\}$ generated by the above algorithm converges to the unique solution λ^* of (D) and the sequence $\{\widehat{x}(\lambda^k)\}$ converges to the unique solution x^* of (P).*

The proof of this theorem is based on [7].

Proof. First of all, we show that the function $\widehat{x}(\lambda)$ is Lipschitz on λ. Since the function $f_0(x)$ is strongly convex, so the Lagrangian $L(x,\lambda)$ is also convex in x, this combined with the property of the minimum lead to

$$(\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T \nabla_x L(\widehat{x}(\lambda^1), \lambda^1) \geq 0$$

and

$$(\widehat{x}(\lambda^1) - \widehat{x}(\lambda^2))^T \nabla_x L(\widehat{x}(\lambda^2), \lambda^2) \geq 0.$$

Summing the two inequalities above, we obtain:

$$\begin{aligned}
& (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^1), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^2), \lambda^2)) \geq 0 \\
\Leftrightarrow\ & (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^1), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^2), \lambda^1) \\
& + \nabla_x L(\widehat{x}(\lambda^2), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^2), \lambda^2)) \geq 0 \\
\Leftrightarrow\ & (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^2), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^2), \lambda^2)) \\
& \geq (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^2), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^1), \lambda^1)).
\end{aligned}$$

Since the function $L(x,\lambda)$ is strongly convex in x with the parameter σ, we have

$$(\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^2), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^1), \lambda^1)) \geq \frac{\sigma}{2} \|\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1)\|^2,$$

where $\|.\|$ is denoted as the Euclidean norm.
Moreover,

$$\begin{aligned}
& (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (\nabla_x L(\widehat{x}(\lambda^2), \lambda^1) - \nabla_x L(\widehat{x}(\lambda^2), \lambda^2)) \\
= \ & (\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1))^T (A^T(\lambda^1 - \lambda^2)) \\
\leq \ & \|\widehat{x}(\lambda^2) - \widehat{x}(\lambda^1)\| \|A^T\| \|\lambda^1 - \lambda^2\|.
\end{aligned}$$

The deductions above implies that $\widehat{x}(\lambda)$ is Lipschitz.
Because of the strong convexity of the Lagrangian $L(x,\lambda)$ in x, the inner minimization in (D) has the unique solution. Therefore, according to Danskin's theorem, $g(\lambda)$ is differentiable and $\nabla g(\lambda) = h(\widehat{x}(\lambda))$.
The Lipschitz property of $\nabla g(\lambda)$ is drawn from that of $h(\lambda)$ and $\widehat{x}(\lambda)$.
We also observe that

$$\begin{aligned}
-g(\lambda) &= \max_{x \in C} (-f_0(x) - \lambda^T (Ax - b)) \\
&= \max_{x \in C} (x^T(-A^T\lambda) - f_0(x) + \lambda^T b)) = f_0^*(-A^T\lambda) + \lambda^T b.
\end{aligned}$$

Since $\nabla f_0(\lambda)$ is Lipschitz, applying Proposition 12.60 in [84], we conclude that the function $f_0^*(\lambda)$ is strongly convex, so is $-g(\lambda)$. As a result, the convergence of the sequence $\{\lambda^k\}$ is drawn from the convergence of the gradient projection method in [58].
Besides, from the strong convexity of $L(x,\lambda)$ in x and $\widehat{x}(\lambda^k) = \arg\min_{x\in \mathrm{C}} L(x,\lambda^k)$ we have

$$L(x,\lambda^k) - L(\widehat{x}(\lambda^k),\lambda^k) \geq \frac{\sigma}{2}\|\widehat{x}(\lambda^k) - x\|^2 \quad \forall x \in C.$$

In particular, if x^* is the optimal solution of the problem (P) then $x^* \in C$, thus

$$\begin{aligned}
&L(x^*,\lambda^k) - L(\widehat{x}(\lambda^k),\lambda^k) \geq \frac{\sigma}{2}\|\widehat{x}(\lambda^k) - x^*\|^2 \\
&\Leftrightarrow f_0(x^*) + (\lambda^k)^T h(x^*) - f_0(\widehat{x}(\lambda^k)) - (\lambda^k)^T h(\widehat{x}(\lambda^k)) \geq \frac{\sigma}{2}\|\widehat{x}(\lambda^k) - x^*\|^2 \\
&\Leftrightarrow g(\lambda^*) - g(\lambda^k) + (\lambda^k)^T h(x^*) \geq \frac{\sigma}{2}\|\widehat{x}(\lambda^k) - x^*\|^2.
\end{aligned}$$

Note that $\lambda^k \geq 0$ and $h(x^*) \leq 0$, thus

$$g(\lambda^*) - g(\lambda^k) \geq \frac{\sigma}{2}\|\widehat{x}(\lambda^k) - x^*\|^2.$$

When $k \to \infty$ then $g(\lambda^k) \to g(\lambda^*)$ (due to the fact that $g(\lambda)$ is differentiable and $\lambda^k \to \lambda^*$), this combined with the above inequality implies that $\|\widehat{x}(\lambda^k) - x^*\|^2 \to 0$. □

A.2 Proposition 12.60 in [84]

Proposition A.1. *For a proper, lower semicontinuous, convex function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and a value $\sigma > 0$, the following properties are equivalent:*
(a) f^* is strongly convex with constant σ;
(b) f is differentiable and ∇f is Lipschitz continuous with constant $1/\sigma$.

A.3 Algorithm for projecting a vector on the intersection of a Hyperplane and a box in $\mathbb{R}^n$ ([69])

In this section, we present the algorithm **BoxProjection** in [69] for projecting the vector $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)^T$ on the set X which is defined by

$$X = \{x \in \mathbb{R}^n : a^T x = b, 0 \leq x_j \leq d_j, j = 1, 2, \ldots, n\}.$$

Let us denote $x^* = (x_1^*, x_2^*, \ldots, x_n^*) = \mathrm{Proj}_X(\bar{x})$ and

$$\begin{aligned}
J &= \{j | a_j \neq 0, j = 1, \ldots, n\}, \\
\tilde{J} &= \{1, 2, \ldots, n\}, \\
\hat{J} &= \{n+1, n+2, \ldots, 2n\}, \\
J^+ &= \{i \in \tilde{J} : a_i > 0\} \cup \{i \in \hat{J} : a_{i-n} > 0\}, \\
J^- &= \{i \in \tilde{J} : a_i < 0\} \cup \{i \in \hat{J} : a_{i-n} < 0\} \\
J_0 &= \{i \in \tilde{J} : a_i = 0\} \cup \{i \in \hat{J} : a_{i-n} = 0\}, \\
U^+(\bar{\alpha}) &= \{j \in \hat{J} \cap J^+ | \bar{\alpha} \leq \alpha_j\}, \\
U^-(\bar{\alpha}) &= \{j \in \tilde{J} \cap J^- | \bar{\alpha} \leq \alpha_j\}, \\
M^+(\bar{\alpha}) &= \{j \in \tilde{J} \cap J^+ | \alpha_{j+n} < \bar{\alpha} \leq \alpha_j\}, \\
M^-(\bar{\alpha}) &= \{j \in \tilde{J} \cap J^- | \alpha_j < \bar{\alpha} \leq \alpha_{j+n}\}, \\
L^+(\bar{\alpha}) &= \{j \in \tilde{J} \cap J^+ | \bar{\alpha} \geq \alpha_j\}, \\
L^-(\bar{\alpha}) &= \{j \in \hat{J} \cap J^- | \bar{\alpha} \geq \alpha_j\}, \\
L_e(\bar{\alpha}) &= \{j \in J^+ \cup J^- | \alpha_j < \bar{\alpha}, \} \\
E_q(\bar{\alpha}) &= \{j \in J^+ \cup J^- | \alpha_j = \bar{\alpha}, \} \\
G_r(\bar{\alpha}) &= \{j \in J^+ \cup J^- | \alpha_j > \bar{\alpha}\}.
\end{aligned}$$

Algorithm BoxProjection.
Step 1: Initialization.

Step 1.1: Build the sets $J, \tilde{J}, \hat{J}, J^+, J^-, J_0$.

Step 1.2: Compute $x_j^* = \max\{0, \min\{\bar{x}_j, d_j\}\}$ for all $j \in J_0$.

Step 1.3: Compute $\alpha_j = \frac{\bar{x}_j}{a_j}$ for all $j \in (J^+ \cup J^-) \cap \tilde{J}$.

Step 1.4: Compute $\alpha_{n+j} = \frac{\bar{x}_j - d_j}{a_j}$ for all $j \in (J^+ \cup J^-) \cap \tilde{J}$.

Step 1.5. Set $\sigma = 0, C^+ = \emptyset, C^- = \emptyset$.

Step 1.6: Compute S, where S is a list of α_j for all $j \in J$.

Step 1.7. set $p^+ = 0, p^- = 0, q^+ = 0, q^- = 0, r^+ = 0, r^- = 0$.
Step 2. If $|S| > 2$, then compute α_m, defined to be the median of the list S; otherwise, set $\alpha_t = \alpha_m, \alpha_m = \alpha_j$ such that $j \in J \setminus \{m\}$; set $\sigma_t = \sigma$.
Step 3: Compute $J = J \setminus \{m\}$.
Step 4: Having defined the value α_m, build the sets $L_e, E_q, G_r, U^+, U^-, L^+, L^-, M^+, M^-$ as described above.

Step 5: Compute

$$\begin{aligned}
p_1^+(\alpha_m, C^+) &= \sum_{j\in M^+(\alpha_m)\setminus C^+} a_j\bar{x}_j - \sum_{j\in U^+(\alpha_m)\cap C^+} a_j\bar{x}_j,\\
q_1^+(\alpha_m, C^+) &= \sum_{j\in M^+(\alpha_m)\setminus C^+} a_j^2 - \sum_{j\in U^+(\alpha_m)\cap C^+} a_j^2,\\
r_1^+(\alpha_m) &= \sum_{j\in U^+(\alpha_m)} a_{j-n}d_{j-n},\\
p_1^-(\alpha_m, C^-) &= \sum_{j\in M^-(\alpha_m)\setminus C^-} a_j\bar{x}_j - \sum_{j\in L^-(\alpha_m)\cap C^-} a_j\bar{x}_j,\\
q_1^-(\alpha_m, C^-) &= \sum_{j\in M^-(\alpha_m)\setminus C^-} a_j^2 - \sum_{j\in L^-(\alpha_m)\cap C^-} a_j^2,\\
r_1^-(\alpha_m) &= \sum_{j\in L^-(\alpha_m)} a_{j-n}d_{j-n},
\end{aligned}$$

Step 6: If $m \in \hat{J}$, then set $r_m = a_{m-n}d_{m-n}$; otherwise set $r_m = 0$.
Step 7: Compute

$$\begin{aligned}
\sigma = \big(p^+ + p_1^+(\alpha_m, C^+) + p_1^-(\alpha_m, C^-) + r^+ + r_1^+(\alpha_m) + r^- + r_1^-(\alpha_m) + r_m\big)\\
-\alpha_m(q^+ + q_1^+(\alpha_m, C^+) + q^- + q_1^-(\alpha_m, C^-)).
\end{aligned}$$

Step 8: if $\sigma > b$, then set

$$\begin{aligned}
J &= G_r(\alpha_m)\cup\{m\},\\
C^- &= C^- + M^-(\alpha_m), p^- = p^- + p_1^-, q^- = q^- + q_1^-, r^- = r^- + r_1^-;
\end{aligned}$$

compute S for all $j \in J$ and go to Step 2.
Step 9: if $\sigma < b$, then set

$$\begin{aligned}
J &= L_e(\alpha_m)\cup\{m\},\\
C^+ &= C^+ + M^+(\alpha_m), p^+ = p^+ + p_1^+, q^+ = q^+ + q_1^+, r^+ = r^+ + r_1^+;
\end{aligned}$$

compute S for all $j \in J$ and go to Step 2.
Step 10: if $\sigma = b$, then set $\alpha^* = \alpha_m$; otherwise, compute

$$D = \frac{\sigma - \sigma_t}{\alpha_m - \alpha_t}, \alpha^* = \frac{b - \sigma + D\alpha_m}{D}.$$

Step 11: if $\alpha^* \notin [\min\{\alpha_m, \alpha_t\}, \max\{\alpha_m, \alpha_t\}]$, then the problem has no solution. Stop.
Step 12: Compute $x_j^* = \max\{0, \min\{\bar{x}_j - \alpha^* a_j, d_j\}\}$ for all $j \in \tilde{J} \setminus J_0$.

Bibliography

[cpl] IBM ILOG CPLEX Optimizer. `http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/`.

[2] Ahn, S., Jung, S., Lee, W., Sung, T. K., Park, J. G., Lee, K. E., and Kang, J. (2016). Enhancing physical-layer security in miso wiretap channel with pilot-assisted channel estimation: Beamforming design for pilot jamming. In *2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–5.

[3] Al-jamali, M., Al-nahari, A., and AlKhawlani, M. M. (2015). Relay selection scheme for Improving the physical layer security in cognitive radio networks. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, pages 495–498.

[4] Al-Shatri, A. and Weber, T. (2012). Achieving the maximum sum rate using DC programming in cellular networks. *IEEE Trans. Signal Process*, 60(3):1331–1341.

[5] Alvarado, A., Scutari, G., and Pang, J.-S. (2014). A new decomposition method for multiuser DC-programming and its application. *IEEE Transactions on Signal Processing*, 62(11):2984–2998.

[6] Banawan, K. and Ulukus, S. (2014). Gaussian MIMO wiretap channel under receiver side power constraints. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 183–190.

[7] Beck, A., Nedic; A., Ozdaglar, A., and Teboulle, M. (2014). An o(1/k) gradient method for network resource allocation problems. *IEEE Transactions on Control of Network Systems*, 1(1):64–73.

[8] Bertsekas, D. P., Nedic, A., and Ozdaglar, E. (2003). *convex analysis and optimization*. Athena Sicientific Belmont.

[9] Bornhorst, N., Pesavento, M., and Gershman, A. B. (2012). Distributed beamforming for multi-group multicasting relay networks. *IEEE Transactions on Signal Processing*, 60(1):221–232.

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

[11] Chen, H., Greshman, B., Shahbazpanahi, S., and Gazor, S. (2010). Filter-and-forward distributed beamforming in relay networks with frequency selective fading. *IEEE Trans. Singal Process*, 58:1251–1262.

[12] Cheng, Y. and Pesavento, M. (2012). Joint Optimization of Source Power Allocation and Distributed Relay Beamforming in Multiuser Peer-to-Peer Relay Networks. *IEEE Trans. on Signal Process*, 60(6):2962–2973.

[13] Csiszár and Korner, J. (1978). Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348.

[14] Czyzyk, J., Mesnier, M. P., and Moré, J. J. (1998). The NEOS Server. *IEEE Journal on Computational Science and Engineering*, 5(3):68–75.

[15] Dartmann, G., Zandi, E., and Ascheid, G. (2013). Equivalent quasi-convex form of the multicast max-min beamforming problem. *IEEE Trans. on Vehicular Tech.*, 62(9):4643–4648.

[16] Deng, H., Wang, H. M., Guo, W., and Wang, W. (2015). Secrecy Transmission With a Helper: To Relay or to Jam. *IEEE Transactions on Information Forensics and Security*, 10(2):293–307.

[17] Dong, L., Han, Z., Petropulu, A., and Poor, H. (2009). Amplify-and-forward based cooperation for secure wireless communications. In *2009. ICASSP 2009. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2613–2616.

[18] Dong, L., Han, Z., Petropulu, A., and Poor, H. (2010). Improving wireless physical layer security via cooperating relays. *IEEE Trans. Signal Process*, 58(3):1875–1888.

[19] Ekrem, E. and Ulukus, S. (2011). The Secrecy Capacity Region of the Gaussian MIMO Multi-Receiver Wiretap Channel. *IEEE Transactions on Information Theory*, 57(4):2083–2114.

[20] Fang, B., Qian, Z., Zhong, W., and Shao, W. (2015). An-aided secrecy precoding for swipt in cognitive mimo broadcast channels. *IEEE Communications Letters*, 19(9):1632–1635.

[21] Fazeli-Dehkordy, S., Shahbazpanahi, S., and Gazor, S. (2009). Multiple peer-to-peer communications using a network of relays. *IEEE Transactions on Signal Processing*, 57(8):3053–3062.

[22] Gershman, A. B., Sidiropoulos, N. D., Shahbazpanahi, S., Bengtsson, M., and Ottersten, B. (2010). Convex optimization-based beamforming. *IEEE Signal Processing Magazine*, 27(3):62–75.

[23] Goeckel, D., Vasudevan, S., Towsley, D., Adams, S., Ding, Z., and Leung, K. (2011). Artificial Noise Generation from Cooperative Relays for Everlasting Secrecy in Two-Hop Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 29(10):2067–2076.

[24] Goel, S. and Negi, R. (2008). Guaranteeing Secrecy using Artificial Noise. *IEEE Transactions on Wireless Communications*, 7(6):2180–2189.

[25] Gomez-Cuba, F., Asorey-Cacheda, R., and Gonzalez-Castano, F. (2012). A Survey on Cooperative Diversity for Wireless Networks. *IEEE Communications Surveys Tutorials*, 14(3):822–835.

[26] Gopala, P. K., Lai, L., and Gamal, H. E. (2008). On the Secrecy Capacity of Fading Channels. *IEEE Transactions on Information Theory*, 54(10):4687–4698.

[27] Grant, M. and Boyd, S. (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, vol 371*, pages 95–110. Springer-Verlag Limited. `http://stanford.edu/~boyd/graph_dcp.html`.

[28] Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`.

[29] Havary, V., Shahbazpanahi, S., Grami, A., and Luo, Z. Q. (2008). Distributed beamforming for relay networks based on second-order statistics of the the channel state information. *IEEE Trans. Signal Process.*, 56(9):4306–4316.

[30] He, H., Ren, P., Du, Q., Sun, L., and Wang, Y. (2016). Secure and energy efficient transmission in multiuser uplink wireless networks. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5.

[31] He, X. and Yener, A. (2010). Cooperation with an Untrusted Relay: A Secrecy Perspective. *IEEE Trans. Inf. Theor.*, 56(8):3807–3827.

[32] Hu, X., Mu, P., Wang, B., Li, Z., Wang, H. M., and Ju, Y. (2016). Secrecy rate maximization for simo wiretap channel with uncoordinated cooperative jamming under secrecy outage probability constraint. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6.

[33] Jean-Baptiste Hiriart-Urruty, C. L. (1993). *Convex Analysis and Minimization Algorithms I: Fundamentals*. Grundlehren der mathematischen Wissenschaften 305. Springer-Verlag Berlin Heidelberg, 1 edition.

[34] Jeong, C., Kim, I.-M., and Kim, D. I. (2012). Joint Secure Beamforming Design at the Source and the Relay for an Amplify-and-Forward MIMO Untrusted Relay System. *IEEE Transactions on Signal Processing*, 60(1):310–325.

[35] Jorswieck, E., Wolf, A., and Gerbracht, S. (2010). *Secrecy on the Physical Layer in Wireless Networks*, chapter 20, pages 413–435. INTECH.

[36] Kha, H. H., Tuan, H. D., and Nguyen, H. H. (2012). Fast global optimal power allocation in wireless network by local DC programming. *IEEE Trans. on Wireless Communications*, 11(2):510–512.

[37] Khisti, A., Tchamkerten, A., and Wornell, G. W. (2008). Secure Broadcasting Over Fading Channels. *IEEE Transactions on Information Theory*, 54(6):2453–2469.

[38] Khisti, A., Wornell, G., Wiesel, A., and Eldar, Y. (2007). On the Gaussian MIMO Wiretap Channel. In *2007 IEEE International Symposium on Information Theory*, pages 2471–2475.

[39] Khisti, A. and Wornell, G. W. (2010a). Secure Transmission With Multiple Antennas I: The MISOME Wiretap Channel. *IEEE Transactions on Information Theory*, 56(7):3088–3104.

[40] Khisti, A. and Wornell, G. W. (2010b). Secure Transmission With Multiple Antennas-Part II: The MIMOME Wiretap Channel. *IEEE Transactions on Information Theory*, 56(11):5515–5532.

[41] Kqripidis, E., Sidiropoulos, N. D., and Luo, Z. Q. (2008). Quality of service and max-min fair transmit beamforming to multiple co-channel multicast group. *IEEE Trans. Signal Processing*, 56:1268–1279.

[42] Lai, L. and Gamal, H. E. (2008). The Relay-Eavesdropper Channel: Cooperation for Secrecy. *IEEE Transactions on Information Theory*, 54(9):4005–4019.

[43] Law, K. L., Wen, X., and Pesavento, M. (2013). General-rank transit beamforming for multi-group multicasting networks using OSTBC. In *Proc. IEEE SPAWC'13*, pages 475–479. IEEE.

[Le Thi] Le Thi, H. A. DC Programming and DCA. *http://www.lita.univ-lorraine.fr/~lethi/*.

[45] Le Thi, H. A. (1994). *Analyse numérique des algorithmes de l'Optimisation DC. Approaches locales et globales. Codes et simulations numérique en grande dimension. Applications.* PhD thesis, Université de Rouen.

[46] Le Thi, H. A. (1997). *Contribution à l'optimisation non convexe et l'optimisation globale: Théorie, Algorithmes et Applications.* PhD thesis, Habilitation à Diriger des Recherches, Université de Rouen.

[47] Le Thi, H. A., Huynh, V. N., and Dinh, T. P. (2014a). DC Programming and DCA for General DC Programs. In *Advanced Computational Methods for Knowledge Engineering, ICCSAMA 2014, Advances in Intelligent Systems and Computing, vol 282*, pages 15–35. Springer.

[48] Le Thi, H. A., Nguyen, M. C., and Pham Dinh, T. (2014b). A DC Programming Approach for Finding Communities in Networks. *Neural Computation*, 26(12):2827–2854.

[49] Le Thi, H. A., Nguyen, M. C., and Pham Dinh, T. (2014c). Self-Organizing Maps by Difference of Convex functions optimization. *Data Mining and Knowledge Discovery*, 28:1336–1365.

[50] Le Thi, H. A., Nguyen, Q. T., Phan, K. T., and Pham Dinh, T. (2013). DC Programming and DCA Based Cross-Layer Optimization in Multi-hop TDMA Networks. In *Intelligent Information and Database Systems, ACIIDS 2013, Lecture Notes in Computer Science, vol 7803*, pages 398–408. Springer.

[51] Le Thi, H. A. and Pham Dinh, T. (2005). The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems. *Annals of Operations Research*, 133:23–46.

[52] Le Thi, H. A. and Pham Dinh, T. (2013). Network utility maximisation: A DC programming approach for Sigmoidal utility function. In *2013 International Conference on Advanced Technologies for Communications (ATC 2013)*, pages 50–54.

[53] Le Thi, H. A. and Pham Dinh, T. (2014). DC programming in communication systems: challenging problems and methods. *Vietnam Journal of Computer Science*, 1(1):15–28.

[54] Le Thi, H. A., Tran, T. T., Pham Dinh, T., and Gély, A. (2016). DC Programming and DCA for Transmit Beamforming and Power Allocation in Multicasting Relay Network. In *Advanced Computational Methods for Knowledge Engineering, ICCSAMA 2016, Advances in Intelligent Systems and Computing, vol 453*, pages 29–41. Springer.

[55] Le Thi, H. A., Vo, X. T., Le, H. M., and Pham Dinh, T. (2015). DC approximation approaches for sparse optimization. *European Journal of Operational Research*, 244(1):26–46.

[56] Le Thi, H. A., Vo, X. T., and Pham Dinh, T. (2014d). Feature selection for linear SVMs under uncertain data: Robust optimization based on difference of convex functions algorithms. *Neural Networks*, 59:36–50.

[57] Leung-Yan-Cheong, S. and Hellman, M. (1978). The Gaussian wire-tap channel. *IEEE Transactions on Information Theory*, 24(4):451–456.

[58] Levitin, E. S. and Polyak, B. T. (1966). Constrained minimization methods. *Computers and Mathematics with applications*, 6(5):1–50.

[59] Li, J., Petropulu, A., and Weber, S. (2011). On cooperative relaying schemes for wireless physical layer security. *IEEE Trans. Signal Process*, 59(10):4985–4997.

[60] Li, L., Chen, Z., and Fang, J. (2014). On Secrecy Capacity of Gaussian Wiretap Channel Aided by A Cooperative Jammer. *IEEE Signal Processing Letters*, 21(11):1356–1360.

[61] Li, L., Li, L., Chen, Z., and Fang, J. (2015). Optimal Transmit Design at Relay Nodes for Secure AF Relay Networks. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–6.

[62] Li, Q. and Han, D. (2016). Sum secrecy rate maximization for full-duplex two-way relay networks. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3641–3645.

[63] Li, Z., Trappe, W., and Yates, R. (2007). Secrete Communication via Multi-antenna Transmission. In *41st Annual Conference on Information Sciences and Systems (CISS)*, pages 905–910.

[64] Liang, Y., Poor, H., and Shamai, S. (2008). Secure Communication Over Fading Channels. *IEEE Transactions on Information Theory*, 54(6):2470–2492.

[65] Liu, J., Hou, Y. T., and Sherali, H. D. (2009). Optimal Power Allocation for Achieving PerfectSecrecy Capacity in MIMO Wire-Tap Channels. In *43rd Annual Conference on Information Sciences and Systems (CISS)*, pages 606–611.

[66] Liu, T. and Shamai, S. (2009). A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel. *IEEE Transactions on Information Theory*, 55(6):2547–2553.

[67] Liu, W., Sarkar, M. Z. I., and Ratnarajah, T. (2014). On the security of cognitive radio networks: Cooperative jamming with relay selection. In *2014 European Conference on Networks and Communications (EuCNC)*, pages 1–5.

[68] Ma, M., Wang, H. M., Liu, F., and Wang, C. (2015). Precoding optimization for secure target user in multi-antenna broadcast channel. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5.

[69] Maculan, N., Santiago, C., Macambira, E., and Jardim, M. (2003). An O(n) Algorithm for Projecting a Vector on the Intersection of a Hyperplane and a Box in $\mathbb{R}^n$. *Journal of Optimization Theory and Applications*, 117(3):553–574.

[70] Mangasarian, O. (1969). *Nonlinear Programming*. McGraw-Hill, New York.

[71] Mangasarian, O. and Fromovitz, S. (1967). The fritz john necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17(1):37–47.

[72] Mu, P., Hu, X., Wang, B., and Li, Z. (2015). Secrecy rate maximization with uncoordinated cooperative jamming by single-antenna helpers under secrecy outage probability constraint. *IEEE Communications Letters*, 19(12):2174–2177.

[73] Murtagh, B. A. and Saunders, M. A. (1983). Minos 5.1 User's Guide. Technical Report SOL 83-20R, Stanford Univ., CA, USA.

[74] Oggier, F. and Hassibi, B. (2011). The Secrecy Capacity of the MIMO Wiretap Channel. *IEEE Transactions on Information Theory*, 57(8):4961–4972.

[75] Ouyang, J., Zhu, W. P., Massicotte, D., and Lin, M. (2016). Energy efficient optimization for physical layer security in cognitive relay networks. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6.

[76] Park, K. H., Wang, T., and Alouini, M. S. (2013). On the Jamming Power Allocation for Secure Amplify-and-Forward Relaying via Cooperative Jamming. *IEEE Journal on Selected Areas in Communications*, 31(9):1741–1750.

[77] Parsaeefard, S. and Le-Ngoc, T. (2015). Improving wireless secrecy rate via full-duplex relay-assisted protocols. *IEEE Transactions on Information Forensics and Security*, 10(10):2095–2107.

[78] Pham Dinh, T. and Le Thi, H. A. (1997). Convex analysis approach to DC programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–357.

[79] Pham Dinh, T. and Le Thi, H. A. (1998). A D.C. Optimization Algorithm for Solving the Trust-Region Subproblem. *SIAM Journal on Optimization*, 8(2):476–505.

[80] Pham Dinh, T. and Le Thi, H. A. (2014). Recent Advances in DC Programming and DCA. In *Transactions on Computational Intelligence XIII, Lecture Notes in Computer Science, Vol 8342*, pages 1–37. Springer.

[81] Phan, K., Le-Ngoc, T., Vorobyov, S. A., and Tellambura, C. (2009). Power allocation in wireless multi-user relay networks. *IEEE Trans. Wireless Commun.*, 8(5):2535–2545.

[82] Poulakis, M. I., Vassaki, S., and Panagopoulos, A. D. (2016). Secure Cooperative Communications Under Secrecy Outage Constraint: A DC Programming Approach. *IEEE Wireless Communications Letters*, 5(3):332–335.

[83] Rockafellar (1970). *Convex Analysis*. Princeton University.

[84] Rockafellar, Tyrrell, R., and Roger, J. B. (1998). *Variational Analysis*, volume 317. Spinger.

[85] Sarma, S., Agnihotri, S., and Kuri, J. (2016). Secure transmission in amplify-and-forward diamond networks with a single eavesdropper. In *2016 Twenty Second National Conference on Communication (NCC)*, pages 1–6.

[86] Schad, A., Law, K., and Pesavento, M. (2015). Rank-two beamforming and power allocation in multicasting relay networks. *IEEE Transactions on Signal Processing*, 63(13):3435–3447.

[87] Schad, A., Law, K. L., and Pesavento, M. (2012). A convex inner approximation techniqe for rank-two beamforming in multicasting relay networks. In *Proc. Eur. Signal Process. Conf.*, pages 1369–1373.

[88] Schad, A. and Pesavento, M. (2011). Multiuser bi-directional communications in cooperative relay networks. In *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 217–220.

[89] Shafiee, S., Liu, N., and Ulukus, S. (2008). Secrecy Capacity of the 2-2-1 Gaussian MIMO Wire-tap Channel. In *3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 207–212.

[90] Sidiropoulos, N. D., Davidson, T. N., and Luo, Z. Q. (2006). Transmit beamforming for physical-layer multicasting. *IEEE Trans. on Signal Process*, 54(6):2239–2251.

[91] Song, B., Lin, Y. H., and Cruz, R. (2008). Weighted max-min fair beamforming, power control and scheduling for a MISO downlink. *IEEE Trans. Wireless Commun.*, 7:464–469.

[92] Sriperumbudur, B. K. and Lanckriet, G. R. G. (2009). On the convergence of concave-convex procedure. In *Advances in Neural Information Processing Systems 22, NIPS2009*, pages 1759–1767. Curran Associates, Inc.

[93] Ta, A. S., Le Thi, H. A., Djamel, K., and Pham Dinh, T. (2010). Solving QoS Routing Problems by DCA. In *Intelligent Information and Database Systems, ACIIDS 2010, Lecture Notes in Computer Science, vol 5991*, pages 460–470. Springer.

[94] Ta, A. S., Le Thi, H. A., Djamel, K., and Pham Dinh, T. (2012a). Solving Partitioning-Hub Location-Routing Problem using DCA. *Journal of Industrial and Management Optimization*, 8(1):87–102.

[95] Ta, A. S., Pham Dinh, T., Le Thi, H. A., and Khadraoui, D. (2012b). Solving Many to many multicast QoS routing problem using DCA and proximal decomposition technique. In *2012 International Conference on Computing, Networking and Communications (ICNC)*, pages 809–814.

[96] Tang, X., Liu, R., Spasojevic, P., and Poor, H. (2007). Multiple Access Channels with Generalized Feedback and Confidential Messages. In *Information Theory Workshop, 2007. ITW '07. IEEE*, pages 608–613.

[97] Tekin, E., Serbetli, S., and Yener, A. (2005). On Secure Signaling for the Gaussian Multiple Access Wire-tap Channel. In *2005 Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, pages 1747–1751.

[98] Tekin, E. and Yener, A. (2008a). The Gaussian Multiple Access Wire-Tap Channel. *IEEE Transactions on Information Theory*, 54(12):5747–5755.

[99] Tekin, E. and Yener, A. (2008b). The General Gaussian Multiple-Access and Two-Way Wiretap Channels: Achievable Rates and Cooperative Jamming. *IEEE Transactions on Information Theory*, 54(6):2735–2751.

[100] Tran, T. T. and Le Thi, H. A. and Pham Dinh, T. (2016). DC programming and DCA for enhancing physical layer security via cooperative jamming. *Computers & Operations Research*.

[101] Vaz, A. I. F. and Vicente, L. (2007). A particle swarm pattern search method for bound constrained global optimization. *Journal of Global Optimization*, 39:197–219.

[102] Vishwakarma, S. and Chockalingam, A. (2014). MIMO decode-and-forward relay beamforming for secrecy with cooperative jamming. In *2014 Twentieth National Conference on Communications (NCC)*, pages 1–6.

[103] Vouyioukas, D. (2013). A Survey on Beamforming Techniques for Wireless MIMO Relay Networks. *International Journal of Antennas and Propagation*, 2013(Article ID 745018):21 pages.

[104] Vucic, N. and Schubert, M. (2010). DC programming approach for resource allocation in wireless networks. In *IEEE, Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pages 380–386.

[105] Wang, C. and Wang, H.-M. (2014). Joint relay selection and artificial jamming power allocation for secure DF relay networks. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 819–824.

[106] Wang, C., Wang, H. M., Ng, D. W. K., Xia, X. G., and Liu, C. (2015a). Joint Beamforming and Power Allocation for Secrecy in Peer-to-Peer Relay Networks. *IEEE Transactions on Wireless Communications*, 14(6):3280–3293.

[107] Wang, C., Wang, H. M., and Xia, X. G. (2015b). Hybrid Opportunistic Relaying and Jamming With Power Allocation for Secure Cooperative Networks. *IEEE Transactions on Wireless Communications*, 14(2):589–605.

[108] Wang, H. M., Liu, F., and Yang, M. (2015c). Joint Cooperative Beamforming, Jamming, and Power Allocation to Secure AF Relay Systems. *IEEE Transactions on Vehicular Technology*, 64(10):4893–4898.

[109] Wang, H. M., Luo, M., Xia, X. G., and Yin, Q. (2013). Joint Cooperative Beamforming and Jamming to Secure AF Relay Systems With Individual Power Constraint and No Eavesdropper's CSI. *IEEE Signal Processing Letters*, 20(1):39–42.

[110] Wang, L., Elkashlan, M., Huang, J., Tran, N. H., and Duong, T. Q. (2014). Secure Transmission with Optimal Power Allocation in Untrusted Relay Networks. *IEEE Wireless Communications Letters*, 3(3):289–292.

[111] Wang, Q., Chen, Z., Mei, W., and Fang, J. (2017). Improving physical layer security using uav-enabled mobile relaying. *IEEE Wireless Communications Letters*, PP(99):1–1.

[112] Wen, X., Law, K. L., Alabed, S. J., and Pesavento, M. (2012). Rank-two beamforming for single-group multicasting network using OSTBC. In *Proc. IEEE SAM'012*, pages 69–72. Hoboken, USA, IEEE.

[113] Wu, S. X., Ma, W. K., and So, A. M. (2013). Physical layer multicasting by stochastic transmit beamforming and Alamouti space-time coding. *IEEE Trans. on Signal Process*, 61(17):4230–4245.

[114] Wyner, A. D. (1975). The wire-tap channel. *Bell Sys. Tech. Journ.*, 54:1355–1387.

[115] Xing, H., Chu, Z., Ding, Z., and Nallanathan, A. (2014). Harvest-and-jam: Improving security for wireless energy harvesting cooperative networks. In *2014 IEEE Global Communications Conference*, pages 3145–3150.

[116] Yamamoto, H. (1986). On secret sharing communication systems with two or three channels. *IEEE Transactions on Information Theory*, 32(3):387–393.

[117] Yamamoto, H. (1991). A coding theorem for secret sharing communication systems with two Gaussian wiretap channels. *IEEE Transactions on Information Theory*, 37(3):634–638.

[118] Yang, J., Kim, I. M., and Kim, D. I. (2013a). Optimal Cooperative Jamming for Multiuser Broadcast Channel with Multiple Eavesdroppers. *IEEE Transactions on Wireless Communications*, 12(6):2840–2852.

[119] Yang, J., Li, Q., Cai, Y., Zou, Y., Hanzo, L., and Champagne, B. (2016). Joint secure af relaying and artificial noise optimization: A penalized difference-of-convex programming framework. *IEEE Access*, 4:10076–10095.

[120] Yang, Y., Li, Q., Ma, W. K., Ge, J., and Ching, P. C. (2013b). Cooperative Secure Beamforming for AF Relay Networks With Multiple Eavesdroppers. *IEEE Signal Processing Letters*, 20(1):35–38.

[121] Yang, Y., Sun, C., Zhao, H., Long, H., and Wang, W. (2014). Algorithms for Secrecy Guarantee With Null Space Beamforming in Two-Way Relay Networks. *IEEE Transactions on Signal Processing*, 62(8):2111–2126.

[122] Zangwill, W. I. (1969). *Nonlinear Programming: A Unified Approach*. Prentice-Hall, Englewood Cliffs, N.J.

[123] Zhang, J. and Gursoy, M. (2010a). Collaborative Relay Beamforming for Secrecy. In *2010 IEEE International Conference on Communications (ICC)*, pages 1–5.

[124] Zhang, J. and Gursoy, M. (2010b). Relay beamforming strategies for physical-layer security. In *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6.

[125] Zhang, Q., Huang, X., Li, Q., and Qin, J. (2015). Cooperative Jamming Aided Robust Secure Transmission for Wireless Information and Power Transfer in MISO Channels. *IEEE Transactions on Communications*, 63(3):906–915.

[126] Zhang, Y., Li, Q., Lin, J., and Wu, S. X. (2016). Robust secrecy rate optimization for full-duplex bidirectional communications. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6.

[127] Zheng, G., Arapoglou, P., and Ottersten, B. (2012). Physical Layer Security in Multibeam Satellite Systems. *IEEE Transactions on Wireless Communications*, 11(2):852–863.

[128] Zhou, J., Cao, R., Gao, H., Zhang, C., and Lv, T. (2015). Secure Beamforming Design in Wiretap MISO Interference Channels. In *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pages 1–5.

[129] Zou, Y., Wang, X., and Shen, W. (2013). Optimal relay selection for physical-layer security in cooperative wireless networks. *IEEE Journal on Selected Areas in Communications*, 31(10):2099–2111.