



Dissertation in fulfillment of the degree of

DOCTOR OF PHILOSOPHY IN APPLIED MATHEMATICS
UNIVERSITY PARIS VI

*Université Pierre et Marie Curie
École Doctorale de Sciences Mathématiques de Paris Centre*

Application of Polynomial Optimization to Electricity Transmission Networks

Cédric JOSZ

under the supervision of

Jean Charles GILBERT

French Institute for Research in Computer Science and Automation
Institut National de Recherche en Informatique et en Automatique

Jean MAEGHT and Patrick PANCIATICI

French Transmission System Operator

Réseau de Transport d'Électricité

funded by

French Transmission System Operator

French Ministry of Higher Education and Research

CIFRE ARNT contract 2013/0179

examined on July 13th 2016 by

Patrick Combettes	University Paris VI
Stéphane Gaubert	Ecole Polytechnique
Jean Charles Gilbert	INRIA Paris
Jean Bernard Lasserre	CNRS Toulouse
Patrick Panciatici	RTE Versailles
Mihai Putinar	UC Santa Barbara
Markus Schweighofer	University of Konstanz
Pascal Van Hentenryck	University of Michigan

« *Cherche et tu trouveras.* »

Angélique Haenecour Josz

Acknowledgements

I wish to thank my thesis advisor Jean Charles Gilbert for his support and guidance throughout my doctoral project. He has given me an invaluable insight into optimization and the world of research. He encouraged me to pursue my ideas and to publish them, as well to attend conferences and to meet with the scientific community. It was a pleasure to be his teaching assistant at the University of Paris-Saclay, ENSTA ParisTech, for two consecutive years. It is thanks to his teachings that I developed a strong background in optimization when I was a master's student at ENSTA ParisTech. Jean Charles mentioned the Lasserre hierarchy at the beginning of our first meeting three years ago. At the time, I didn't understand its relevance for electricity, but it turned out to be the cornerstone of this dissertation!

I would like to thank Jean Maeght, my scientific advisor at RTE, for mentoring me over the three years of a half years I spent at RTE. I was fortunate to share my office with someone who works on the optimal power flow problem in a European project and who has a very strong background in mathematics. He invested a lot of time in me to teach me about the company, its various projects, its key people, its history, and the way it functions on a daily basis. He always made sure that I was in the loop concerning events and meetings at RTE. I would have liked to spend more time sharing the office, but as he said so often after a long day of day of work, "*toute bonne chose ayant une fin ...*".

I am grateful to Patrick Panciatici, the chief scientific advisor of R&D at RTE, for being the main promoter of my doctoral project. Thanks to his daring vision, I was able to work on new and exciting ideas in the field of power systems. He gave me the opportunity to participate in a workshop in Dublin at IBM Research Ireland and to meet with top researchers in various universities and laboratories across the world. These include ETH Zürich, Berkeley, Caltech, and Tokyo Tech. It was thrilling to work with someone so passionate and knowledgeable about energy.

I thank my three advisors for the trust they put in me during the entire project. I was given a great amount of freedom to work on what pleased me most and to start collaborations with researchers in my field.

I wish to thank Stéphane Fliscounakis, research engineer at RTE, for his collaboration and assistance during my doctoral project. Thanks to his expertise on the issues and complexity of modeling transmission networks, we successfully made a large-scale representation of the European network easily available to the public for the first time. He taught me a great deal about the applications of optimization to power systems, and about the problems that arise when dealing with data in industry. I truly hope to pursue our joint work in the future.

I would like to thank Robert Gonzalez, optimization expert at RTE, for allowing me to participate in the optimization workshops he organizes. It was a great opportunity to learn about how optimization is used by RTE, as well as a chance to present my work in detail. That allowed me to get valuable feedback on my research.

I am grateful to my managers Frédérique Verrier and Lucian Baléa at RTE for supporting

my project. They always made sure that I was well integrated within my group at RTE. They gave me the opportunity to share my work with the R&D through various presentations.

I wish to thank Françoise Sericola, Lydie Pendu and Nathalie Lucazeau from HR at RTE for being very helpful on numerous occasions. They were crucial in setting up my contract as well as organizing my travels for RTE. I also wish to thank Nathalie Bonte from HR at INRIA for helping me get started quickly at INRIA.

I wish to thank Jean-Pierre Restoux, head of IT at RTE Versailles, as well as Gauthier Plouvier, Léon Amirkhanian, and Yacine Chaoui at IT support from RTE for being a great help throughout my years at RTE. Thanks to them, that I was able to work efficiently on my computer and use the latest versions of the softwares I needed.

I would like to thank all my colleagues at RTE and INRIA for creating an energetic environment to work in. I wish to thank Alexandre Debetencourt for his relentless support during my doctoral project. He stunned me by his natural curiosity and interest in my work.

I am fortunate to have collaborated with Professor Didier Henrion from the University of Toulouse. Working with a leader in the domain of optimization and control was an enriching experience. I am also fortunate to have collaborated with post-doctoral fellow Daniel K. Molzahn (now a research engineer at Argonne National Laboratory) and Professor Ian A. Hiskens from the University of Michigan. Many of the results in this dissertation were found thanks to our close collaboration. It was a very exciting experience that I hope to build on in the future.

I wish to thank Professors Pascal Gourdel and Bruno Nazaret for giving me the opportunity to teach at the University of Paris I Sorbonne-Tolbiac. It was a great opportunity for me to learn how to teach mathematics. Also from that university, I would like to thank Professor Jean-Bernard Baillon for his kindness and his help on some tough mathematics.

Many thanks to Professor Mihai Putinar at the University of California, Santa Barbara, for his encouragements, his valuable advice, and his generous help with my work.

Last but not least, many thanks to my brothers Tanguy and Jérôme, to my parents, and to my friends for their support.

Abstract

Transmission system operators need to adapt their decision-making tools to the technological evolutions of the twenty first century. A computation inherent to most tools seeks to find alternating-current power flows that minimize power loss or generation cost. Mathematically, it consists in an optimization problem that can be described using only addition and multiplication of complex numbers. The objective of this thesis is to find global solutions, in other words the best solutions to the problem. One of the outcomes of this highly collaborative doctoral project is to use recent results from algebraic geometry to compute globally optimal power flows in the European high-voltage transmission network.

Keywords: polynomial optimization, semidefinite optimization, optimal power flow, Lasserre hierarchy

Summary

This dissertation is motivated by an encouraging discovery made in the field of power systems during the first decade of the twenty-first century. Numerical experiments on several benchmark transmission networks showed that it is possible to find global solutions to the optimal power flow problem using semidefinite optimization. The optimal power flow problem seeks to find a steady-state operating point of an alternating-current transmission network that is optimal under some criteria such as power loss or generation costs. After five decades of research on this highly nonconvex problem, a method for finding global solutions was thought to be out of reach. The concept used was to omit nonconvexities and solve a convex problem instead. This is known as the Shor relaxation, in reference to the Ukrainian mathematician Naum Zuselevich Shor. However, the Shor relaxation does not provide global solutions to many networks of interest. Bridging this gap is the starting point of this dissertation.

The first step that was achieved (cf. Chapter 2) was to show that low orders of the Lasserre hierarchy find the global solution to small-scale networks that the Shor relaxation cannot solve. To do so, we realized that the optimal power flow problem is a particular instance of polynomial optimization. Thankfully, any polynomial optimization problem with a bounded feasible set can be approximated as close as desired by a sequence of semidefinite optimization problems. This sequence is called the Lasserre hierarchy, in reference to the French mathematician Jean Bernard Lasserre. This is remarkable because polynomial optimization problems encompass many non-deterministic polynomial-time hard problems such as quadratically-constrained quadratic programming, mixed-integer linear programming, and in particular the traveling salesman problem.

To further prove the numerical applicability of the Lasserre hierarchy, we proved that there is zero duality gap in each semidefinite optimization problem in the hierarchy in the case of the optimal power flow problem (cf. Chapter 3). This property is essential for efficient solvers to work. More generally, we proved that for any polynomial optimization problem containing a ball constraint, there is no duality gap. Adding a redundant ball constraint to a problem with a bounded feasible set guarantees the global convergence of the Lasserre hierarchy, hence the relevance of our result.

Having shown the applicability of the Lasserre hierarchy to small instances, the next task was to be able to tackle large-scale problems. However, there were few large-scale benchmark networks on which to test new approaches. Networks are considered large-scale if they contain several thousand buses. We filled this gap by providing data for the entire European synchronous grid, with a little over 9,000 buses. To make it possible to work progressively on the data, we provided four instances corresponding to larger and larger parts of the European network (cf. Chapter 4). The data stems from a European project involving many transmission system operators whose purpose was to develop new tools for the pan-European grid.

Since it had been discovered in 2000, the Lasserre hierarchy had never been able to solve practical problems with more than several dozens of variables. This changed when Daniel K.

Molzahn and Ian A. Hiskens at the University of Michigan developed an algorithm to exploit sparsity in the Lasserre hierarchy for the optimal power flow problem. This enabled them to solve networks with several hundred buses. At around the same time, Ramtin Madani, Morteza Ashraphijuo and Javad Lavaei at the University of Columbia showed that the Shor relaxation succeeded on some large-scale networks provided two penalty terms were added to the objective function. We proposed to combine both approaches to systematically provide nearly global solutions to large-scale networks (cf. Chapter 5). This work was carried out in collaboration with the University of Michigan. In the combined approach, only one penalty parameter has to be specified, instead of two. In the case of active power loss minimization, the objective function is convex (in function of the voltage variables) and we observed that no penalization term is needed. This means that the approach finds the global solution. In the case of generation cost minimization, the objective is not a convex function and a penalty parameter must be specified, yielding a nearly global solution.

Specifying a penalization parameter is problematic because there is no general method for doing so. To overcome this, we realized that successful penalizations of the optimal power flow were related to the Laplacian matrix of the graph of the power network. We thus proposed a Laplacian-based Shor relaxation to obtain nearly global solutions without the need to specify any parameter (cf. Chapter 6). An issue that emerged while trying to solve large-scale optimal power flow problems is that the data are ill-conditioned. Some power lines have very low impedance, i.e., opposition to current, while others have up to one thousand times larger impedance. As a result, in all large-scale numerical experiments in this dissertation, the data is preprocessed to have more homogenous line characteristics.

Having shown the applicability of the Lasserre hierarchy to large-scale networks, we next enhanced its tractability by transposing it from real to complex numbers (cf. Chapter 7). What prompts us to do so is that the optimal power flow problem is written using complex numbers. They are used to model an oscillatory phenomena, namely alternating-current. We realized that omitting nonconvexities and converting from complex to real numbers are two non-commutative operations. This led us to propose a general approach for finding global solutions to polynomial optimization problems with bounded feasible sets where variables and data are complex numbers. It is based on recent results in algebraic geometry concerning positive polynomials with complex indeterminates. By exploiting sparsity, it succeeds in finding global solutions to problems with several thousand complex variables. In addition to the operation and planning of future power systems, the complex moment/sum-of-squares hierarchy we developed can be applied to signal processing, imaging science, automatic control, and quantum mechanics.

The dissertation is organized as follows.

Chapter 1 describes the optimal power flow problem and the underlying mathematical concepts.

Chapter 2 numerically illustrates that low orders of the Lasserre hierarchy find the global solution to small-scale networks. Associated publication: C. JOSZ, J. MAEGHT, P. PANCIATICI, AND J.C. GILBERT, *Application of the Moment-SOS Approach to Global Optimization of the OPF Problem*, Institute of Electrical and Electronics Engineers, Transactions on Power Systems, 30, pp. 463–470, May 2014. [doi] [preprint]

Chapter 3 proves that there is no duality gap between the primal and dual versions of an instance of the Lasserre hierarchy in the presence of a ball constraint in the original polynomial problem. Associated publication: C. JOSZ AND D. HENRION, *Strong Duality in Lasserre’s Hierarchy for Polynomial Optimization*, Springer Optimization Letters, February 2015. [doi] [preprint]

Chapter 4 provides data of large-scale networks representing the European high-voltage transmission network. Associated public data: C. JOSZ, S. FLISCOUNAKIS, J. MAEGHT, AND P. PANCIATICI, *Power Flow Data of the European High-Voltage Transmission Network: 89, 1354, 2869, and 9241-bus PEGASE Systems*, MATPOWER 5.1, March 2015. [link]

Chapter 5 computes nearly global solutions to large-scale networks using the Lasserre hierarchy and a penalization parameter. Associated publication: D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *Solution of Optimal Power Flow Problems using Moment Relaxations Augmented with Objective Function Penalization*, 54th Annual Conference on Decision and Control, Osaka, December 2015. [preprint]

Chapter 6 computes nearly global solutions to large-scale networks using Laplacian matrices instead of a penalization parameter. Associated preprint: D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *A Laplacian-Based Approach for Finding Near Globally Optimal Solutions to OPF Problems*, submitted to Institute of Electrical and Electronics Engineers, Transactions on Power Systems. [preprint]

Chapter 7 transposes the Lasserre hierarchy to complex numbers to enhance its tractability when dealing with complex variables instead of real ones. Associated preprint: C. JOSZ, D. K. MOLZAHN, *Moment/Sum-of-Squares Hierarchy for Complex Polynomial Optimization*, submitted to Society for Industrial and Applied Mathematics, Journal on Optimization. [preprint]

Chapter 8 suggests future research directions and is followed by references.

The abstract and summary are translated in French in the following pages.

Abstract

Les gestionnaires des réseaux de transport d'électricité doivent adapter leurs outils d'aide à la décision aux avancées technologiques du XXI^{ème} siècle. Une opération sous-jacente à beaucoup d'outils est de calculer les flux en actif/réactif qui minimisent les pertes ou les coûts de production. Mathématiquement, il s'agit d'un problème d'optimisation qui peut être décrit en utilisant seulement l'addition et la multiplication de nombres complexes. L'objectif de cette thèse est de trouver des solutions globales. Un des aboutissements de ce projet doctoral hautement collaboratif est d'utiliser des résultats récents en géométrie algébrique pour calculer des flux optimaux dans le réseau Européen à haute tension.

Mots-clefs: hiérarchie de Lasserre, réseau de transport d'électricité, optimisation polynomiale, optimisation semidéfinie

Résumé

Cette thèse est motivée par une découverte encourageante faite dans le domaine des réseaux électriques durant la première décennie du XXI^{ème} siècle. Des expériences numériques sur certains cas tests ont montré qu'il était possible de trouver des solutions globales au problème d'écoulement des flux en utilisant l'optimisation semidéfinie positive. Le problème d'écoulement des flux recherche un point stationnaire du réseau qui est optimal au sens des pertes d'énergie ou des coûts de production. Après cinquante années de recherches sur ce problème non convexe, une méthode pour trouver des solutions globales semblait hors de portée. Le concept utilisé a été d'omettre les non convexités et de résoudre un problème convexe à la place. Ce procédé est connu sous le nom de relaxation de Shor, en référence au mathématicien ukrainien Naum Zuselevich Shor. Cependant, la relaxation de Shor ne fournit pas de solutions globales dans tous les cas. Pallier ce manque est le point de départ de cette dissertation.

La première étape qui a été franchie (cf. Chapitre 2) a été de montrer que l'on peut résoudre des petits réseaux à l'aide de la hiérarchie de Lasserre avec des ordres faibles lorsque la relaxation de Shor échoue. Nous nous sommes en effet aperçus que le problème de calcul des flux optimaux est une instance particulière d'optimisation polynomiale. Or tout problème d'optimisation polynomiale dont le domaine d'admissibilité est borné peut être approché d'aussi près que l'on veuille par une suite de problèmes d'optimisation semidéfinie positive. Cette suite est connue sous le nom de hiérarchie de Lasserre, en référence au mathématicien français Jean Bernard Lasserre. Ceci est remarquable car l'optimisation polynomiale englobe de nombreux problèmes NP-ardus tels que l'optimisation quadratique sous contraintes quadratiques, l'optimisation linéaire en nombres entiers, et en particulier le problème du voyageur de commerce.

Pour prouver davantage l'applicabilité de la hiérarchie de Lasserre d'un point de vue numérique, nous avons prouvé qu'il n'y a pas de saut de dualité pour chaque problème d'optimisation semidéfinie positive dans la hiérarchie, pour le cas du problème d'écoulement des flux (cf. Chapitre 3). Cette propriété est essentielle pour que des solveurs efficaces fonctionnent. Plus généralement, nous avons prouvé que pour tout problème d'optimisation contenant une contrainte de boule, il n'y a pas de saut de dualité. Ajouter une contrainte de boule redondante à un problème avec un ensemble admissible borné garantit la convergence de la hiérarchie de Lasserre, d'où la pertinence de notre résultat.

Après que l'applicabilité de la hiérarchie de Lasserre ait été démontrée pour des petits réseaux, la prochaine étape était de pouvoir traiter des réseaux de grande taille. Cependant, il y avait peu de cas tests sur lesquels tester de nouvelles approches. Les réseaux sont considérés de grande taille s'il contiennent plusieurs milliers de nœuds. Nous avons pallié ce manque en fournissant des données du réseau Européen synchrone, contenant un peu plus de 9.000 nœuds. Afin de pouvoir travailler progressivement sur ces données, nous avons fourni quatre instances correspondant à des parties de plus en plus grandes du réseau Européen (cf. Chapitre 4). Les données émanent d'un projet européen impliquant nombreux gestionnaires

de réseaux dont le but était de développer de nouveaux outils pour le réseau supra-national Européen.

Depuis sa découverte en 2000, la hiérarchie de Lasserre n'avait jamais résolu des problèmes provenant des applications avec plus de quelques dizaines de variables. Ceci changea lorsque Daniel K. Molzahn et Ian A. Hiskens à l'université du Michigan développèrent un algorithme pour exploiter le creux dans la hiérarchie Lasserre pour le problème d'écoulement des flux optimaux. Cela leur permit de s'attaquer à des réseaux avec quelques centaines de nœuds. A peu près au même moment, Ramtin Madani, Morteza Ashraphijuo, et Javad Lavaei à l'université de Columbia ont montré que la relaxation de Shor permet de résoudre certains réseaux de grande taille à condition d'ajouter deux termes de pénalisation à l'objectif. Nous avons proposé de combiner les deux approches afin d'apporter des solutions proches de l'optimum global de façon systématique (cf. Chapitre 5). Ce travail a été effectué en collaboration avec l'université du Michigan. Dans l'approche combinée, seul un paramètre de pénalisation doit être spécifié, au lieu de deux. Dans le cas de la minimisation des pertes, l'objectif est convexe (en fonction des variables de tensions) et nous avons observé qu'aucun terme de pénalisation n'est nécessaire. Cela signifie que l'approche trouve l'optimum global. Dans le cas de la minimisation des coûts de production, l'objectif n'est pas convexe et un terme de pénalisation doit être spécifié, ce qui donne lieu à une solution proche de l'optimum global.

Spécifier un paramètre de pénalisation est problématique car il n'existe pas de méthode générale pour le faire. Pour contourner ce problème, nous nous sommes aperçus que les pénalisations réussites étaient liées à la matrice de Laplace du graphe du réseau électrique. Nous avons donc proposé une relaxation de Shor basée sur la matrice de Laplace afin d'obtenir des solutions proches de l'optimum global sans avoir à spécifier un paramètre (cf. Chapitre 6). Un problème qui est survenu lorsque nous avons essayé de résoudre des problèmes de grande taille est que les données sont mal conditionnées. Certaines lignes ont des impédances très faibles alors que d'autres ont des impédances jusqu'à mille fois plus grandes. En conséquence, dans toutes les expérimentations à grande échelle, les données subissent un traitement préalable afin d'avoir des caractéristiques de lignes plus homogènes.

Ayant prouvé l'applicabilité de la hiérarchie de Lasserre aux réseaux de grande taille, nous avons ensuite réduit son temps de calcul en la transposant des nombres réels aux nombres complexes (cf. Chapitre 7). Ce qui nous a poussé à le faire est que le problème d'écoulement des flux est écrit en nombres complexes. Ceux-ci sont utilisés pour modéliser un phénomène oscillatoire, à savoir le courant alternatif. Nous nous sommes aperçus qu'omettre les non convexités et convertir des nombres complexes aux réels sont deux opérations non commutatives. Cela nous a conduit à proposer une approche générale pour trouver des solutions globales à des problèmes d'optimisation avec un domaine admissible borné où les variables et les données sont des nombres complexes. Elle est basée sur des résultats récents en géométrie algébrique concernant des polynômes strictement positifs avec des indéterminées complexes. En exploitant l'aspect creux, elle parvient à trouver des solutions globales à des problèmes à

plusieurs milliers de variables complexes. En plus de la gestion et la planification des réseaux d'électricité du futur, la hiérarchie complexe des moments et sommes de carrés que nous avons développée pourra être appliquée en traitement du signal, en imagerie, en automatique, et en mécanique quantique.

La thèse est organisée comme suit.

Le chapitre 1 décrit le problème de l'écoulement des flux dans un réseau de transport et les concepts mathématiques sous-jacents.

Le chapitre 2 illustre numériquement que des ordres faibles de la hiérarchie de Lasserre permettent de résoudre des réseaux de petite taille. Publication associée: C. JOSZ, J. MAEGHT, P. PANCIATICI, AND J.C. GILBERT, *Application of the Moment-SOS Approach to Global Optimization of the OPF Problem*, Institute of Electrical and Electronics Engineers, Transactions on Power Systems, 30, pp. 463–470, May 2014. [doi] [preprint]

Le chapitre 3 prouve qu'il n'y a pas de saut de dualité entre les versions primales et duales de la hiérarchie de Lasserre en présence d'une contrainte de boucle dans le problème d'optimisation initial. Publication associée: C. JOSZ AND D. HENRION, *Strong Duality in Lasserre's Hierarchy for Polynomial Optimization*, Springer Optimization Letters, February 2015. [doi] [preprint]

Le chapitre 4 fournit des données de grande taille représentant le réseau Européen à haute tension. Données publiques associées: C. JOSZ, S. FLISCOUNAKIS, J. MAEGHT, AND P. PANCIATICI, *Power Flow Data of the European High-Voltage Transmission Network: 89, 1354, 2869, and 9241-bus PEGASE Systems*, MATPOWER 5.1, March 2015. [link]

Le chapitre 5 calcule des solutions proches de l'optimum global pour des réseaux de grande taille à l'aide de la hiérarchie de Lasserre et d'un paramètre de pénalisation. Publication associée: D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *Solution of Optimal Power Flow Problems using Moment Relaxations Augmented with Objective Function Penalization*, 54th Annual Conference on Decision and Control, Osaka, December 2015. [preprint]

Le chapitre 6 calcule des solutions proches de l'optimum global pour des réseaux de grande taille à l'aide de matrices de Laplace au lieu d'un paramètre de pénalisation. Papier soumis associé: D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *A Laplacian-Based Approach for Finding Near Globally Optimal Solutions to OPF Problems*, submitted to Institute of Electrical and Electronics Engineers, Transactions on Power Systems. [preprint]

Le chapitre 7 transpose la hiérarchie de Lasserre aux nombres complexes afin de réduire les

temps de calculs lorsqu'on s'intéresse à des variables complexes au lieu de variables réelles. Papier soumis associé: C. JOSZ, D. K. MOLZAHN, *Moment/Sum-of-Squares Hierarchy for Complex Polynomial Optimization*, submitted to Society for Industrial and Applied Mathematics, Journal on Optimization. [preprint]

Le chapitre 8 suggère des pistes de recherches futures et est suivi des références.

Contents

1	Background and motivations	16
1.1	Optimal power flow problem	16
1.2	Optimality conditions in optimization	20
1.3	Convex relaxation of the optimal power flow problem	23
1.4	Definitions of complex numbers	29
2	Lasserre hierarchy for small-scale networks	31
2.1	Introduction	31
2.2	Polynomial optimization formulation	33
2.3	Moment-sos approach	34
2.4	Numerical results	37
2.5	Conclusion	43
3	Zero duality gap in the Lasserre hierarchy	44
3.1	Introduction	44
3.2	Proof	46
3.3	Conclusion	50
4	Data of European transmission network	52
4.1	case89pegase	53
4.2	case1354pegase	53
4.3	case2869pegase	53
4.4	case9241pegase	53
5	Penalized Lasserre hierarchy	56
5.1	Introduction	56
5.2	Preprocessing low-impedance lines	58
5.3	Moment relaxations and penalization	59
5.4	Penalization of reactive power generation	59
5.5	Moment+penalization approach	61
5.6	Numerical results	62

<i>CONTENTS</i>	15
5.7 Conclusion	65
6 Laplacian matrix approach	67
6.1 Introduction	67
6.2 Laplacian objective function	69
6.3 Determining Laplacian weights	71
6.4 Numerical results	74
6.5 Conclusion	78
7 Complex hierarchy for enhanced tractability	81
7.1 Introduction	81
7.2 Motivation	83
7.3 Complex moment/sum-of-squares hierarchy	87
7.4 Numerical results	103
7.5 Conclusion	108
8 Conclusion and perspectives	110
A Ring Homomorphism	113
B Rank-2 Condition	114
C Invariance of Shor Relaxation Bound	116
D Invariance of SDP-\mathbb{R} Relaxation Bound	117
E Discrepancy Between Second-Order Conic Relaxation Bounds	118
F Five-Bus Illustrative Example for Exploiting Sparsity	119
G Complex Hierarchy Applied to Optimal Power Flow	121

Chapter 1

Background and motivations

The industrial problem which motivates this work can be viewed as an optimization problem. In this chapter, the equations that define this problem are presented and their relevance in practice is discussed. Next, optimality conditions are presented for general optimization problems. These are crucial to current methods used by industry, and remain important for the approach investigated in this thesis. This approach consists in solving convex relaxations of the original nonconvex problem. The original problem is written using complex numbers and this thesis advocates the use of convex relaxations in complex numbers. To that end, several definitions of complex numbers are provided.

1.1 Optimal power flow problem

The optimal power flow is a central problem in electric power systems introduced half a century ago by Carpentier [25]. It seeks to find a steady state operation point of an alternating current transmission network that respects Kirchoff's laws, Ohm's law, and power balance equations. In addition, the point has to be optimal under a criteria such as generation costs. It must also satisfy operational constraints which include narrow voltage ranges around nominal values and line ratings to keep Joule heating to acceptable levels.

While many nonlinear methods [27, 50, 88, 132] have been developed to solve this notoriously difficult problem, there is a strong motivation for producing more robust and reliable tools. Firstly, electric power systems are growing in complexity due to the increase in the share of renewables, the increase in the peak load, and the expected wider use of demand response and storage. This could hamper power systems reliability if decision-making tools do not evolve. Costly power interruptions could occur more often. Secondly, new tools are needed to profit from high-performance computing and advances in telecommunications such as phasor measurement units and dynamic line ratings. This will reduce operation costs and help keep power supply affordable at a time when expensive investments are being made for renewables. Lastly, system operators face large-scale optimization problems with combinato-

rial complexity due to phase-shifting transformers, high-voltage direct current transmission lines, and special protection schemes. Solving the continuous case to global optimality would be of great benefit for a more automated decision process.

Electricity transmission networks are meshed networks in which buses not only inject or retrieve power from the network, but also serve as a relay for other buses. Topologically, there exists cycles in the network. This is not the case for distribution networks where the topology of the network is a tree¹. A simple model of high-voltage power lines in transmission networks uses a resistance R , an inductance L , and a capacitance C (cf. figure 1.1).

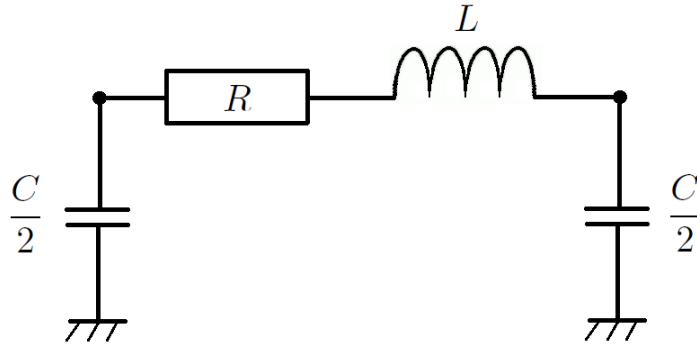


Figure 1.1: Π model of a high voltage transmission line

Continental Europe uses alternating current (AC) at a frequency of $50 \text{ Hz} \pm 0,5 \text{ Hz}$, which makes for an angular speed of $\omega \approx 2\pi \text{ rad} \times 50 \text{ Hz} \approx 314 \text{ rad.s}^{-1}$. The total impedance of a resistance R and an inductance L in series is $R + jL\omega$. 110 mH is a typical value for inductance in a 100 km long line operating at 400 kV line, so reactance $L\omega$ is roughly equal to $110 \text{ mH} \times 314 \text{ rad.s}^{-1} \approx 35 \Omega$. Divided by a hundred, the value lies in the range given in table 1.1.

	overhead line			underground cable	
	63 - 90 kV	225 kV	400 kV	63 kV	225 kV
resistance (Ω/km)	0.10 - 0.16	0.022 - 0.065	0.022 - 0.039	0.028 - 0.225	0.028 - 0.110
reactance (Ω/km)	0.4	0.29 - 0.41	0.32 - 0.43	0.104 - 0.134	0.107 - 0.134
capacitance (nF/km)	9.1 - 9.5	8.9 - 12.5	8.7 - 11.5	158 - 289	131 - 320

Table 1.1: Range of physical values in high voltage transmission lines and cables [15]

In order to switch from one of the voltage levels shown in table 1.1 to another, electricity transmission networks are equipped with an electrical device called *transformer*. In this work, it is assumed that power entering a transformer is equal to power exiting it. Such a

¹Optimization over distribution networks may involve graphs that are not trees however. This is due to different possible configurations of the connections between buses in the network.

transformer is said to be an *ideal transformer*. It is modeled by a complex number called *ratio*. The output voltage is equal to the input voltage divided by the ratio while the output current is equal to the input current multiplied by the conjugate of the ratio. This is visible in figure 1.2 (where $(\cdot)^H$ denotes the conjugate transpose). Regular transformers have a real ratio and some special transformers called *phase-shifting transformers* have a complex ratio.

Consider a non-zero integer $n \in \mathbb{N}^*$. We model an electricity transmission network by a set of buses $\mathcal{N} := \{1, \dots, n\}$ of which a subset $\mathcal{G} \subset \mathcal{N}$ is connected to generators. Let $s_k^{\text{gen}} = p_k^{\text{gen}} + jq_k^{\text{gen}} \in \mathbb{C}$ denote generated power at bus $k \in \mathcal{G}$. All buses are connected to a load (i.e. power demand). Let $s_k^{\text{dem}} = p_k^{\text{dem}} + jq_k^{\text{dem}} \in \mathbb{C}$ denote power demand at bus $k \in \mathcal{N}$. Let $v_k \in \mathbb{C}$ denote voltage at bus $k \in \mathcal{N}$ and $i_k \in \mathbb{C}$ denote current injected into the network at bus $k \in \mathcal{N}$. The convention used for current means that $v_k i_k^H$ is the power injected into the network at bus $k \in \mathcal{N}$. This means that $v_k i_k^H = -s_k^{\text{dem}}$ at bus $k \in \mathcal{N} \setminus \mathcal{G}$ and $v_k i_k^H = s_k^{\text{gen}} - s_k^{\text{dem}}$ at bus $k \in \mathcal{G}$.

The network links buses to one another through a set of lines $\mathcal{L} \subset \mathcal{N} \times \mathcal{N}$. A link between two buses is described in figure 1.2. In this figure, $y_{lm} \in \mathbb{C}$ denotes the mutual admittance between buses $(l, m) \in \mathcal{L}$ ($y_{ml} = y_{lm}$ for all $(l, m) \in \mathcal{L}$); $y_{lm}^{\text{gr}} \in \mathbb{C}$ denotes the admittance-to-ground at end l of line $(l, m) \in \mathcal{L}$; $\rho_{lm} \in \mathbb{C}$ denotes the ratio of the phase-shifting transformer at end l of line $(l, m) \in \mathcal{L}$ ($\rho_{lm} = 1$ if there is no transformer); and $i_{lm} \in \mathbb{C}$ denotes current injected in line $(l, m) \in \mathcal{L}$ at bus l .

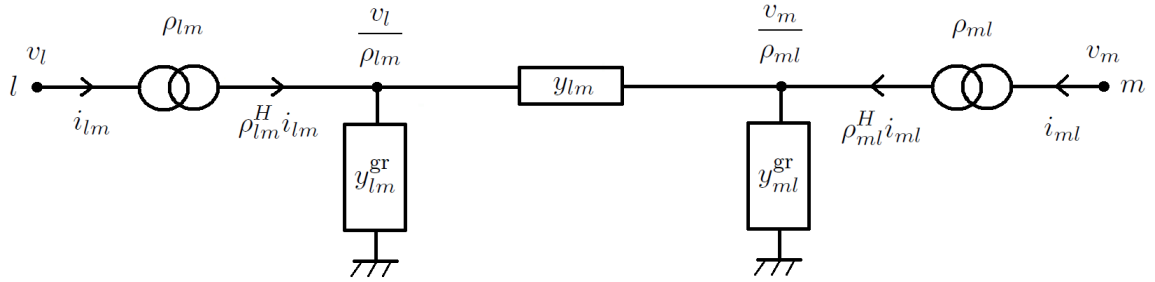


Figure 1.2: Link between buses l and m

A formulation of the optimal power flow problem is given in tables 1.2 (where $a_k, b_k, c_k \in \mathbb{R}$) and 1.3.

objective	description
$\min \sum_{k \in \mathcal{G}} a_k (p_k^{\text{gen}})^2 + b_k p_k^{\text{gen}} + c_k$	generation cost
variables	description
$(i_k)_{k \in \mathcal{N}}$ $(i_{lm})_{(l,m) \in \mathcal{L}}$ $(p_k^{\text{gen}})_{k \in \mathcal{N}}$ $(q_k^{\text{gen}})_{k \in \mathcal{N}}$ $(v_k)_{k \in \mathcal{N}}$	injected current current flow active generation reactive generation voltage
data	description
$(y_{lm})_{(l,m) \in \mathcal{L}}$ $(y_{lm}^{\text{gr}})_{(l,m) \in \mathcal{L}}$ $(\rho_{lm})_{(l,m) \in \mathcal{L}}$ $(p_k^{\text{dem}})_{k \in \mathcal{N}}$ $(q_k^{\text{dem}})_{k \in \mathcal{N}}$ $v_k^{\min}, v_k^{\max}, p_k^{\min}, p_k^{\max}, q_k^{\min}, q_k^{\max}$ $i_{lm}^{\max}, v_{lm}^{\max}, s_{lm}^{\max}, p_{lm}^{\max}$	mutual admittance admittance-to-ground ratio of (phase-shifting) transformer active power demand reactive power demand bounds at buses bounds on line flow

Table 1.2: Objective, variables, and data

$k \in \dots$ $(l, m) \in \dots$	constraints	description
\mathcal{N}	$i_l = \sum_{m \in \mathcal{N}(l)} i_{lm}$	Kirchoff's first law
\mathcal{L}	$\rho_{lm}^H i_{lm} = y_{lm}^{\text{gr}} \frac{v_l}{\rho_{lm}} + y_{lm} \left(\frac{v_l}{\rho_{lm}} - \frac{v_m}{\rho_{ml}} \right)$	Kirchoff's first law and Ohm's law
$\mathcal{N} \setminus \mathcal{G}$	$v_k i_k^H = -p_k^{\text{dem}} - j q_k^{\text{dem}}$	power demand
\mathcal{G}	$v_k i_k^H = p_k^{\text{gen}} - p_k^{\text{dem}} + j(q_k^{\text{gen}} - q_k^{\text{dem}})$	power demand and generation
\mathcal{G}	$p_k^{\min} \leq p_k^{\text{gen}} \leq p_k^{\max}$	bounds on active generation
\mathcal{G}	$q_k^{\min} \leq q_k^{\text{gen}} \leq q_k^{\max}$	bounds on reactive generation
\mathcal{N}	$v_k^{\min} \leq v_k \leq v_k^{\max}$	bounds on voltage amplitude
\mathcal{L}	$ v_l - v_m \leq v_{lm}^{\max}$	bound on voltage difference
\mathcal{L}	$ i_{lm} \leq i_{lm}^{\max}$	bound on current flow
\mathcal{L}	$ v_l i_{lm}^H \leq s_{lm}^{\max}$	bound on apparent power flow
\mathcal{L}	$ \text{Re}(v_l i_{lm}^H) \leq p_{lm}^{\max}$	bound on active power flow

Table 1.3: Constraints

According to the second constraint in table 1.3, for all $(l, m) \in \mathcal{L}$:

$$i_{lm} = \frac{y_{lm} + y_{lm}^{\text{gr}}}{|\rho_{lm}|^2} v_l - \frac{y_{lm}}{\rho_{ml} \rho_{lm}^H} v_m \quad (1.1)$$

Together with the first constraint in table 1.3, relationship (1.1) yields that for all $l \in \mathcal{N}$:

$$\begin{aligned} i_l &= \sum_{m \in \mathcal{N} \setminus \{l\}} i_{lm} \\ &= \sum_{m \in \mathcal{N} \setminus \{l\}} \frac{y_{lm} + y_{lm}^{\text{gr}}}{|\rho_{lm}|^2} v_l - \frac{y_{lm}}{\rho_{ml} \rho_{lm}^H} v_m \\ &= \left(\sum_{m \in \mathcal{N} \setminus \{l\}} \frac{y_{lm} + y_{lm}^{\text{gr}}}{|\rho_{lm}|^2} \right) v_l - \sum_{m \in \mathcal{N} \setminus \{l\}} \frac{y_{lm}}{\rho_{ml} \rho_{lm}^H} v_m \end{aligned}$$

Define the *admittance matrix* Y as the complex matrix of size $n \times n$ by:

$$Y_{lm} := \begin{cases} \sum_{m \in \mathcal{N} \setminus \{l\}} \frac{y_{lm} + y_{lm}^{\text{gr}}}{|\rho_{lm}|^2} & \text{if } l = m \\ -\frac{y_{lm}}{\rho_{ml} \rho_{lm}^H} & \text{if } l \neq m \end{cases}$$

Also, define $\mathbf{i} := (i_k)_{k \in \mathcal{N}}$ and $\mathbf{v} := (v_k)_{k \in \mathcal{N}}$. It follows that :

$$\mathbf{i} = Y \mathbf{v}$$

1.2 Optimality conditions in optimization

As mentioned in Section 1.1, current methods for solving the optimal power flow problem use nonlinear optimization techniques. These aim to find at least a solution to the optimality conditions, which we present in this section. Satisfaction of the optimality conditions does not guarantee global optimality for nonconvex problems, but they do for convex problems. The proposed approach in this dissertation uses the optimality conditions to solve convex relaxations of the optimal power flow problem, a concept presented in the next section.

To discuss optimality conditions, we consider a general framework that encompasses both the nonconvex and convex cases. Consider a finite dimensional normed vector space \mathbb{E} over \mathbb{R} or \mathbb{C} and an objective function $f : \mathbb{E} \rightarrow \mathbb{R}$. Also, consider a feasible set $X \subset \mathbb{E}$ described by a single function $c : \mathbb{E} \rightarrow \mathbb{F}$ where \mathbb{F} is a Hilbert space over \mathbb{R} or \mathbb{C} . The feasible set is also defined by a nonempty closed convex cone $K \subset \mathbb{F}$. Using these notations, the problem to be solved can be written:

$$\boxed{\inf_{x \in \mathbb{E}} f(x) \quad \text{subject to } c(x) \in K}$$

Objective function f and constraint function c will be considered twice differentiable. This is valid for the optimal power flow problem and its convex relaxations.

Relationship between local optimality and derivatives

Let's illustrate the relationship between local optimality and derivatives with a simple example. Consider a function $f : \mathbb{R} \rightarrow \mathbb{R}$ with a local minimum in 0 equal to 0. The first order Taylor series reads:

$$0 \leq f(t) = f'(0)t + o(t) = t[f'(0) + o(1)]$$

Thus $f'(0) = 0$. The second order Taylor series reads:

$$0 \leq f(t) = f'(0)t + \frac{f''(0)}{2}t^2 + o(t^2) = \frac{1}{2}t^2 [f''(0) + o(1)]$$

Thus $f''(0) \geq 0$. If $f''(0) > 0$ then the above line of equations tells us that 0 is a strict local minimum. Else if $f''(0) = 0$, the third order Taylor series reads:

$$0 \leq f(t) = f'(0)t + \frac{f''(0)}{2}t^2 + \frac{f'''(0)}{6}t^3 + o(t^3) = \frac{1}{6}t^3 [f'''(0) + o(1)]$$

Thus $f'''(0) = 0$. The fourth order Taylor series will then imply that $f''''(0) \geq 0$. We can again distinguish between strictly inequality and equality, and so on ...

In practice, only first order and second order conditions are considered because higher order derivatives are expensive to compute. Moreover, higher order derivatives are unrelated to local optimality unless all lower order derivatives zero out in one point. An example is $f(t) = t^4$ where $f'(0) = f''(0) = f'''(0) = 0$ and $f''''(0) = \frac{1}{4} > 0$. The fourth order strict inequality indicates that 0 is a strict local minimum. On the other hand, if one considers $\alpha \in \mathbb{R}$ and $f(t) = t^2 + \alpha t^3$, this yields $f'(0) = 0$, $f''(0) = 2 > 0$, and $f'''(0) = 6\alpha$ so that 0 is a strict local minimum regardless of the value of the third order derivative.

First and second order necessary optimality conditions

Let's go back to the general case where $f : \mathbb{E} \rightarrow \mathbb{R}$. When there are no constraints, a locally optimal point must be stationary, that is to say that first order derivatives must zero out in that point. Indeed, consider an optimal point x and write the first order Taylor series for h close to but different from zero:

$$0 \leq f(x+h) - f(x) = f'(x).h + o(\|h\|) = \|h\| \left[f'(x) \left(\frac{h}{\|h\|} \right) + o(1) \right]$$

Hence for all $\|d\| = 1$, one has $f'(x).d \geq 0$. Taking $d = -\nabla f(x)$ leads to $f'(x) = 0$.

Consider the second order Taylor series for h close to but different from zero:

$$0 \leq f(x+h) - f(x) = f'(x).h + f''(x)(h, h) + o(\|h\|^2) = \|h\|^2 \left[f''(x) \left(\frac{h}{\|h\|}, \frac{h}{\|h\|} \right) + o(1) \right]$$

Hence for all $\|d\| = 1$, one has $f''(x)(d, d) \geq 0$.

When there are constraints, establishing necessary optimality conditions requires the concept of duality in some way or another. Let's establish these conditions using the notion of saddle point in min-max duality. This implies an assumption of global rather than local

optimality. Min-max duality consists of writing the objective function as the supremum of a coupling function, that is to say : $f(x) = \sup_{y \in Y} \varphi(x, y)$. The optimization problem can thus be written, for some $X \subset \mathbb{E}$ and some $Y \subset \mathbb{F}$:

$$\inf_{x \in X} \sup_{y \in Y} \varphi(x, y)$$

If it is legitimate to swap inf and sup, i.e. $\inf_{x \in X} \sup_{y \in Y} \varphi(x, y) = \sup_{y \in Y} \inf_{x \in X} \varphi(x, y)$ (called *no duality gap*), and there exists a solution to the right-hand problem, then solving the original problem is equivalent to finding a saddle point of the coupling function. Indeed, (\bar{x}, \bar{y}) is saddle point of φ if and only if there is no duality gap and \bar{x} solves $\inf_{x \in X} \sup_{y \in Y} \varphi(x, y)$ and \bar{y} solves $\sup_{y \in Y} \inf_{x \in X} \varphi(x, y)$.

The Lagrange function $\varphi(x, y) = f(x) + \langle c(x), y \rangle$ where $X := \mathbb{E}$ and $Y := K^- := \{ y \in \mathbb{F} \mid \langle z, y \rangle \leq 0, \forall z \in K \}$ is an example of a coupling function with the enviable property that $\varphi(x, \cdot)$ is an affine function. Let's consider a saddle point $(\bar{x}, \bar{y}) \in X \times Y$ of the Lagrange function. By definition :

$$\forall x \in X, \forall y \in Y, \quad \varphi(\bar{x}, y) \leq \varphi(\bar{x}, \bar{y}) \leq \varphi(x, \bar{y})$$

The right-hand side inequality means that \bar{x} is an optimal solution of the unconstrained problem $\inf_{x \in \mathbb{E}} \varphi(x, \bar{y})$. Thus $\varphi'_x(\bar{x}, \bar{y}) = f'(\bar{x}) + c'(\bar{x})^* \bar{y} = 0$. Proceeding in the same fashion with the left-hand side inequality leads to a constrained optimization problem, so we will proceed differently. The left-hand side inequality implies:

$$\forall y \in K^-, \quad \langle c(\bar{x}), y \rangle \leq \langle c(\bar{x}), \bar{y} \rangle$$

Plugging in for $y = 0$ and $y = 2\bar{y}$ yields $\langle c(\bar{x}), \bar{y} \rangle = 0$, known as *complementary slackness*.

To sum up, if \bar{x} is primal optimal and \bar{y} is dual optimal and there is no duality gap, then:

$$\left\{ \begin{array}{l} \text{first order condition: } f'(\bar{x}) + c'(\bar{x})^* \bar{y} = 0 \\ \text{primal feasibility: } c(\bar{x}) \in K \\ \text{dual feasibility: } \bar{y} \in K^- \\ \text{complementary slackness: } \langle c(\bar{x}), \bar{y} \rangle = 0 \end{array} \right.$$

Note that min-max duality originates from the minimax theorem proven by Von Neumann in his 1928 paper *Zur Theorie der Gesellschaftsspiele*. A generalization of the minimax theorem states that if X and Y are nonempty convex sets, X is compact, and φ is continuous and convex-concave, then:

$$\inf_{x \in X} \sup_{y \in Y} \varphi(x, y) = \sup_{y \in Y} \inf_{x \in X} \varphi(x, y)$$

In general, the KKT conditions do not guarantee global optimality, nor even local optimality in fact. If the objective and constraint functions f and c are convex, then they guarantee global optimality. In the next section, nonconvexities are removed from the optimal power flow problem, yielding a relaxed convex problem. This convex problem is solved using interior-point methods which involve solving for KKT conditions.

1.3 Convex relaxation of the optimal power flow problem

Lavaei and Low [66] proposed a formulation of the optimal power flow problem where the variables are the real and imaginary parts of voltages at each bus. To do so, they defined a real vector $x = [\text{Re}(v) \ \text{Im}(v)]^T$ where $v \in \mathbb{C}^n$ are the complex voltages. Next they proposed a convex relaxation of the optimal power flow problem. We illustrate their work by considering an example of power loss minimization. The system of Figure 1.3 links a

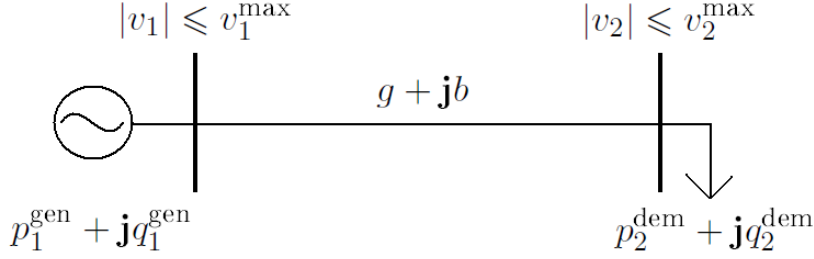


Figure 1.3: Two-Bus System

generator to a load via a line of admittance $g + \mathbf{i}b$ while respecting upper voltage constraints. Minimizing power loss reads

$$\inf_{v_1, v_2 \in \mathbb{C}} g |v_1|^2 - g \bar{v}_1 v_2 - g \bar{v}_2 v_1 + g |v_2|^2, \quad (1.2)$$

subject to

$$-\frac{g - \mathbf{i}b}{2} \bar{v}_1 v_2 - \frac{g + \mathbf{i}b}{2} \bar{v}_2 v_1 + g |v_2|^2 = -p_2^{\text{dem}}, \quad (1.3)$$

$$\frac{b + \mathbf{i}g}{2} \bar{v}_1 v_2 + \frac{b - \mathbf{i}g}{2} \bar{v}_2 v_1 - b |v_2|^2 = -q_2^{\text{dem}}, \quad (1.4)$$

$$|v_1|^2 \leq (v_1^{\text{max}})^2, \quad (1.5)$$

$$|v_2|^2 \leq (v_2^{\text{max}})^2, \quad (1.6)$$

where \mathbf{i} denotes the imaginary number. Identifying real and imaginary parts of the variables $v_1 =: x_1 + \mathbf{i}x_3$ and $v_2 =: x_2 + \mathbf{i}x_4$ leads to

$$\inf_{x_1, x_2, x_3, x_4 \in \mathbb{R}} gx_1^2 + gx_3^2 - 2gx_1x_2 - 2gx_3x_4 - gx_2^2 - gx_4^2, \quad (1.7)$$

subject to

$$-gx_1x_2 - gx_3x_4 - bx_1x_4 + bx_2x_3 + gx_2^2 + gx_4^2 + p_2^{\text{dem}} = 0, \quad (1.8)$$

$$bx_1x_2 + bx_3x_4 - gx_1x_4 + gx_2x_3 - bx_2^2 - bx_4^2 + q_2^{\text{dem}} = 0, \quad (1.9)$$

$$x_1^2 + x_2^2 \leq (v_1^{\text{max}})^2, \quad (1.10)$$

$$x_3^2 + x_4^2 \leq (v_2^{\text{max}})^2. \quad (1.11)$$

$$(1.12)$$

This problem can be rewritten as

$$\inf_y \quad gy_{11} + gy_{33} - 2gy_{12} - 2gy_{34} - gy_{22} - gy_{44}, \quad (1.13)$$

subject to

$$-gy_{12} - gy_{34} - by_{14} + by_{23} + gy_{22} + gy_{44} + p_2^{\text{dem}} = 0, \quad (1.14)$$

$$by_{12} + by_{34} - gy_{14} + gy_{23} - by_{22} - by_{44} + q_2^{\text{dem}} = 0, \quad (1.15)$$

$$y_{11} + y_{22} \leq (v_1^{\text{max}})^2, \quad (1.16)$$

$$y_{33} + y_{44} \leq (v_2^{\text{max}})^2, \quad (1.17)$$

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{14} \\ y_{12} & y_{22} & y_{23} & y_{24} \\ y_{13} & y_{23} & y_{33} & y_{34} \\ y_{14} & y_{24} & y_{34} & y_{44} \end{pmatrix} \succeq 0, \quad (1.18)$$

$$\text{rank}(y) \leq 1. \quad (1.19)$$

Removing the rank constraint leads to a convex relaxation of the optimal power flow problem.

As an extension of Lavaei and Low's work, Sojoudi and Lavaei [110] studied the theory behind optimization over graphs. To help with future work, we provide fully detailed proofs of results found in [110]. The reader may skip these by moving to Section 1.4 and still understand the rest of the dissertation.

Consider

$$\mathbb{H}_{+1}^n := \{ \mathbf{v}\mathbf{v}^H \mid \mathbf{v} \in \mathbb{C}^n \}.$$

Let \mathbf{i} denote the imaginary number. The notation of set \mathbb{H}_{+1}^n stems from the following proposition:

Proposition 1.1.

$$M \in \mathbb{H}_{+1}^n \iff M \in \mathbb{H}_+^n \text{ and } \text{rk}(M) \leq 1$$

Proof. (\implies) Consider a matrix $M \in \mathbb{H}_{+1}^n$, that is, there exists $\mathbf{v} \in \mathbb{C}^n$ such that $M = \mathbf{v}\mathbf{v}^H$. Firstly, observe that $M^H = (\mathbf{v}\mathbf{v}^H)^H = \mathbf{v}\mathbf{v}^H = M$. Secondly, for all $\mathbf{z} \in \mathbb{C}^n$, $\mathbf{z}^H M \mathbf{z} = \mathbf{z}^H \mathbf{v}\mathbf{v}^H \mathbf{z} = (\mathbf{v}^H \mathbf{z})^H \mathbf{v}^H \mathbf{z} = |\mathbf{v}^H \mathbf{z}|^2 \geq 0$. Lastly, each column of $M = \mathbf{v}\mathbf{v}^H$ is a linear combination of \mathbf{v} , so that the rank of M is at most 1.

(\impliedby) Matrix M is of rank at most 1 so there exists two vectors $\mathbf{t}, \mathbf{u} \in \mathbb{C}^n$ such that $M = \mathbf{t}\mathbf{u}^H$. Define $\mathbf{v} \in \mathbb{C}^n$ such that for all $1 \leq i \leq n$:

$$v_i := \sqrt{|t_i u_i|} e^{\mathbf{i} \arg(u_i)}$$

Thus :

$$\begin{aligned}
 v_i v_j^H &= \sqrt{|t_i u_i t_j u_j|} e^{i \arg(u_i)} e^{-i \arg(u_j)} \\
 &= \sqrt{|t_i u_j t_j u_i|} e^{i \arg(u_i)} e^{-i \arg(u_j)} \\
 &= \sqrt{|t_i u_j|^2} e^{i \arg(u_i)} e^{-i \arg(u_j)} \quad (|t_i u_j| = |t_j u_i| \text{ since } \mathbf{t} \mathbf{u}^H \in \mathbb{H}^n) \\
 &= |t_i u_j| e^{i \arg(u_i)} e^{-i \arg(u_j)} \\
 &= |t_i| e^{i \arg(u_i)} |u_j| e^{-i \arg(u_j)} \\
 &= |t_i| e^{i \arg(u_i)} u_j^H \\
 &= |t_i| e^{i \arg(t_i)} u_j^H \quad (\arg(t_i) \equiv \arg(u_i) [2\pi] \text{ since } t_i u_i^H \in \mathbb{R}_+ \text{ since } \mathbf{t} \mathbf{u}^H \succcurlyeq 0) \\
 &= t_i u_j^H
 \end{aligned}$$

Vector $\mathbf{v} \in \mathbb{C}^n$ thereby satisfies $\mathbf{v} \mathbf{v}^H = \mathbf{t} \mathbf{u}^H = M$. □

Given a set of edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$, define the following \mathbb{C} -linear operator :

$$\begin{aligned}
 \phi^{\mathcal{E}} : \mathcal{M}_n(\mathbb{C}) &\longmapsto \mathcal{M}_n(\mathbb{C}) \\
 M &\longrightarrow \phi^{\mathcal{E}}(M)_{ij} = \begin{cases} M_{ij} & \text{if } (i, j) \in \mathcal{E} \text{ or } i = j \\ 0 & \text{else} \end{cases}
 \end{aligned}$$

Given a graph \mathcal{L} , graph theory can be used to decompose the constraint $M \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}})$ into several smaller constraints. (The graph \mathcal{L} typically corresponds to the sparsity pattern. In a such sparse optimization problem, the constraint $M \in \mathbb{H}_{+1}^n$ may be replaced by $M \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}})$.) First, two lemmas are presented.

Lemma 1.2. *Any undirected connected graph has a spanning tree.*

Proof. Let \mathcal{L} be the set of edges of an undirected connected graph. Define the following set :

$$\mathcal{S} := \{ \mathcal{T} \in \mathcal{P}(\mathcal{L}) \mid \mathcal{T} \text{ is a tree } \}$$

(\mathcal{S}, \subset) is a partially ordered set since $(\mathcal{P}(\mathcal{L}), \subset)$ is a partially ordered set and $\mathcal{S} \subset \mathcal{P}(\mathcal{L})$. Consider a totally ordered subset of \mathcal{S} and name it \mathcal{U} . $\cup_{\mathcal{T} \in \mathcal{U}} \mathcal{T}$ is a bound of \mathcal{U} in \mathcal{S} . Zorn's lemma implies that \mathcal{S} contains a maximal element. If the maximal element is not a spanning tree, there exists a vertex not contained in it. Since the graph is connected, there exists a path in \mathcal{L} linking this vertex to a vertex in the maximum tree. The union of the maximum tree and that path forms a tree of \mathcal{L} that contradicts the maximality of the maximum tree. Thus the maximum tree is a spanning tree of \mathcal{L} . □

Lemma 1.3. *Let \mathcal{T} denote a spanning tree of a finite, undirected, and connected graph $(\mathcal{N}, \mathcal{L})$ and let $(\theta_{ij})_{(i,j) \in \mathcal{T}}$ denote some real numbers. Assume that $\theta_{ij} + \theta_{ji} \equiv 0 [2\pi]$ for all $(i, j) \in \mathcal{T}$. Then there exists some real numbers $(\theta_i)_{i \in \mathcal{N}}$ such that:*

$$\theta_i - \theta_j \equiv \theta_{ij} [2\pi] \quad , \quad \forall (i, j) \in \mathcal{T}$$

Proof. Simply define $(\theta_i)_{i \in \mathcal{N}}$ by choosing a random real number θ_{i_0} for some $i_0 \in \mathcal{N}$ and for all $i \in \mathcal{N} \setminus \{i_0\}$, choose $\theta_i \equiv \theta_{i_0} + \sum_{l,m} \theta_{lm} [2\pi]$ where the sum is taken over the path in \mathcal{T} linking i to i_0 . This is possible because $(\mathcal{N}, \mathcal{L})$ is a finite graph. Thus defined, the real numbers $(\theta_i)_{i \in \mathcal{N}}$ satisfy for all $(i, j) \in \mathcal{T}$:

$$\theta_i - \theta_j \equiv \theta_{ij} [2\pi]$$

because all terms in the sums associated respectively to θ_i and θ_j cancel each other out except for θ_{ij} . Indeed, $(i, j) \in \mathcal{T}$ so the paths starting at i and j and both ending at i_0 are the same except that one contains (i, j) or (j, i) and the other doesn't. If the former contains (i, j) , the only term that does not cancel out is θ_{ij} . If the latter contains (j, i) , the only term that does not cancel out is $-\theta_{ji} \equiv \theta_{ij} [2\pi]$. \square

Let $(c_i)_{1 \leq i \leq p}$ denote a cycle basis of \mathcal{L} and let $(b_i)_{1 \leq i \leq q}$ denote the set of bridge edges of \mathcal{L} . Define the following set :

$$\Omega := \underbrace{\{c_1, \dots, c_p\}}_{\text{cycle basis}} \cup \underbrace{\{b_1, \dots, b_q\}}_{\text{bridge edges}}$$

Proposition 1.4. *The following statement holds:*

$$\mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}}) = \bigcap_{\mathcal{E} \in \Omega} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{E}})$$

Proof. (C) It suffices to see that $\text{Ker}(\phi^{\mathcal{L}}) \subset \text{Ker}(\phi^{\mathcal{E}})$ for all $\mathcal{E} \in \Omega$. Indeed, $\mathcal{E} \in \Omega$ implies that $\mathcal{E} \subset \mathcal{L}$.

(D) Consider $M \in \bigcap_{\mathcal{E} \in \Omega} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{E}})$. For all $\mathcal{E} \in \Omega$, there exists $\mathbf{v}^{\mathcal{E}} \in \mathbb{C}^n$ and $N^{\mathcal{E}} \in \text{Ker}(\phi^{\mathcal{E}})$ such that :

$$M = \mathbf{v}^{\mathcal{E}} (\mathbf{v}^{\mathcal{E}})^H + N^{\mathcal{E}}$$

Thus, for all $i, j \in \mathcal{N}$:

$$M_{ij} = v_i^{\mathcal{E}} (v_j^{\mathcal{E}})^H + N_{ij}^{\mathcal{E}}$$

Moreover, $\phi^{\mathcal{E}}(N^{\mathcal{E}}) = 0$ so $N_{ij}^{\mathcal{E}} = 0$ if $(i, j) \in \mathcal{E}$ or $i = j$. It follows that for all $\mathcal{E} \in \Omega$:

$$\begin{cases} M_{ii} = v_i^{\mathcal{E}} (v_i^{\mathcal{E}})^H & , \text{ for all vertices } i \text{ of } \mathcal{E} & (1.20a) \\ M_{ij} = v_i^{\mathcal{E}} (v_j^{\mathcal{E}})^H & , \forall (i, j) \in \mathcal{E} & (1.20b) \end{cases}$$

\mathcal{L} is an undirected connected graph so there exists a spanning tree \mathcal{T} of \mathcal{L} according to lemma 1.2. The real numbers $(\arg(M_{ij}))_{(i,j) \in \mathcal{T}}$ satisfy $\arg(M_{ij}) + \arg(M_{ji}) \equiv 0 [2\pi]$ due to (1.20b). Indeed, given $(i, j) \in \mathcal{T}$, there exists some $\mathcal{E} \in \Omega$ such that $(i, j) \in \mathcal{E}$ and $(j, i) \in \mathcal{E}$ so that

$M_{ij} = v_i^{\mathcal{E}}(v_j^{\mathcal{E}})^H$ and $M_{ji} = v_j^{\mathcal{E}}(v_i^{\mathcal{E}})^H$. Moreover, the set of vertices \mathcal{N} is finite. Lemma 1.3 can thereby be applied to prove that there exists a vector $\mathbf{v} \in \mathbb{C}^n$ such that :

$$\begin{cases} |v_i| = \sqrt{M_{ii}} \quad , \quad \forall i \in \mathcal{N} & (1.21a) \\ \arg(v_i) - \arg(v_j) \equiv \arg(M_{ij}) [2\pi] \quad , \quad \forall (i, j) \in \mathcal{T} & (1.21b) \end{cases}$$

Notice that for all $(i, j) \in \mathcal{L}$,

$$\begin{aligned} |v_i v_j^H| &= |v_i| |v_j| \\ &= \sqrt{M_{ii}} \sqrt{M_{jj}} \quad \text{due to (1.21a)} \\ &= |v_i^{\mathcal{E}}| |v_j^{\mathcal{E}}| \quad \text{due to (1.20a) where } (i, j) \in \mathcal{E} \text{ for some } \mathcal{E} \in \Omega \\ &= |v_i^{\mathcal{E}}(v_j^{\mathcal{E}})^H| \\ &= |M_{ij}| \quad \text{due to (1.20b)} \end{aligned}$$

Moreover, for all $(i, j) \in \mathcal{L}$,

$$\begin{aligned} \arg(v_i v_j^H) &\equiv \arg(v_i) - \arg(v_j) [2\pi] \\ &\equiv \sum_{l,m} \arg(v_l) - \arg(v_m) [2\pi] && \text{telescoping sum along a path in } \mathcal{T} \text{ from } i \text{ to } j \\ &\equiv \sum_{l,m} \arg(M_{lm}) [2\pi] && \text{due to (1.21b) and } (l, m) \in \mathcal{T} \\ &\equiv \sum_{\mathcal{C}} \sum_{(l,m) \in \mathcal{C}} \arg(M_{lm}) [2\pi] && \text{sum over cycles } \mathcal{C} \in \Omega \text{ for which cycle defined} \\ &&& \text{by path in } \mathcal{T} \text{ has nonzero coordinates} \\ &\equiv \sum_{\mathcal{C}} \sum_{(l,m) \in \mathcal{C}} \arg(v_l^{\mathcal{C}}(v_m^{\mathcal{C}})^H) [2\pi] && \text{due to (1.20b)} \\ &\equiv \sum_{\mathcal{C}} \sum_{(l,m) \in \mathcal{C}} \arg(v_l^{\mathcal{C}}) - \arg(v_m^{\mathcal{C}}) [2\pi] && \text{telescoping sums over cycles are equal to zero} \\ &\equiv \sum_{l,m} \arg(v_l^{\mathcal{C}}) - \arg(v_m^{\mathcal{C}}) [2\pi] && \text{telescoping sum along path in } \mathcal{T} \text{ from } i \text{ to } j \\ &\equiv \arg(v_i^{\mathcal{C}}) - \arg(v_j^{\mathcal{C}}) [2\pi] \\ &\equiv \arg\{v_i^{\mathcal{C}}(v_j^{\mathcal{C}})^H\} [2\pi] \\ &\equiv \arg(M_{ij}) [2\pi] && \text{due to (1.20b)} \end{aligned}$$

To sum up :

$$\begin{cases} M_{ii} = v_i v_i^H \quad , \quad \forall i \in \mathcal{N} \\ M_{ij} = v_i v_j^H \quad , \quad \forall (i, j) \in \mathcal{L} \end{cases}$$

Therefore there exists $N^{\mathcal{L}} \in \text{Ker}(\phi^{\mathcal{L}})$ such that :

$$M = \mathbf{v} \mathbf{v}^H + N^{\mathcal{L}} \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}})$$

□

Lemma 1.5. *Let \mathcal{C} denote a cycle of a finite and undirected graph $(\mathcal{N}, \mathcal{L})$ and let $(\theta_{ij})_{(i,j) \in \mathcal{C}}$ denote some real numbers. Assume that $\theta_{ij} + \theta_{ji} \equiv 0 [2\pi]$ for all $(i, j) \in \mathcal{C}$ and that $\sum_{(i,j) \in \mathcal{C}} \theta_{ij} \equiv 0 [2\pi]$. Then for each vertex i of \mathcal{C} , there exists a real number θ_i such that:*

$$\theta_i - \theta_j \equiv \theta_{ij} [2\pi] \quad , \quad \forall (i, j) \in \mathcal{C}$$

Proof. Simply define the set of all θ_i 's for each vertex i of \mathcal{C} by choosing a random real number θ_{i_0} for some vertex i_0 of \mathcal{C} and for all other vertices $i \in \mathcal{N} \setminus \{i_0\}$, choose $\theta_i \equiv \theta_{i_0} + \sum_{l,m} \theta_{lm} [2\pi]$ where the sum is taken over the path in $\vec{\mathcal{C}}$ linking i to i_0 for some orientation of \mathcal{C} . This is possible because $(\mathcal{N}, \mathcal{L})$ is a finite graph. Thus defined, the real numbers θ_i satisfy for all $(i, j) \in \mathcal{C}$:

$$\theta_i - \theta_j \equiv \theta_{ij} [2\pi]$$

Indeed, for all $(i, j) \in \vec{\mathcal{C}}$ but the edge ending in i_0 , all terms in the sums associated respectively to θ_i and θ_j cancel each other out except for θ_{ij} . As for the edge (i, i_0) ending in i_0 , notice that $\theta_i \equiv \theta_{i_0} + \sum_{l,m} \theta_{lm} [2\pi]$ where the sum is taken over the path in $\vec{\mathcal{C}}$ linking i to i_0 so that $\sum_{(l,m) \in \vec{\mathcal{C}}} \theta_{lm} \equiv \theta_{i_0 i} + \sum_{l,m} \theta_{lm} [2\pi]$. It is assumed that $\sum_{(l,m) \in \vec{\mathcal{C}}} \theta_{lm} \equiv 0 [2\pi]$ thus $\sum_{l,m} \theta_{lm} \equiv -\theta_{i_0 i} [2\pi]$. It follows that $\theta_i \equiv \theta_{i_0} - \theta_{i_0 i} [2\pi]$, ie $\theta_i - \theta_{i_0} \equiv \theta_{i i_0} [2\pi]$. \square

Proposition 1.6. *The following statement holds:*

$$M \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}}) \iff M \in \bigcap_{(i,j) \in \mathcal{L}} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{(i,j)}) \ \& \ \sum_{(i,j) \in \vec{\mathcal{C}}} \arg(M_{ij}) \equiv 0 [2\pi] \text{ for all cycles } \mathcal{C} \in \Omega$$

Proof. (C) Consider $M \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}})$. For all $(i, j) \in \mathcal{L}$, $\text{Ker}(\phi^{\mathcal{L}}) \subset \text{Ker}(\phi^{(i,j)})$ thus $M \in \bigcap_{(i,j) \in \mathcal{L}} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{(i,j)})$.

Consider a cycle $\mathcal{C} \in \Omega$. There exists $\mathbf{v} \in \mathbb{C}^n$ such that for all $(i, j) \in \vec{\mathcal{C}}$, $M_{ij} = v_i v_j^H$. Therefore:

$$\begin{aligned} \sum_{(i,j) \in \vec{\mathcal{C}}} \arg(M_{ij}) &\equiv \sum_{(i,j) \in \vec{\mathcal{C}}} \arg(v_i v_j^H) [2\pi] \\ &\equiv \sum_{(i,j) \in \vec{\mathcal{C}}} \arg(v_i) - \arg(v_j) [2\pi] \quad \text{telescoping sum} \\ &\equiv 0 [2\pi] \end{aligned}$$

(D) Consider $M \in \bigcap_{(i,j) \in \mathcal{L}} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{(i,j)})$ such that $\sum_{(i,j) \in \vec{\mathcal{C}}} \arg(M_{ij}) \equiv 0 [2\pi]$ for all cycles $\mathcal{C} \in \Omega$. The following is true for all $(i, j) \in \mathcal{L}$:

$$\begin{cases} M_{kk} = v_k^{(i,j)} (v_k^{(i,j)})^H, & \text{for } k = i, j \end{cases} \quad (1.22a)$$

$$\begin{cases} M_{lm} = v_l^{(i,j)} (v_m^{(i,j)})^H, & \text{for } (l, m) = (i, j), (j, i) \end{cases} \quad (1.22b)$$

Consider a cycle $\mathcal{C} \in \Omega$. The real numbers $(\arg(M_{ij}))_{(i,j) \in \mathcal{C}}$ satisfy $\arg(M_{ij}) + \arg(M_{ji}) \equiv 0 [2\pi]$ due to (1.22b). Moreover, it is assumed that $\sum_{(i,j) \in \vec{\mathcal{C}}} \arg(M_{ij}) \equiv 0 [2\pi]$. Lemma 1.5 can therefore be used to show that there exists $\mathbf{v} \in \mathbb{C}^n$ such that:

$$\begin{cases} |v_i| = \sqrt{M_{ii}}, & \text{for all vertices } i \text{ of } \mathcal{C} \end{cases} \quad (1.23a)$$

$$\begin{cases} \arg(v_i) - \arg(v_j) \equiv \arg(M_{ij}) [2\pi], & \forall (i, j) \in \mathcal{C} \end{cases} \quad (1.23b)$$

Notice that for all $(i, j) \in \mathcal{C}$:

$$\begin{aligned}
|v_i v_j^H| &= |v_i| |v_j| \\
&= \sqrt{M_{ii}} \sqrt{M_{jj}} && \text{due to (1.23a)} \\
&= |v_i^{(i,j)}| |v_j^{(i,j)}| && \text{due to (1.22a)} \\
&= |v_i^{(i,j)} (v_j^{(i,j)})^H| \\
&= |M_{ij}| && \text{due to (1.22b)}
\end{aligned}$$

Thus for all $\mathcal{C} \in \Omega$:

$$M \in \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{C}})$$

Moreover, it is assumed that :

$$M \in \bigcap_{(i,j) \in \mathcal{L}} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{(i,j)})$$

Therefore :

$$M \in \bigcap_{\mathcal{E} \in \Omega} \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{E}}) = \mathbb{H}_{+1}^n + \text{Ker}(\phi^{\mathcal{L}})$$

where the equality follows from proposition 1.4. □

1.4 Definitions of complex numbers

Complex numbers are a central aspect of the thesis. They are used to model an oscillatory phenomenon, namely alternating current. We now consider several definitions of complex numbers.

A complex number $x + \mathbf{i}y$ can be thought of as the matrix:

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

The additions and multiplications of complex numbers translate into additions and multiplications of real matrices. Concerning addition, we have

$$\begin{aligned}
(a + \mathbf{i}b) + (c + \mathbf{i}d) &= (a + c) + \mathbf{i}(b + d) \\
\begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} c & -d \\ d & c \end{pmatrix} &= \begin{pmatrix} a + c & -b - d \\ b + d & a + c \end{pmatrix}
\end{aligned}$$

Concerning multiplication, we have:

$$\begin{aligned}
(a + \mathbf{i}b) \times (c + \mathbf{i}d) &= ac - bd + \mathbf{i}(ad + bc) \\
\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \times \begin{pmatrix} c & -d \\ d & c \end{pmatrix} &= \begin{pmatrix} ac - bd & -ad - bc \\ ad + bc & ac - bd \end{pmatrix}
\end{aligned}$$

In particular, we have

$$\begin{array}{ccccccc} \mathbf{i} & \times & \mathbf{i} & = & -1 \\ \left(\begin{array}{cc} 0 & -1 \\ 1 & 0 \end{array} \right) & \times & \left(\begin{array}{cc} 0 & -1 \\ 1 & 0 \end{array} \right) & = & \left(\begin{array}{cc} -1 & 0 \\ 0 & -1 \end{array} \right) \end{array}$$

As stated above, a complex number can be viewed as a real matrix of size 2×2 . More generally, a complex square matrix can be viewed as a real matrix of double its size. This will become very helpful when we consider optimization over complex matrix variables in Chapter 7.

We note that complex numbers can also be defined using the Euclidian division of polynomials. The idea is to build a solution to the equation $x^2 = -1$ though it has no real solution. To do so, consider the ring of polynomials $\mathbb{R}[X]$ with one real indeterminate X . The remainder of X^2 when divided by $1 + X^2$ is equal to -1 , which is written $X^2 \equiv -1 [1 + X^2]$. More generally, given a polynomial $P \in \mathbb{R}[X]$, its division by $1 + X^2$ has a remainder of the form $a + bX$ where a and b are some real numbers. Indeed, the degree of the remainder must be strictly less than the degree of $1 + X^2$. Let $\text{cl}(P)$ denote the equivalence class modulo $1 + X^2$ represented by P . The set of equivalence classes is thus equal to $\{ \text{cl}(a + bX) \mid a, b \in \mathbb{R} \}$. This set may be identified with the set of complex numbers because it is isomorphic to it. Indeed, consider two classes $\text{cl}(a + bX)$ and $\text{cl}(c + dX)$ with $a, b, c, d \in \mathbb{R}$. Concerning addition, we have

$$\text{cl}(a + bX) + \text{cl}(c + dX) = \text{cl}((a + c) + (b + d)X)$$

since

$$(a + bX) + (c + dX) \equiv (a + c) + (b + d)X [1 + X^2].$$

Concerning multiplication, we have

$$\text{cl}(a + bX) \times \text{cl}(c + dX) = \text{cl}(ac - bd + (ad + bc)X)$$

since

$$\begin{aligned} (a + bX)(c + dX) &\equiv ac + bdX^2 + (ad + bc)X && [1 + X^2] \\ &\equiv ac - bd + (ad + bc)X && [1 + X^2] \\ &(\text{because } X^2 \equiv -1 [1 + X^2]). \end{aligned}$$

Chapter 2

Lasserre hierarchy for small-scale networks

Finding a global solution to the optimal power flow (OPF) problem is difficult due to its nonconvexity. A convex relaxation in the form of semidefinite optimization (SDP) had attracted much attention when I started my Ph.D. Indeed, it yielded a global solution in several practical cases. However, it did not in all cases, and such cases had been documented in several publications. Here we present another SDP method known as the moment-sos (sum of squares) approach, which generates a sequence that converges towards a global solution to the OPF problem at the cost of higher runtime. Our finding is that in the small examples where the previously studied SDP method fails, this approach finds the global solution. The higher cost in runtime is due to an increase in the matrix size of the SDP problem, which can vary from one instance to another. Numerical experiment shows that the size is very often a quadratic function of the number of buses in the network, whereas it is a linear function of the number of buses in the case of the previously studied SDP method. The material in this chapter is based on the publication:

C. JOSZ, J. MAEGHT, P. PANCIATICI, AND J.C. GILBERT, *Application of the Moment-SOS Approach to Global Optimization of the OPF Problem*, Institute of Electrical and Electronics Engineers, Transactions on Power Systems, 30, pp. 463–470, May 2014. [doi] [preprint]

2.1 Introduction

The optimal power flow can be cast as a nonlinear optimization problem which is NP-hard, as was shown in [66]. So far, the various methods [50, 88] that have been investigated to solve the OPF can only guarantee local optimality, due to the nonconvexity of the problem. Recent progress suggests that it may be possible to design a method, based on semidefinite optimization (SDP), that yields global optimality rapidly.

The first attempt to use SDP to solve the OPF problem was made by Bai et al. [13] in 2008. In [66], Lavaei and Low show that the OPF can be written as an SDP problem, with an additional constraint imposing that the rank of the matrix variable must not exceed 1. They discard the rank constraint, as it is done in Shor's relaxation [106], a procedure which applies to quadratically constrained quadratic problems (see [71, 116] and the references therein). They also accept quartic terms that appear in some formulations of the OPF, transforming them by Schur's complement. Their finding is that for all IEEE benchmark networks, namely the 9, 14, 30, 57, 118, and 300-bus systems, the rank constraint is satisfied if a small resistance is added in the lines of the network that have zero resistance. Such a modification to the network is acceptable because in reality, resistance is never equal to zero.

There are cases when the rank constraint is not satisfied and a global solution can thus not be found. Lesieutre et al. [67] illustrate this with a practical 3-bus cyclic network. Gopalakrishnan et al. [45] find yet more examples by modifying the IEEE benchmark networks. Bukhsh et al. [22] provide a 2-bus and a 5-bus example. In addition, they document the local solutions to the OPF in many of the above-mentioned examples where the rank constraint is not satisfied [23].

Several papers propose ways of handling cases when the rank constraint is not satisfied. Gopalakrishnan et al. [45] propose a branch and reduce algorithm. It is based on the fact that the rank relaxation gives a lower bound of the optimal value of the OPF. But according to the authors, using the classical Lagrangian dual to evaluate a lower bound is about as efficient. Sojoudi and Lavaei [110] prove that if one could add controllable phase-shifting transformers to every loop in the network and if the objective is an increasing function of generated active power, then the rank constraint is satisfied. Though numerical experiments confirm this [41], such a modification to the network is not realistic, as opposed to the one mentioned earlier.

Cases where the rank constraint holds have been identified. Authors of [19, 111, 130] prove that the rank constraint is satisfied if the graph of the network is acyclic and if load over-satisfaction is allowed. This is typical of distribution networks but it is not true of transmission networks.

This paper examines the applicability of the moment-sos (sum of squares) approach to the OPF. This approach [61, 62, 89] aims at finding global solutions to polynomial optimization problems, of which the OPF is a particular instance. The approach can be viewed as an extension of the SDP method of [66]. Indeed, it proposes a sequence of SDP relaxations whose first element is the rank relaxation in many cases. The subsequent relaxations of the sequence become more and more accurate. When the rank relaxation fails, it is therefore natural to see whether the second order relaxation provides the global minimum, then the third, and so on.

The limit to this approach is that the complexity of the relaxations rapidly increases. The matrix size of the SDP relaxation of order d is roughly equal to the number of buses in the network to the power d . Surprisingly, in the 2, 3, and 5-bus systems found in [22, 67] where

the rank relaxation fails, the second order relaxation nearly always finds the global solution.

Below, section 2.2 shows that the OPF can be viewed as a polynomial optimization problem. The moment-sos approach which aims at solving such problems is described in section 2.3. In section 2.4, numerical results show that this approach successfully finds the global solution to the 2, 3, and 5-bus systems mentioned earlier. Conclusions are given in section 2.5.

2.2 Polynomial optimization formulation

In order to obtain a polynomial formulation of the OPF, we proceed in 3 steps. First, we write a formulation in complex numbers. Second, we use it to write a formulation in real numbers. Third, we use the real formulation to write a polynomial formulation.

Let \mathbf{a}^H and A^H denote the conjugate transpose of a complex vector \mathbf{a} and of a complex matrix A respectively. It can be deduced from [110] that there exist finite sets \mathcal{I} and \mathcal{J} , Hermitian matrices $(A_k)_{k \in \mathcal{G}}$ of size n , complex matrices $(B_i)_{i \in \mathcal{I}}$ and $(C_i)_{i \in \mathcal{J}}$ of size n , and complex numbers $(b_i)_{i \in \mathcal{I}}$ and $(c_i)_{i \in \mathcal{J}}$ such that the OPF can be written as

$$\min_{\mathbf{v} \in \mathbb{C}^n} \sum_{k \in \mathcal{G}} c_{k2} (\mathbf{v}^H A_k \mathbf{v})^2 + c_{k1} \mathbf{v}^H A_k \mathbf{v} + c_{k0}, \quad (2.1)$$

subject to

$$\forall i \in \mathcal{I}, \quad \mathbf{v}^H B_i \mathbf{v} \leq b_i, \quad (2.2)$$

$$\forall i \in \mathcal{J}, \quad |\mathbf{v}^H C_i \mathbf{v}| \leq c_i. \quad (2.3)$$

Constraints (2.3) correspond to bounds on apparent power flow. Constraints (2.2) correspond to all other constraints.

Let $\mathbf{x} \in \mathbb{R}^{2n}$ denote $[\operatorname{Re}(\mathbf{v})^T \operatorname{Im}(\mathbf{v})^T]^T$ as is done in [66]. In order to transform the complex formulation of the OPF (2.1)-(2.3) into a real number formulation, observe that $\mathbf{v}^H M \mathbf{v} = (\mathbf{x}^T M^{\operatorname{re}} \mathbf{x}) + j(\mathbf{x}^T M^{\operatorname{im}} \mathbf{x})$, where the superscript T denotes transposition,

$$M^{\operatorname{re}} := \begin{bmatrix} \operatorname{Re}(M) & -\operatorname{Im}(M) \\ \operatorname{Im}(M) & \operatorname{Re}(M) \end{bmatrix}, \quad \text{and}$$

$$M^{\operatorname{im}} := \begin{bmatrix} \operatorname{Im}(M) & \operatorname{Re}(M) \\ -\operatorname{Re}(M) & \operatorname{Im}(M) \end{bmatrix}.$$

Then (2.1)-(2.3) becomes

$$\min_{\mathbf{x} \in \mathbb{R}^{2n}} \sum_{k \in \mathcal{G}} c_{k2} (\mathbf{x}^T A_k^{\operatorname{re}} \mathbf{x})^2 + c_{k1} \mathbf{x}^T A_k^{\operatorname{re}} \mathbf{x} + c_{k0}, \quad (2.4)$$

subject to

$$\forall i \in \mathcal{I}, \quad \mathbf{x}^T B_i^{\operatorname{re}} \mathbf{x} \leq \operatorname{Re}(b_i), \quad (2.5)$$

$$\forall i \in \mathcal{I}, \quad \mathbf{x}^T B_i^{\text{im}} \mathbf{x} \leq \text{Im}(b_i), \quad (2.6)$$

$$\forall i \in \mathcal{J}, \quad (\mathbf{x}^T C_i^{\text{re}} \mathbf{x})^2 + (\mathbf{x}^T C_i^{\text{im}} \mathbf{x})^2 \leq c_i^2. \quad (2.7)$$

We recall that a polynomial is a function $p : \mathbf{x} \in \mathbb{R}^n \mapsto \sum_{\alpha \in \mathcal{A}} p_\alpha \mathbf{x}^\alpha$, where $\mathcal{A} \subset \mathbb{N}^n$ is a finite set of integer multi-indices, the coefficients p_α are real numbers, and \mathbf{x}^α is the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Its degree, denoted $\deg p$, is the largest $|\alpha| = \sum_{i=1}^n \alpha_i$ associated with a nonzero p_α .

The formulation of the OPF in real numbers (2.4)-(2.7) is said to be a polynomial optimization problem since the functions that define it are polynomials. Indeed, the objective (2.4) is a polynomial of $\mathbf{x} \in \mathbb{R}^n$ of degree 4, the constraints (2.5)-(2.6) are polynomials of \mathbf{x} of degree 2, and the constraints (2.7) are polynomials of \mathbf{x} of degree 4.

Formulation (2.4)-(2.7) will however not be used below because it has infinitely many global solutions. Indeed, formulation (2.1)-(2.3) from which it derives is invariant under the change of variables $\mathbf{v} \rightarrow \mathbf{v}e^{j\theta}$ where $\theta \in \mathbb{R}$. This invariance property transfers to (2.4)-(2.7). An optimization problem with non isolated solutions is generally more difficult to solve than one with a unique solution [17]. This feature manifests itself in some properties of the moment-sos approach described in section 2.3. For this reason, we choose to arbitrarily set the voltage phase at bus n to zero. Bearing in mind that $v_n^{\min} \geq 0$, this can be done by replacing voltage constraint (2.8) at bus n by (2.9):

$$(v_n^{\min})^2 \leq x_n^2 + x_{2n}^2 \leq (v_n^{\max})^2, \quad (2.8)$$

$$x_{2n} = 0 \quad \text{and} \quad v_n^{\min} \leq x_n \leq v_n^{\max}. \quad (2.9)$$

In light of (2.9), a polynomial optimization problem where there are $2n - 1$ variables instead of $2n$ variables can be formulated. More precisely, the OPF can be cast as the following polynomial optimization problem

PolyOPF:

$$\min_{\mathbf{x} \in \mathbb{R}^{2n-1}} f_0(\mathbf{x}) := \sum_{\alpha} f_{0,\alpha} \mathbf{x}^\alpha, \quad (2.10)$$

subject to

$$\forall i = 1, \dots, m, \quad f_i(\mathbf{x}) := \sum_{\alpha} f_{i,\alpha} \mathbf{x}^\alpha \geq 0, \quad (2.11)$$

where m is an integer, $f_{i,\alpha}$ denotes the real coefficients of the polynomial functions f_i , and summations take place over \mathbb{N}^{2n-1} . The summations are nevertheless finite because only a finite number of coefficients are nonzero.

2.3 Moment-sos approach

We first review some theoretical aspects of the moment-sos approach (a nice short account can be found in [9], and more in [16, 64]). Next, we present a set of relaxations of PolyOPF

obtained by this method and illustrate it on a simple example. Finally, we emphasize the relationship between the moment-sos approach and the rank relaxation of [66].

The moment-sos approach has been designed to find global solutions to polynomial optimization problems. It is grounded on deep results from real algebraic geometry. The term *moment-sos* derives from the fact that the approach has two dual aspects: the moment and the sum of squares approaches. Both approaches are dual of one another in the sense of Lagrangian duality [99]. Below, we focus on the moment approach because it leads to SDP problems that have a close link with the previously studied SDP method in [66].

Let \mathbf{K} be a subset of \mathbb{R}^{2n-1} . The moment approach rests on the surprising (though easy to prove) fact that the problem $\min\{f_0(\mathbf{x}): \mathbf{x} \in \mathbf{K}\}$ is equivalent to the convex optimization problem in μ :

$$\min_{\substack{\mu \text{ positive measure on } \mathbf{K} \\ \int d\mu=1}} \int f_0 d\mu. \quad (2.12)$$

Although the latter problem has a simple structure, it cannot be solved directly, since its unknown μ is an infinite dimensional object. Nevertheless, the realized transformation suggests that the initial difficult global optimization problem can be structurally simplified by judiciously expressing it on a space of larger dimension. The moment-sos approach goes along this way by introducing a hierarchy of more and more accurate approximations of problem (2.12), hence (2.10)-(2.11), defined on spaces of larger and larger dimension.

When f_0 is a polynomial and $\mathbf{K} := \{x \in \mathbb{R}^{2n-1}: f_i(\mathbf{x}) \geq 0, \text{ for } i = 1, \dots, m\}$ is defined by polynomials f_i like in PolyOPF, it becomes natural to approximate the measure μ by a finite number of its moments. The *moment* of μ , associated with $\alpha \in \mathbb{N}^{2n-1}$, is the real number $y_\alpha := \int \mathbf{x}^\alpha d\mu$. Then, when f_0 is the polynomial in (2.10), the objective of (2.12) becomes $\int f_0 d\mu = \int (\sum_\alpha f_{0,\alpha} \mathbf{x}^\alpha) d\mu = \sum_\alpha f_{0,\alpha} \int \mathbf{x}^\alpha d\mu = \sum_\alpha f_{0,\alpha} y_\alpha$, whose linearity in the new unknown y is transparent. The constraint $\int d\mu = 1$ is also readily transformed into $y_0 = 1$. In contrast, expressing which are the vectors y that are moments of a positive measure μ on \mathbf{K} (the other constraint in (2.12)) is a much more difficult task known as the *moment problem*, which has been studied for over a century [97]. It is that constraint that is approximated in the moment-sos approach, with more and more accuracy in spaces of higher and higher dimension.

The sum of squares approach is dual to the moment approach in the sense of Lagrangian duality [99]. It relies on the fact that minimizing a function f_0 over a set \mathbf{K} is equivalent to maximizing a real number λ under the constraints $f_0(\mathbf{x}) - \lambda \geq 0$ for all $\mathbf{x} \in \mathbf{K}$. These trivial *linear* constraints are intractable because there is an *infinite* number of them. In the case of polynomial optimization, one recovers the problem of finding certificates ensuring the positivity of the polynomial $f_0 - \lambda$ on the semi-algebraic set \mathbf{K} , which involves sums of squares of polynomials [76]. Relaxations consist in imposing degree bounds on these sos polynomials.

Lasserre [62] proposes a sequence of relaxations for any polynomial optimization problem like PolyOPF that grow better in accuracy and bigger in size when the order d of the relax-

ation increases. Here and below, d is an integer larger than or equal to each $v_i := \lceil (\deg f_i)/2 \rceil$ for all $i = 0, \dots, m$ (we have denote by $\lceil \cdot \rceil$ the ceiling operator).

Let $Z \succcurlyeq 0$ denote that Z is a symmetric positive semidefinite matrix. Define $\mathbb{N}_q^p := \{\alpha \in \mathbb{N}^p : |\alpha| \leq q\}$, whose cardinality is $|\mathbb{N}_q^p| = \binom{p+q}{q} := (p+q)!/(p!q!)$, and denote by $(z_{\alpha,\beta})_{\alpha,\beta \in \mathbb{N}_q^p}$ a matrix indexed by the elements of \mathbb{N}_q^p .

Relaxation of order d :

$$\min_{(y_\alpha)_{\alpha \in \mathbb{N}_{2d}^{2n-1}}} \sum_{\alpha} f_{0,\alpha} y_\alpha, \quad (2.13)$$

subject to

$$y_0 = 1, \quad (2.14)$$

$$(y_{\alpha+\beta})_{\alpha,\beta \in \mathbb{N}_d^{2n-1}} \succcurlyeq 0, \quad (2.15)$$

$$\forall i = 1, \dots, m, \quad \sum_{\gamma} f_{i,\gamma} (y_{\alpha+\beta+\gamma})_{\alpha,\beta \in \mathbb{N}_{d-v_i}^{2n-1}} \succcurlyeq 0. \quad (2.16)$$

We have already discussed the origin of (2.13)-(2.14) in the above SDP problem, while (2.15)-(2.16) are necessary conditions to ensure that y is formed of moments of some positive measure on \mathbf{K} . When d increases, these problems form a *hierarchy of semidefinite relaxations*, called that way because the objective (2.13) is not affected and the feasible set is reduced as the size of the matrices in (2.15)-(2.16) increases. These properties show that the optimal value of problem (2.13)-(2.16) increases with d and remains bounded by the optimal value of (2.10)-(2.11).

For the method to give better results, a ball constraint $\|\mathbf{x}\|^2 \leq M$ must be added according to the technical assumption 1.1 in [9]. For the OPF problem, this can be done easily by setting M to $\sum_{k \in \mathcal{N}} (v_k^{\max})^2$ without modifying the problem. The following two properties hold in this case [9, theorem 1.12]:

1. the optimal values of the hierarchy of semidefinite relaxations increasingly converge toward the optimal value of PolyOPF,
2. let \mathbf{y}^d denote a global solution to the relaxation of order d and $(\mathbf{e}^i)_{1 \leq i \leq 2n-1}$ denotes the canonical basis of \mathbb{N}^{2n-1} ; if PolyOPF has a unique global solution, then $(\mathbf{y}_{\mathbf{e}^i}^d)_{1 \leq i \leq 2n-1}$ converges towards the global solution to PolyOPF as d tends to $+\infty$.

The largest matrix size of the moment relaxation appears in (2.14) and has the value $|\mathbb{N}_d^{2n-1}| = \binom{2n-1+d}{d}$, where n is the number of buses. For a fixed d , matrix size is therefore equal to $O(n^d)$. This makes high order relaxations too large to compute with currently available SDP software packages. Consequently, the success of the moment-sos approach relies wholly upon its ability to find a global solution with a low order relaxation,

for which there is no guarantee. Note that the global solution is found by a finite order relaxation under conditions that include the convexity of the problem [63] (not the case of PolyOPF though) or the positive definiteness of the Hessian of the Lagrangian at the saddle points of the Lagrangian [39] (open question in the case of PolyOPF).

Moment-sos relaxations and rank relaxation

When the polynomials f_i defining PolyOPF are quadratic, the first order ($d = 1$) relaxation (2.13)-(2.16) is equivalent to Shor's relaxation [60]. To make the link with the rank relaxation of [66], consider now the case when the varying part of the f_i 's are quadratic and *homogeneous* like in [66], that is $f_i(\mathbf{x}) = \mathbf{x}^T A_i \mathbf{x} - a_i$ for all $i = 0, \dots, m$, with symmetric matrices A_i and scalars a_i . Then introducing the vector \mathbf{s} and the matrix Y defined by $\mathbf{s}_i = y_{\mathbf{e}^i}$ and $Y_{kl} = y_{\mathbf{e}^k + \mathbf{e}^l}$, reads

$$\min_{(\mathbf{s}, Y)} \text{trace}(A_0 Y) - a_0, \quad (2.17)$$

subject to

$$\begin{bmatrix} 1 & \mathbf{s}^T \\ \mathbf{s} & Y \end{bmatrix} \succcurlyeq 0 \quad \text{and} \quad \text{trace}(A_i Y) \geq a_i \quad (\forall i = 1, \dots, m). \quad (2.18)$$

Using Schur's complement, the positive semidefiniteness condition in (2.18) is equivalent to $Y - \mathbf{s}\mathbf{s}^T \succcurlyeq 0$. Since \mathbf{s} does not intervene elsewhere in (2.17)-(2.18), it can be eliminated and the constraints of the problem can be replaced by

$$Y \succcurlyeq 0 \quad \text{and} \quad \text{trace}(A_i Y) \geq a_i \quad (\forall i = 1, \dots, m). \quad (2.19)$$

The pair made of (2.17) and (2.19) is the rank relaxation of [66]. We have just shown that the equivalence between that the SDP relaxation of [66] and to the first-order moment relaxation holds when the varying part of the f_i 's are quadratic and homogeneous. For the OPF problem, this certainly occurs when

1. the objective of the OPF is an affine function of active power,
2. there are no constraints on apparent power flow,
3. (2.8) is not replaced by (2.9).

Point 1 ensures that the objective is quadratic and has a homogeneous varying part. Points 2 and 3 guarantee the same property for the constraint functions.

2.4 Numerical results

We present numerical results for the moment-sos approach applied to instances of the OPF for which the rank relaxation method of [66] fails to find the global solution. We focus on the

WB2 2-bus system, LMBM3 3-bus system, and the WB5 5-bus system that are described in [22]. Note that LMBM3 is also found in [67]. For each of the three systems, the authors of [22] modify a bound in the data and specify a range for which the rank relaxation fails. We consider 10 values uniformly distributed in the range in order to verify that the rank relaxation fails and to assess the moment-sos approach. We proceed in accordance with the discussion of section 2.3 by adding the redundant ball constraint. Surprisingly, the second order relaxation whose greatest matrix size is equal to $(2n+1)n$ nearly always finds the global solution.

The materials used are:

- Data of WB2, LMBM3, WB5 systems available online [23],
- Intel® Xeon™ MP CPU 2.70 GHz 7.00 Go RAM,
- MATLAB version 7.7 2008b,
- MATLAB-package MATPOWER version 3.2 [132],
- SeDuMi 1.02 [115] with tolerance parameter `pars.eps` set to 10^{-12} for all computations,
- MATLAB-based toolbox YALMIP [69] to compute Optimization 4 (Dual OPF) in [66] that yields the solution to the rank relaxation,
- MATLAB-package GloptiPoly version 3.6.1 [48] to compute solutions to a hierarchy of SDP relaxations (2.13)-(2.16).

The same precision is used as in the solutions of the test archives [23]. In other words, results are precise up to 10^{-2} p.u. for voltage phase, 10^{-2} degree for angles, 10^{-2} MW for active power, 10^{-2} MVA for reactive power, and cent per hour for costs. Computation time is several seconds.

GloptiPoly can guarantee that it has found a global solution to a polynomial optimization problem, up to a given precision. This is certainly the case when it finds a feasible point \mathbf{x} giving to the objective a value sufficiently close to the optimal value of the relaxation.

2-bus network: WB2

Authors of [22] observe that in the WB2 2-bus system of figure 2.1, the rank constraint is not satisfied in the rank relaxation method of [66] when $0.976 \text{ p.u.} < v_2^{\max} < 1.035 \text{ p.u.}$ In table 2.1, the first column is made up of 10 points in that range that are uniformly distributed. The second column contains the lowest order of the relaxations that yield a global solution. The optimal value of the relaxation of that order is written in the third column. The fourth column contains the optimal value of the rank relaxation (it is put between parentheses when the relaxation is inexact).

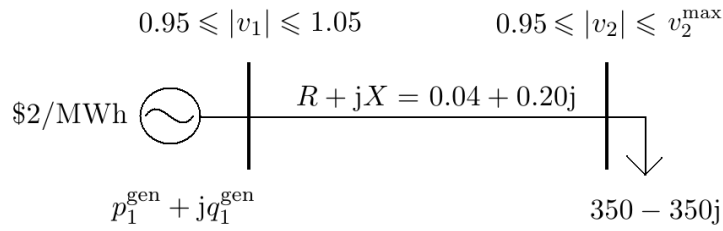


Figure 2.1: WB2 2-bus system

Table 2.1: Order of hierarchy needed to reach global solution to WB2 when rank relaxation fails

v_2^{\max} (p.u.)	relax. order	optimal value (\$/h)	rank relax. value (\$/h)
0.976	2	905.76	905.76
0.983	2	905.73	(903.12)
0.989	2	905.73	(900.84)
0.996	2	905.73	(898.17)
1.002	2	905.73	(895.86)
1.009	2	905.73	(893.16)
1.015	2	905.73	(890.82)
1.022	3	905.73	(888.08)
1.028	3	905.73	(885.71)
1.035	2	882.97	882.97

The hierarchy of SDP relaxations is defined for $d \geq 1$ because the objective is an affine function and there are no apparent flow constraints. Let's explain how it works in the case where $v_2^{\max} = 1.022$ p.u. The optimal value of the first order relaxation is 861.51 \$/h, that of the second order relaxation is 901.38 \$/h, and that of the third is 905.73 \$/h. This is coherent with point 1 of the discussion of section 2.3 that claims that the optimal values increase with d . Computing higher orders is not necessary because GloptiPoly numerically proves global optimality for the third order.

Notice that for $v_2^{\max} = 1.022$ p.u. the value of the rank relaxation found in table 2.1 (888.08 \$/h) is different from the value of the first order relaxation (861.51 \$/h). If we run GloptiPoly with (2.8) instead of (2.9), the optimal value of the first order relaxation is equal 888.08 \$/h as expected according to section 2.3.

For $v_2^{\max} = 0.976$ p.u. and $v_2^{\max} = 1.035$ p.u. (see the first and last rows of table 2.1), the rank constraint is satisfied in the rank relaxation method so its optimal value is equal to the one of the successful moment-sos method. In between those values, the rank constraint is not satisfied since the optimal value is less than the optimal value of the OPF. Notice the correlation between the results of table 2.1 and the upper half of figure 8 in [22]. Indeed, the

figure shows the optimal value of the OPF is constant whereas the optimal value of the rank relaxation decreases in a linear fashion when $0.976 \text{ p.u.} < v_2^{\max} < 1.035 \text{ p.u.}$

Surprisingly and encouragingly, according to the second column of table 2.1, the second order moment-sos relaxation finds the global solution in 8 out of 10 times, and the third order relaxation always find the global solution.

Remark: The fact that the rank constraint is not satisfied for the WB2 2-bus system of [22] seems in contradiction with the results of papers [19, 111, 130]. Indeed, the authors of the papers state that the rank is less than or equal to 1 if the graph of the network is acyclic and if load over-satisfaction is allowed. However, load over-satisfaction is not allowed in this network. For example, for $v_2^{\max} = 1.022 \text{ p.u.}$, adding 1 MW of load induces the optimal value to go down from 905.73 \$/h to 890.19 \$/h. One of the sufficient conditions in [18] for the rank is less than or equal to 1 relies on the existence of a strictly feasible point. It is not the case here because equality constraints must be enforced in the power balance equation.

3-bus network: LMBM3

We observe that in the LMBM3 3-bus system of figure 2.2, the rank constraint is not satisfied in the rank relaxation method of [66] when $28.35 \text{ MVA} \leq s_{23}^{\max} = s_{32}^{\max} < 53.60 \text{ MVA}$. Below 28.35 MVA, no solutions can be found by the OPF solver `runopf` in MATPOWER nor by the hierarchy of SDP relaxations. At 53.60 MVA, the rank constraint is satisfied in the rank relaxation method so its optimal value is equal to the optimal value of the OPF found by the second order relaxation; see to the last row of table 2.2.

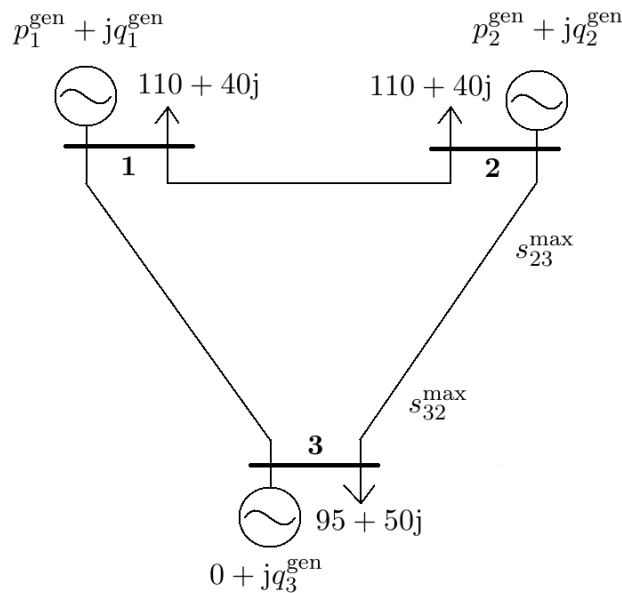


Figure 2.2: LMBM3 3-bus system

Table 2.2: Order of hierarchy needed to reach global solution to LMBM3 when rank relaxation fails

$s_{23}^{\max} = s_{32}^{\max}$ (MVA)	relax. order	optimal value (\$/h)	rank relax. value (\$/h)
28.35	2	10294.88	(6307.97)
31.16	2	8179.99	(6206.78)
33.96	2	7414.94	(6119.71)
36.77	2	6895.19	(6045.33)
39.57	2	6516.17	(5979.38)
42.38	2	6233.31	(5919.12)
45.18	2	6027.07	(5866.68)
47.99	2	5882.67	(5819.02)
50.79	2	5792.02	(5779.34)
53.60	2	5745.04	5745.04

The objective of the OPF is a quadratic function of active power so the hierarchy of SDP relaxations is defined for $d \geq 2$. Again, it is surprising that the second order moment-sos relaxation always finds the global solution to the LMBM3 system, as can be seen in the second column of table 2.2.

Authors of [66] make the assumption that the objective of the OPF is an increasing function of generated active power. The moment-sos approach does not require such an assumption. For example, when $s_{23}^{\max} = s_{32}^{\max} = 50$ MVA, active generation at bus 1 is equal to 148.07 MW and active generation at bus 2 is equal to 170.01 MW using the increasing cost function of [23,67]. Suppose we choose a different objective which aims at reducing deviation from a given active generation plan at each generator. Say that this plan is $p_1^{\text{plan}} = 170$ MW at bus 1 and $p_2^{\text{plan}} = 150$ MW at bus 2. The objective function is equal to $(p_1^{\text{gen}} - p_1^{\text{plan}})^2 + (p_2^{\text{gen}} - p_2^{\text{plan}})^2$. It is not an increasing function of p_1^{gen} and p_2^{gen} . The second order relaxation yields a global solution in which active generation at bus 1 is equal to 169.21 MW and active generation at bus 2 is equal to 149.19 MW.

5-bus network: WB5

Authors of [22] observe that in the WB5 5-bus system of figure 2.3, the rank constraint is not satisfied in the rank relaxation method of [66] when $q_5^{\min} > -30.80$ MVAR. Above 61.81 MVAR, no solutions can be found by the OPF solver `runopf` in MATPOWER. At -30.80 MVAR, the rank constraint is satisfied in the rank relaxation method so its optimal value is equal to the optimal value of the OPF found by the second order moment-sos relaxation; see the first row of table 2.3. As for the 9 values considered greater than -30.80 MVAR, the rank constraint is not satisfied since the optimal value is not equal to the optimal value of the OPF. Notice that the objective of the OPF is a linear function of active power and

there are bounds on apparent flow so the hierarchy of SDP relaxations is defined for $d \geq 1$.

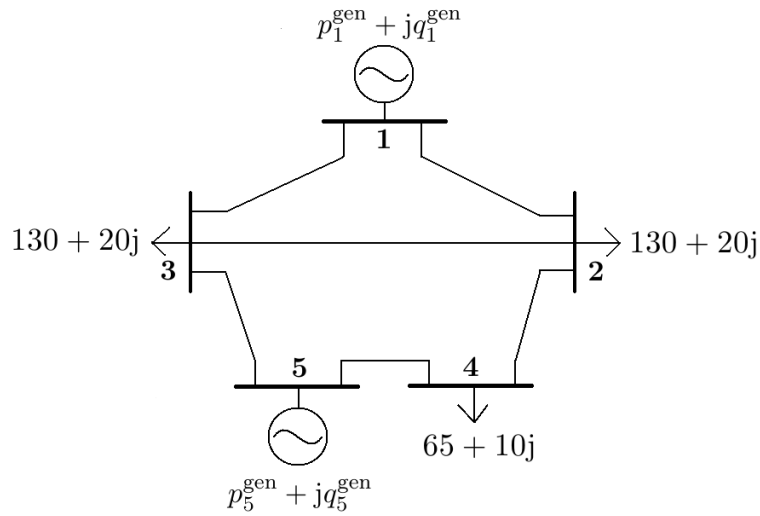


Figure 2.3: WB5 5-bus system

Table 2.3: Order of hierarchy needed to reach global solution to WB5 when rank relaxation fails

q_5^{\min} (MVAR)	relax. order	optimal value (\$/h)	rank relax. value (\$/h)
-30.80	2	945.83	945.83
-20.51	2	1146.48	(954.82)
-10.22	2	1209.11	(963.83)
00.07	2	1267.79	(972.85)
10.36	2	1323.86	(981.89)
20.65	2	1377.97	(990.95)
30.94	2	1430.54	(1005.13)
41.23	2	1481.81	(1033.07)
51.52	2	1531.97	(1070.39)
61.81	-	-	(1114.90)

When $q_5^{\min} = 61.81$ MVAR, the hierarchy of SDP relaxations is unable to find a feasible point, hence the empty slots in the last row of table 2.3. Apart from that value, the second order moment-sos relaxation again always finds the global solution according to the second column of table 2.3.

Waki *et al.* [125] have produced a piece of software called SparsePOP [127] similar to GloptiPoly only that it seeks to reduce problem size in Lasserre’s relaxations using matrix

completion theory in semidefinite programming. SparsePOP successfully solves the systems studied in this paper to global optimality but fails to reduce the size of the moment-sos relaxations and to solve problems with a larger number of buses.

2.5 Conclusion

This chapter examined the application of the moment-sos (sum of squares) approach to the global optimization of the optimal power flow (OPF) problem. The result is that the OPF can be successfully convexified in the case of several small networks where a previously known SDP method fails. The SDP problems considered in this paper can be viewed as extensions of the previously used rank relaxation. It is guaranteed to be more accurate than the previous one but requires more runtime.

Interestingly, Daniel K. Molzahn and Ian A. Hiskens independently made very similar findings [79] as presented in this chapter. They successfully solved networks with up to 10 buses. A group at IBM research Ireland was also working on the same ideas [44]. They managed to solve networks with up to 40 buses. To do so, they formulated the OPF problem as a quadratically-constrained quadratic problem and used SparsePOP. This works better than what we had tried, namely using SparsePOP with an OPF formulation with monomials of order 4.

We next focus on a property of the moment-sos approach to further prove its applicability in practice. In the small examples considered in this chapter, there is no duality gap at each order of the moment-sos hierarchy according to numerical results. This property is necessary for efficient solvers to work such as interior-point solvers. However, in the existing literature, there were no results guaranteeing this property. In the next chapter, we prove there is no duality gap in the moment-sos hierarchy in the presence of a ball constraint. We also explain why there is no duality gap when applying the moment-sos hierarchy to the OPF without a ball constraint.

Chapter 3

Zero duality gap in the Lasserre hierarchy

A polynomial optimization problem (POP) consists of minimizing a multivariate real polynomial on a semi-algebraic set K described by polynomial inequalities and equations. In its full generality it is a nonconvex, multi-extremal, difficult global optimization problem. More than an decade ago, J. B. Lasserre proposed to solve POPs by a hierarchy of convex semidefinite optimization (SDP) relaxations of increasing size. Each problem in the hierarchy has a primal SDP formulation (a relaxation of a moment problem) and a dual SDP formulation (a sum-of-squares representation of a polynomial Lagrangian of the POP). In this chapter, we show that there is no duality gap between each primal and dual SDP problem in Lasserre's hierarchy, provided one of the constraints in the description of set K is a ball constraint. Our proof uses elementary results on SDP duality, and it does not assume that K has a strictly feasible point. The material in this chapter is based on the publication:

C. JOSZ AND D. HENRION, *Strong Duality in Lasserre's Hierarchy for Polynomial Optimization*, Optimization Letters, February 2015. [doi] [preprint]

3.1 Introduction

Consider the following polynomial optimization problem (POP)

$$\begin{aligned} \inf_x \quad & f(x) := \sum_{\alpha} f_{\alpha} x^{\alpha} \\ \text{s.t.} \quad & g_i(x) := \sum_{\alpha} g_{i,\alpha} x^{\alpha} \geq 0, \quad i = 1, \dots, m \end{aligned} \tag{3.1}$$

where we use the multi-index notation $x^{\alpha} := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for $x \in \mathbb{R}^n$, $\alpha \in \mathbb{N}^n$ and where the data are polynomials $f, g_1, \dots, g_m \in \mathbb{R}[x]$ so that in the above sums only a finite number of coefficients f_{α} and $g_{i,\alpha}$ are nonzero. Let K denote its feasible set:

$$K := \{x \in \mathbb{R}^n : g_i(x) \geq 0, i = 1, \dots, m\}$$

To solve POP (3.1), Lasserre [61, 62] proposed a semidefinite optimization (SDP) relaxation hierarchy with guaranteed asymptotic global convergence provided an algebraic assumption holds:

Assumption 3.1. There exists a polynomial $u \in \mathbb{R}[x]$ such that $\{x \in \mathbb{R}^n : u(x) \geq 0\}$ is bounded and $u = u_0 + \sum_{i=1}^m u_i g_i$ where polynomials $u_i \in \mathbb{R}[x]$, $i = 0, 1, \dots, m$ are sums of squares (SOS) of other polynomials.

Nie *et al.* [87] have proven that Assumption 3.1 also implies generically finite convergence, that is to say that for almost every instance of POP, there exists a finite-dimensional SDP relaxation in the hierarchy whose optimal value is equal to the optimal value of the POP. Assumption 3.1 can be difficult to check computationally (as the degrees of the SOS multipliers can be arbitrarily large), and it is often replaced by the following slightly stronger assumption:

Assumption 3.2. The description of K contains a ball constraint, say $g_m(x) = R^2 - \sum_{i=1}^n x_i^2$ for some real number R .

Indeed, under Assumption 3.2, simply choose $u = g_m$, $u_1 = \dots = u_{m-1} = 0$, and $u_m = 1$ to conclude that Assumption 3.1 holds as well. In practice, it is often easy to see to it that Assumption 3.2 holds. In the case of a POP with a bounded feasible set, a redundant ball constraint can be added.

More generally, if the intersection of the sublevel set $\{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ with the feasible set of the POP is bounded for some feasible point x_0 , then a redundant ball constraint can also be added. As an illustration, a reviewer suggested the example of the minimization of $f(x) = x_1^2 + x_2^2 - 3x_1x_2$ on the unbounded set defined on \mathbb{R}^2 by the constraint $g_1(x) = 1 - 3x_1x_2 \geq 0$. The intersection of the feasible set with the set defined by the constraint $f(x) \leq f(0) = 0$ is included in the ball defined by $g_2(x) = 1 - x_1^2 - x_2^2 \geq 0$ so that the POP can be equivalently defined on the bounded set $K = \{x \in \mathbb{R}^2 : g_1(x) \geq 0, g_2(x) \geq 0\}$.

Each problem in Lasserre's hierarchy consists of a primal-dual SDP pair, called SDP relaxation, where the primal corresponds to a convex moment relaxation of the original (typically nonconvex) POP, and the dual corresponds to a SOS representation of a polynomial Lagrangian of the POP. The question arises of whether the duality gap vanishes in each SDP relaxation. This is of practical importance because numerical algorithms to solve SDP problems are guaranteed to converge only where there is a zero duality gap, and sometimes under the stronger assumption that there is a primal or/and dual SDP interior point.

In [104, Example 4.9], Schweighofer provides a two-dimensional POP with no interior point for which Assumption 3.1 holds, yet a duality gap exists at the first SDP relaxation: $\inf x_1x_2$ s.t. $x \in K = \{x \in \mathbb{R}^2 : -1 \leq x_1 \leq 1, x_2^2 \leq 0\}$, with primal SDP value equal to zero and dual SDP value equal to minus infinity. This shows that a stronger assumption is required to ensure a zero SDP duality gap. A sufficient condition for strong duality has been given in [62]: set K should contain an interior point. However, this may be too restrictive: in the proof of Lemma 1 in [47] the authors use notationally awkward arguments involving truncated moment matrices to prove the absence of SDP duality gap for a certain set K that

contains no interior point. This shows that the existence of an interior point is not necessary for a zero SDP duality gap. More generally, it is not possible to assume the existence of an interior point for POPs with explicit equality constraints, and a weaker assumption for zero SDP duality gap is welcome.

Motivated by these observations, in this note we prove that under the basic Assumption 3.2 on the description of set K , there is no duality gap in the SDP hierarchy. Our interpretation of this result, and the main message of this contribution, is that in the context of Lasserre's hierarchy for POP, a practically relevant description of a bounded semialgebraic feasibility set must include a redundant ball constraint.

3.2 Proof

For notational convenience, let $g_0(x) = 1 \in \mathbb{R}[x]$ denote the unit polynomial. Define the localizing matrix

$$M_{d-d_i}(g_i y) := \left(\sum_{\gamma} g_{i,\gamma} y_{\alpha+\beta+\gamma} \right)_{|\alpha|, |\beta| \leq d-d_i} = \sum_{|\alpha| \leq 2d} A_{i,\alpha} y_{\alpha}$$

where d_i is the smallest integer greater than or equal to half the degree of g_i , for $i = 0, 1, \dots, m$, and $|\alpha| = \sum_{i=1}^n \alpha_i$. The Lasserre hierarchy for POP (3.1) consists of a primal moment SDP problem

$$(P_d) : \begin{array}{l} \inf_y \quad \sum_{\alpha} f_{\alpha} y_{\alpha} \\ \text{s.t.} \quad y_0 = 1 \\ \quad \quad M_{d-d_i}(g_i y) \succeq 0, \quad i = 0, 1, \dots, m \end{array}$$

and a dual SOS SDP problem

$$(D_d) : \begin{array}{l} \sup_{z, Z} \quad z \\ \text{s.t.} \quad f_0 - z = \sum_{i=0}^m \langle A_{i,0}, Z_i \rangle \\ \quad \quad f_{\alpha} = \sum_{i=0}^m \langle A_{i,\alpha}, Z_i \rangle, \quad 0 < |\alpha| \leq 2d \\ \quad \quad Z_i \succeq 0, \quad i = 0, 1, \dots, m, \quad z \in \mathbb{R} \end{array}$$

where $A \succeq 0$ stands for matrix A positive semidefinite, $\langle A, B \rangle = \text{trace } AB$ is the inner product between two matrices. The Lasserre hierarchy is indexed by an integer $d \geq d_{\min} := \max_{i=0,1,\dots,m} d_i$. The primal-dual pair (P_d, D_d) is called the SDP relaxation of order d for POP (3.1). The size of the primal variable $(y_{\alpha})_{|\alpha| \leq 2d}$ is $\binom{n+2d}{n}$ and the size of the dual

variable Z_i is $\binom{n+d-d_i}{n}$.

Let us define the following sets:

- \mathcal{P}_d : feasible points for P_d ;
- \mathcal{D}_d : feasible points for D_d ;
- $\text{int } \mathcal{P}_d$: strictly feasible points for P_d ;
- $\text{int } \mathcal{D}_d$: strictly feasible points for D_d ;
- \mathcal{P}_d^* : optimal solutions for P_d ;
- \mathcal{D}_d^* : optimal solutions for D_d .

Finally, let us denote by $\text{val } P_d$ the infimum in problem P_d and by $\text{val } D_d$ the supremum in problem D_d .

Lemma 3.3. *$\text{int } \mathcal{P}_d$ nonempty or $\text{int } \mathcal{D}_d$ nonempty implies $\text{val } P_d = \text{val } D_d$.*

Lemma 3.3 is classical in convex optimization, and it is generally called Slater's condition, see e.g. [105, Theorem 4.1.3].

Lemma 3.4. *The two following statements are equivalent :*

1. \mathcal{P}_d is nonempty and $\text{int } \mathcal{D}_d$ is nonempty;
2. \mathcal{P}_d^* is nonempty and bounded.

A proof of Lemma 3.4 can be found in [122]. According to Lemmas 3.3 and 3.4, \mathcal{P}_d^* nonempty and bounded implies strong duality. This result is also mentioned without proof at the end of [105, Section 4.1.2].

Lemma 3.5. *Under Assumption 3.2, set \mathcal{P}_d is included in the Euclidean ball of radius*

$$\sqrt{\binom{n+d}{n}} \sum_{k=0}^d R^{2k} \text{ centered at the origin.}$$

Proof. If $\mathcal{P}_d = \emptyset$, the result is obvious. If not, consider a feasible point $(y_\alpha)_{|\alpha| \leq 2d} \in \mathcal{P}_d$. Let $k \in \{1, \dots, d\}$. In the SDP problem P_k , the localizing matrix associated to the ball constraint $g_m(x) = R^2 - \sum_{i=1}^n x_i^2 \geq 0$ reads

$$M_{k-1}(g_m y) = \left(\sum_{\gamma} g_{m,\gamma} y_{\alpha+\beta+\gamma} \right)_{|\alpha|, |\beta| \leq k-1}$$

with trace equal to

$$\begin{aligned}
\text{trace } M_{k-1}(g_m y) &= \sum_{|\alpha| \leq k-1} \sum_{\gamma} g_{m,\gamma} y_{2\alpha+\gamma} \\
&= \sum_{|\alpha| \leq k-1} \left(g_{m,0} y_{2\alpha} + \sum_{|\gamma|=1} g_{m,2\gamma} y_{2\alpha+2\gamma} \right) \\
&= \sum_{|\alpha| \leq k-1} \left(R^2 y_{2\alpha} - \sum_{|\gamma|=1} y_{2(\alpha+\gamma)} \right) \\
&= \sum_{|\alpha| \leq k-1} R^2 y_{2\alpha} - \sum_{|\alpha| \leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)} \\
&\leq^1 R^2 (\sum_{|\alpha| \leq k-1} y_{2\alpha}) + y_0 - \sum_{|\alpha| \leq k} y_{2\alpha} \\
&\leq R^2 \text{trace } M_{k-1}(y) + 1 - \text{trace } M_k(y).
\end{aligned}$$

From the structure of the localizing matrix, it holds $M_{k-1}(g_m y) \succeq 0$ hence $\text{trace } M_{k-1}(g_m y) \geq 0$ and

$$\text{trace } M_k(y) \leq 1 + R^2 \text{trace } M_{k-1}(y)$$

from which we derive

$$\text{trace } M_d(y) \leq \sum_{k=1}^d R^{2(k-1)} + R^{2d} \text{trace } M_0(y) = \sum_{k=0}^d R^{2k}$$

since $\text{trace } M_0(y) = y_0 = 1$. The operator norm $\|M_d(y)\|$, equal to the maximum eigenvalue of $M_d(y)$, is upper bounded by $\text{trace } M_d(y)$, the sum of the eigenvalues of $M_d(y)$, which are all nonnegative. Moreover the Frobenius norm satisfies

$$\begin{aligned}
\|M_d(y)\|_F^2 &:= \langle M_d(y), M_d(y) \rangle \\
&= \langle \sum_{|\delta| \leq 2d} A_{0,\delta} y_\delta, \sum_{|\delta| \leq 2d} A_{0,\delta} y_\delta \rangle \\
&= \sum_{|\delta| \leq 2d} \langle A_{0,\delta}, A_{0,\delta} \rangle y_\delta^2 \quad \text{by orthogonality of matrices } (A_{0,\delta})_{|\delta| \leq 2d} \\
&\geq \sum_{|\delta| \leq 2d} y_\delta^2 \quad \text{because } \langle A_{0,\delta}, A_{0,\delta} \rangle \geq 1
\end{aligned}$$

¹In the associated publication [56], there is an equality instead of an inequality, which is an error. Indeed, $\sum_{|\alpha| \leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)} = \sum_{0 < |\alpha| \leq k} y_{2\alpha}$ is false whereas $\sum_{|\alpha| \leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)} \geq \sum_{0 < |\alpha| \leq k} y_{2\alpha}$ is true. The reason for this is that each term $y_{2\alpha}$ with $0 < |\alpha| \leq k$ appears at least once in $\sum_{|\alpha| \leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)}$, but can potentially appear more than once. More precisely, we have $\sum_{|\alpha| \leq k-1, |\gamma|=1} y_{2(\alpha+\gamma)} = \sum_{|\alpha| \leq k} |\alpha| y_{2\alpha} \geq \sum_{0 < |\alpha| \leq k} y_{2\alpha}$. This error thankfully has no impact on the rest of the proof other than the inequality right below it.

where matrices $(A_{0,\delta})_{|\delta|\leq 2d}$ can be written using column vectors $(e_\alpha)_{|\alpha|\leq d}$, containing only zeros apart from the value 1 at index α , via the explicit formula

$$A_{0,\delta} = \sum_{\substack{\alpha + \beta = \delta \\ |\alpha|, |\beta| \leq d}} e_\alpha e_\beta^T.$$

The proof follows then from

$$\sqrt{\sum_{|\delta|\leq 2d} y_\delta^2} \leq \|M_d(y)\|_F \leq \sqrt{\binom{n+d}{n}} \|M_d(y)\| \leq \sqrt{\binom{n+d}{n}} \sum_{k=0}^d R^{2k}.$$

□

Theorem 3.6. *Assumption 3.2 implies that $-\infty < \text{val } P_d = \text{val } D_d$ for all $d \geq d_{\min}$.*

Proof. Let $d \geq d_{\min}$. Firstly, let us consider the case when \mathcal{P}_d is nonempty. According to Lemma 3.5, \mathcal{P}_d is bounded and closed, and the objective function in P_d is linear, so we conclude that \mathcal{P}_d^* is nonempty and bounded. According to Lemma 3.4, $\text{int } \mathcal{D}_d$ is nonempty, and from Lemma 3.3, $\text{val } P_d = \text{val } D_d$.

Secondly, let us consider the case when \mathcal{P}_d is empty. An infeasible SDP problem can be either weakly infeasible or strongly infeasible, see [40, Section 5.2] for definitions. Let us prove by contradiction that P_d cannot be weakly infeasible. If P_d is weakly infeasible, there must exist a sequence $(y^p)_{p \in \mathbb{N}}$ such that

$$\forall p \in \mathbb{N}, \quad \begin{cases} 1 - \frac{1}{p+1} \leq y_0^p \leq 1 + \frac{1}{p+1} \\ \lambda_{\min}(M_{d-d_i}(g_i y^p)) \geq -\frac{1}{p+1}, \quad i = 0, 1, \dots, m \end{cases}$$

where λ_{\min} denotes the minimum eigenvalue of a symmetric matrix. According to the proof of Lemma 3.5, for all $1 \leq k \leq d$ and all real numbers $(y_\alpha)_{|\alpha|\leq 2d}$, one has

$$\text{trace } M_{k-1}(g_m y) = R^2 \text{trace } M_{k-1}(y) + y_0 - \text{trace } M_k(y).$$

Clearly, $\text{trace } M_{k-1}(g_m y) \geq -\frac{c}{1+p}$ where $c := \binom{n+d}{n}$ denotes the size of the moment matrix $M_d(y)$. The following holds

$$\text{trace } M_k(y^p) \leq R^2 \text{trace } M_{k-1}(y^p) + 1 + \frac{1+c}{1+p}$$

from which we derive

$$\text{trace } M_d(y^p) \leq \left(1 + \frac{1+c}{1+p}\right) \sum_{k=0}^d R^{2k}.$$

Together with $\lambda_{\min}(M_d(y^p)) \geq -\frac{1}{1+p}$, this yields

$$\lambda_{\max}(M_d(y^p)) \leq \left(1 + \frac{1+c}{1+p}\right) \sum_{k=0}^d R^{2k} + \frac{c-1}{1+p}$$

where λ_{\max} denotes the maximum eigenvalue of a symmetric matrix. Hence for all $p \in \mathbb{N}$, the spectrum of the moment matrix $M_d(y^p)$ is lower bounded by $l := -1$ and upper bounded by $u := (2+c) \sum_{k=0}^d R^{2k} + c - 1$. Therefore:

$$\sqrt{\sum_{|\delta| \leq 2d} (y_\delta^p)^2} \leq \|M_d(y^p)\|_F \leq \sqrt{c} \max(|l|, |u|)$$

The sequence $(y^p)_{p \in \mathbb{N}}$ is hence included in a compact set. Thus there exists a subsequence which converges towards y^{lim} such that $y_0^{\text{lim}} = 1$ and $\lambda_{\min}(M_{d-d_i}(g_i y^{\text{lim}})) \geq 0$, $i = 0, 1, \dots, m$. The limit y^{lim} is thus included in \mathcal{P}_d , which is a contradiction.

SDP problem P_d is strongly infeasible which means that its dual problem D_d has an improving ray [40, Definition 5.2.2]. To conclude that $\text{val } D_d = +\infty$, all that is left to prove is that $\mathcal{D}_d \neq \emptyset$. Consider the primal problem P_d discarding all constraints but $y_0 = 1$, $M_d(y) \succeq 0$, and $M_{d-1}(g_m y) \succeq 0$. It is a feasible and bounded SDP problem owing to Lemma 3.5. According to Lemma 3.4, its dual problem must contain a feasible point (z, Z_0, Z_m) and hence $(z, Z_0, 0, \dots, 0, Z_m) \in \mathcal{D}_d$. □

3.3 Conclusion

We prove that there is no duality gap in Lasserre's SDP hierarchy for POPs whose description of the feasible set contains a ball constraint. Prior results ensuring zero duality gap required the existence of a strictly feasible point, which excludes POPs with equality constraints. A zero duality gap is an important property for interior-point solvers to successfully find solutions. A slight adaption of the proof we propose shows that in the case of the optimal power flow problem, upper bounds on voltage imply that there is no duality gap at each order of the Lasserre hierarchy. The adaption consists of summing the traces of each localizing matrix associated to an upper voltage constraint. The sum is equal to the trace of the localizing matrix of a sphere constraint that would be obtained by adding all upper voltage constraints. This means that the computation in Lemma 3.5 is still valid and that the overall proof still holds.

In Chapter 2, we've proven the applicability of Lasserre's hierarchy from numerical perspective. In this chapter, its applicability was enforced from a theoretical perspective. To prove its applicability to large-scale networks, new test cases needed to be made publicly

available. Indeed, the only large-scale networks available so far were Polish networks, each corresponding to a different period in the year, and the Great Britain network. These networks contain roughly two to three thousand buses each. In the next chapter, we present new data of European networks from various countries and with up to nine thousand buses. Figure 3.1 gives a brief history of the computations prior to the introduction of these networks. Only computations leading to physically meaningful results are presented, in other words those that lead to feasible solutions. These are either global solutions or nearly global solutions (in the case of penalization). Working in collaboration with Daniel K. Molzahn and Ian Hiskens, we were to find global solutions in the case of active power minimization for well-conditioned networks with up to two thousand variables (Chapters 5 and 7).

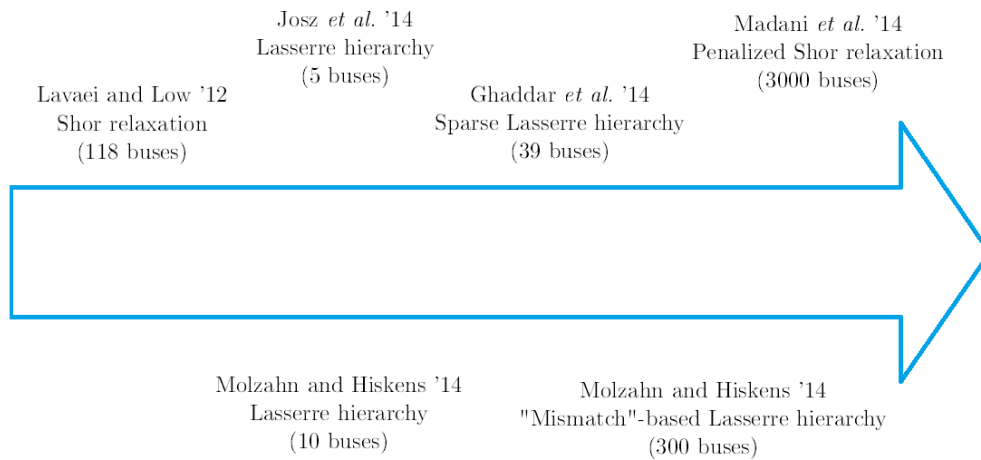


Figure 3.1: Timeline of computational advances [44, 57, 66, 72, 79, 81]

Chapter 4

Data of European transmission network

We contributed four new test cases [55] to MATPOWER [132], a package of MATLAB. They represent parts and all of the European high voltage AC transmission network. The data stems from the Pan European Grid Advanced Simulation and State Estimation (PEGASE) project, part of the 7th Framework Program of the European Union (<http://www.fp7-pegase.com/>). Its goal was to develop new tools for the real-time control and operational planning of the pan-Europeean transmission network. Specifically, new approaches were implemented for state estimation, dynamic security analysis, steady state optimization. A dispatcher training simulator was also created. The associated public data we provided is:

C. Jozs, S. Fliscounakis, J. Maeght, and P. Panciatici, *Power Flow Data for European High-Voltage Transmission Network: 89, 1354, 2869, and 9241-bus PEGASE Systems*, MATPOWER 5.1, March 2015. [link]

PEGASE data contains asymmetric shunt conductance and susceptance in the PI transmission line model of branches. However, MATPOWER format do not allow for asymmetry. As a result, we set the total line charging susceptance of branches to 0 per unit in the MATPOWER files. We used the nodal representation of shunt conductance and susceptance. This procedure leaves the power flow equations unchanged compared with the original PEGASE data. However, line flow constraints in the optimal power flow problem are modified.

The data includes negative resistances and negative lower bounds on active power generation. This is due to sections of the network that are not represented. These sections may include generators, which accounts for the negative resistances. Imports and exports with these sections account for negative bounds on generation. These sections may be a country connected to the European network that is represented by the data. It may also be a section within a country for which data was not provided. A big challenge when making realistic test cases is to account for missing data in a sensible manner. See [6,86] for work on the subject. Note that the non-represented sections also account for buses where voltage magnitudes have

very large lower and upper bounds. All others buses have tight constraints, plus or minus 10% of the nominal value. This is visible in figure 4.2. It represents the voltages at each bus along with the lower and upper constraints resulting in an annulus.

4.1 case89pegase

This case accurately represents the size and complexity of a small part of the European high voltage transmission network. The network contains 89 buses, 12 generators, and 210 branches. It operates at 380, 220, and 150 kV.

4.2 case1354pegase

This case accurately represents the size and complexity of a medium part of the European high voltage transmission network. The network contains 1,354 buses, 260 generators, and 1,991 branches. It operates at 380, and 220 kV.

4.3 case2869pegase

This case accurately represents the size and complexity of a large part of the European high voltage transmission network. The network contains 2,869 buses, 510 generators, and 4,582 branches. It operates at 380, 220, 150, and 110 kV.

4.4 case9241pegase

This case accurately represents the size and complexity of the European high voltage transmission network. The network contains 9,241 buses, 1,445 generators, and 16,049 branches. It operates at 750, 400, 380, 330, 220, 154, 150, 120 and 110 kV. It represents 23 countries as can be seen in figure 4.1. The numbers between the countries correspond to the sum of the active power flows traded at the interconnections for the voltage profile of figure 4.2. This voltage profile was found with the nonlinear solver KNITRO. It is contained in the case9241 MATPOWER file.

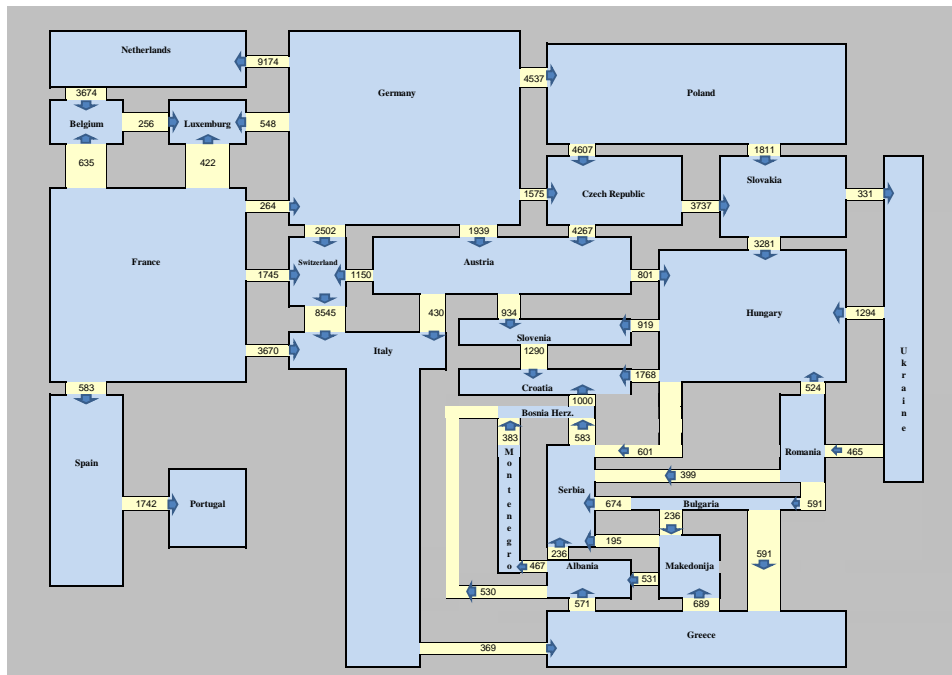


Figure 4.1: PEGASE network with active power flows in MW (courtesy of Stéphane Fliscounakis)

Voltages in Complex Plane: Local Optimum of 9241-bus European Network

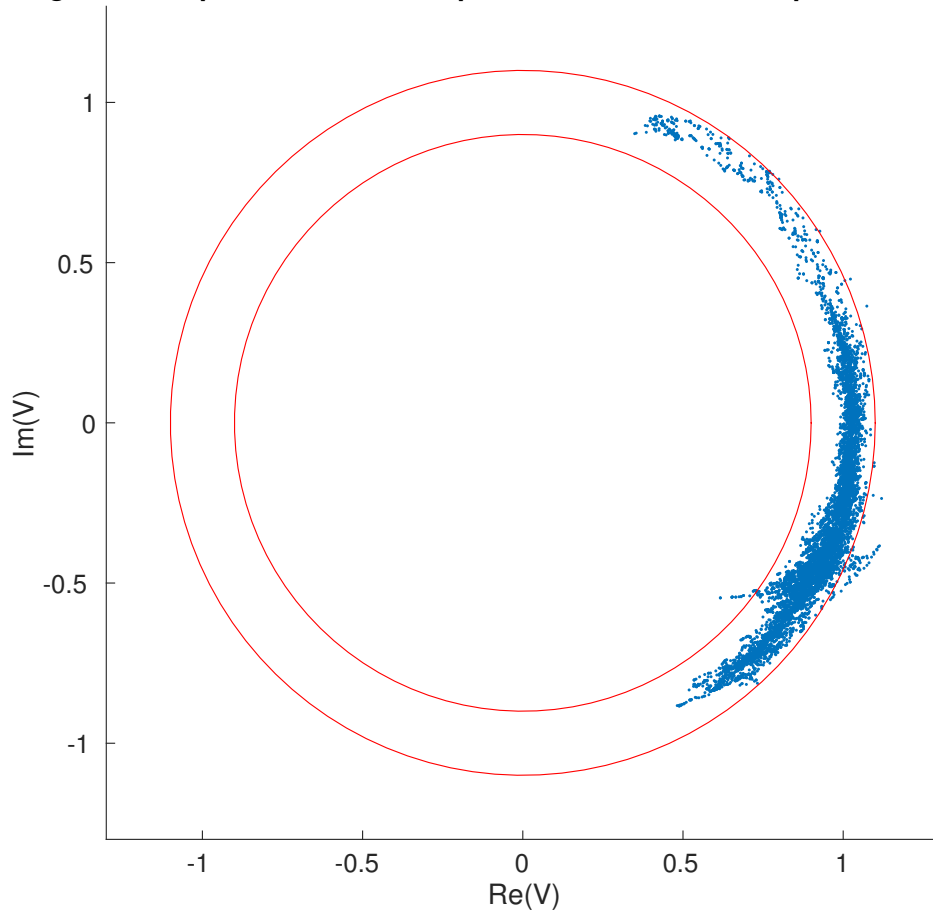


Figure 4.2: Local optimum

The PEGASE data provides large-scale test cases on which to test new methods and tools. After having shown the applicability of moment/sum-of-squares approach to small-scale systems, the main challenge was to make the approach tractable for large-scale systems. A significant turn took place when Daniel K. Molzahn and Ian A. Hiskens proposed a novel way to exploit sparsity when applying the moment-sos hierarchy to the OPF problem [81]. The main idea is that a global solution can be obtained by only applying high-order moments to some constraints. These constraints are deduced by looking at the moment matrix when the relaxation fails. The authors of [81] were thus able to solve networks to global optimality with up to three hundred buses. In the next chapter, their method is combined with a penalization approach due to Madani *et al* [72]. Tests are made with the PEGASE data in the next three chapters.

Chapter 5

Penalized Lasserre hierarchy

Applications of convex relaxation techniques to the nonconvex OPF problem have been of recent interest, including work using the Lasserre hierarchy of “moment” relaxations to globally solve many OPF problems. By preprocessing the network model to eliminate low-impedance lines, this chapter demonstrates the capability of the moment relaxations to globally solve large OPF problems that minimize active power losses for portions of several European power systems. Large problems with more general objective functions have thus far been computationally intractable for current formulations of the moment relaxations. To overcome this limitation, this chapter proposes the combination of an objective function penalization with the moment relaxations. This combination yields feasible points with objective function values that are close to the global optimum of several large OPF problems. Compared to an existing penalization method, the combination of penalization and the moment relaxations eliminates the need to specify one of the penalty parameters and solves a broader class of problems. The material presented in this chapter is based on the publication:

D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *Solution of Optimal Power Flow Problems using Moment Relaxations Augmented with Objective Function Penalization*, 54th Annual Conference on Decision and Control, Osaka, December 2015. [preprint]

5.1 Introduction

The SDP relaxation of [66] has been generalized to a family of “moment relaxations” using the Lasserre hierarchy [64] for polynomial optimization [44, 57, 79]. The moment relaxations take the form of SDPs, and the first-order relaxation in this hierarchy is equivalent to the SDP relaxation of [66]. Increasing the relaxation order in this hierarchy enables global solution of a broader class of OPF problems.

The ability to globally solve a broader class of OPF problems has a computational cost; the moment relaxations quickly become intractable with increasing order. Fortunately, second-

and third-order moment relaxations globally solve many small problems for which the first-order relaxation fails to yield the globally optimal decision variables.

However, increasing system size results in computational challenges even for low-order moment relaxations. The second-order relaxation is computationally intractable for OPF problems with more than ten buses. Exploiting network sparsity enables solution of the first-order relaxation for systems with thousands of buses [52, 82] and the second-order relaxation for systems with about forty buses [44, 81]. Recent work [81] solves larger problems (up to 300 buses) by both exploiting sparsity and only applying the computationally intensive higher-order moment relaxations to specific buses in the network. Other recent work improves tractability using a second-order cone programming (SOCP) relaxation of the higher-order moment constraints [80].

Solving larger problems using moment relaxations is often limited by numerical convergence issues rather than computational performance. We present a preprocessing method that improves numerical convergence by removing low-impedance lines from the network model. Similar methods are commonly employed (e.g., PSS/E [107]), but more extensive modifications are needed for adequate convergence due to the limited numerical capabilities of current SDP solvers.

After this preprocessing, the moment relaxations globally solve several large OPF problems which minimize active power losses for European power systems. Directly using the moment relaxations to globally solve more general large OPF problems with objective functions that minimize generation cost has been computationally intractable thus far.

To solve these OPF problems, we form moment relaxations using a penalized objective function. Previous literature [72, 74] augments the SDP relaxation [66] with penalization terms for the total reactive power generation and the apparent power loss of certain lines. For many problems, this penalization finds feasible points with objective function values that are very close to the lower bounds obtained from the SDP relaxation. Related work [83] uses a Laplacian-based objective function with a constraint on generation cost to find feasible points are very near the global optima. This section analyzes the physical and convexity properties of the reactive power penalization.

There are several disadvantages of the penalization method of [72]. This penalization often requires choosing multiple parameters. (See [83] for a related approach that does not require choosing penalty parameters.) Also, there are OPF problems that are globally solved by the moment relaxations, but no known penalty parameters yield a feasible solution.

We propose a “moment+penalization” approach that augments the moment relaxations with a reactive power penalty. Typical penalized OPF problems only require higher-order moment constraints at a few buses. Thus, for a variety of large test cases, augmenting the moment relaxation with the proposed single-parameter penalization achieves feasible solutions that are at least very near the global optima (within at least 1% for a variety of example problems). The moment+penalization approach enables solution of a broader class of problems than either method individually.

Below, Section 5.2 describes the low-impedance line preprocessing. Section 5.3 discusses the existing penalization and the proposed moment+penalization approaches. Section 5.6 demonstrates the moment+penalization approach using several large test cases, and Section 5.7 concludes.

5.2 Preprocessing low-impedance lines

By exploiting sparsity and selectively applying the higher-order constraints, the moment relaxations globally solve many OPF problems with up to 300 buses. Solution of larger problems with higher-order relaxations is typically limited by numerical convergence issues rather than computational concerns. This section describes a preprocessing method for improving numerical properties of the moment relaxations.

Low-impedance lines, which often represent “jumpers” between buses in the same physical location, cause numerical problems for many algorithms. Low line impedances result in a large range of values in the bus admittance matrix \mathbf{Y} , which causes numerical problems in the constraint equations.

To address these numerical problems, many software packages remove lines with impedances below a threshold. For instance, lines with impedance below a parameter `thrshz` are removed prior to applying other algorithms in PSS/E [107].

We use a slightly modified version of the low-impedance line removal procedure in PSS/E [107].¹ Low-impedance lines are eliminated by grouping buses that are connected by lines with impedances below a specified threshold `thrshz`. Each group of buses is replaced by one bus that is connected to all lines terminating outside the group. Generators, loads, and shunts (including the shunt susceptances of lines connecting buses within a group) are aggregated. The series parameters of lines connecting buses within a group are eliminated.

Removing low-impedance lines typically has a small impact on the solution. To recover an approximate solution to the original power system model, assign identical voltage phasors to all buses in each group and distribute flows on lines connecting buses within a group under the approximation that all power flows through the low-impedance lines.

A typical low-impedance line threshold `thrshz` is 1×10^{-4} per unit. However, the numerical capabilities of SDP solvers are not as mature as other optimization tools. Therefore, we require a larger `thrshz` = 1×10^{-3} per unit to obtain adequate convergence of the moment relaxations. This larger threshold typically introduces only small errors in the results, although non-negligible errors are possible.

MATPOWER solutions obtained for the Polish [132] and most PEGASE systems [42, 55] were the the same before and after low-impedance line preprocessing to within 0.0095 per unit voltage magnitude and 0.67° voltage angle difference across each line. Operating costs for all test problems were the same to within 0.04%. The 2869-bus PEGASE system had

¹Lines with non-zero resistances are not considered to be “low impedance” by PSS/E. We consider both the resistance and the reactance.

larger differences: 0.0287 per unit voltage magnitude and 1.37° angle difference. A power flow solution for the full network using the solution to the OPF problem after low-impedance line preprocessing yields smaller differences: 0.0059 per unit voltage magnitude and 1.17° angle difference. Thus, the differences from the preprocessing for the 2869-bus PEGASE system can be largely attributed to the sensitivity of the OPF problem itself to small changes in the low-impedance line parameters. Preprocessing reduced the number of buses by between 21% and 27% for the PEGASE and between 9% and 26% for the Polish systems.

These results show the need for further study of the sensitivity of OPF problems to low-impedance line parameters as well as additional numerical improvements of the moment relaxations and SDP solvers to reduce `thrshz`.

5.3 Moment relaxations and penalization

As will be shown in Section 6.4, the moment relaxations globally solve many large OPF problems with active power loss minimization objectives after removing low-impedance lines as described in Section 5.2. Directly applying the moment relaxations to many large OPF problems with more general cost functions has so far been computationally intractable. This section describes the nonconvexity associated with more general cost functions and proposes a method to obtain feasible solutions that are at least near the global optimum, if not, in fact, globally optimal for many problems.

Specifically, we propose augmenting the moment relaxations with a penalization in the objective function. Previous literature [72, 74] adds terms to the first-order moment relaxation that penalize the total reactive power injection and the apparent power line loss (i.e., $\sqrt{(f_{Plm} + f_{Pml})^2 + (f_{Qlm} + f_{Qml})^2}$) for “problematic” lines identified by a heuristic. This penalization often finds feasible points that are at least nearly globally optimal.

However, the penalization in [72] requires choosing two penalty parameters and fails to yield a feasible solution to some problems. This section describes progress in addressing these limitations by augmenting the moment relaxations with a reactive power penalization. The proposed “moment+penalization” approach only requires a single penalty parameter and finds feasible points that are at least nearly globally optimal for a broader class of OPF problems. This section also analyzes the convexity properties and provides a physical intuition for reactive power penalization.

5.4 Penalization of reactive power generation

The penalization method proposed in [72] perturbs the objective function to include terms that minimize the total reactive power loss and the apparent power loss on specific lines determined by a heuristic method. These terms enter the objective function with two scalar

parameter coefficients. Obtaining a feasible point near the global solution requires appropriate choice of these parameters.

For typical operating conditions, reactive power is strongly associated with voltage magnitude. Penalizing reactive power injections tends to reduce voltage magnitudes, which also tends to increase active power losses since a larger current flow, with higher associated loss, is required to deliver a given quantity of power at a lower voltage magnitude.

For many problems for which the first-order moment relaxation fails to yield the global optimum, the relaxation “artificially” increases the voltage magnitudes to reduce active power losses. This results in voltage magnitudes and power injections that are feasible for the SDP relaxation but infeasible for the OPF problem.

By choosing a reactive power penalty parameter that balances these competing tendencies (increasing voltage magnitudes to reduce active power losses vs. decreasing voltage magnitudes to reduce the penalty), the penalized relaxation finds a feasible solution to many OPF problems. Since losses typically account for a small percentage of active power generation and active and reactive power are typically loosely coupled, the reactive power penalization often results in a feasible point that is near the global optimum.

We next study the convexity properties of the cost function and the reactive power penalization. The cost function is convex in terms of active power generation but not necessarily in terms of the real and imaginary voltage components.² Thus, the objective function is a potential source of nonconvexity which may result in the relaxation’s failure to globally solve the OPF problem.

Consider the eigenvalues of the symmetric matrices \mathbf{C} and \mathbf{D} , where, for the vector \hat{x} containing the voltage components, $\hat{x}^T \mathbf{C} \hat{x}$ is a linear cost of active power generation and $\hat{x}^T \mathbf{D} \hat{x}$ calculates the reactive power losses. For the 2383-bus Polish system [132], which has linear generation costs, the most negative eigenvalue of \mathbf{C} is -8.53×10^7 . Thus, the objective function of the original OPF problem is nonconvex in terms of the voltage components, which can cause the SDP relaxation to fail to yield the global optimum. Conversely, active power loss minimization is convex in terms of the voltage components due to the absence of negative resistances.

As indicated by the potential for negative eigenvalues of \mathbf{D} (e.g., the matrix \mathbf{D} for the 2383-bus Polish system has a pair of negative eigenvalues at -0.0175), penalizing reactive power losses is generally nonconvex due to capacitive network elements (i.e., increasing voltage magnitudes may decrease the reactive power loss). See [83] for related work that uses a convex objective based on a Laplacian matrix.

Further work is needed to investigate the effects of reactive power penalization on OPF problems with more realistic generator models that explicitly consider the trade-off between active and reactive power outputs (i.e., generator “D-curves”). A tighter coupling between active and reactive power generation may cause the reactive power penalization to yield

²The cost function of the *moment relaxation* is always convex. This section studies the convexity of the penalized objective function for the *original* nonconvex OPF problem.

solutions that are far from the global optimum.

The apparent power line loss penalty’s effects are not as easy to interpret as the reactive power penalty. Ongoing work includes understanding the effects of the line loss penalty.

5.5 Moment+penalization approach

Although the reactive power penalization often yields a near rank-one solution, this penalization alone is not sufficient to obtain a feasible point for many problems. Reference [72] penalizes the apparent power line loss associated with certain lines to address the few remaining non-rank-one “problematic” submatrices. However, this approach has several disadvantages.

First, penalizing apparent power line losses introduces another parameter.³ Introducing parameters is problematic, especially when lacking an intuition for appropriate values.

Second, the combination of reactive power and line loss penalization may not yield a feasible solution to some problems. For instance, the OPF problems case9mod and case39mod1 from [22] are globally solved with low-order moment relaxations, but there is no known penalization of reactive power and/or apparent power line loss that yields a feasible solution for these problems. Also, the penalization approach is not guaranteed to yield a feasible solution that is close to the global optimum.

Unlike the penalization approach, the moment relaxation approach does not require the choice of penalty parameters, globally solves a broader class of OPF problems, and is guaranteed to yield the global optimum when the rank-one condition is satisfied. However, direct application of the moment relaxations to large problems has so far been limited to active power loss minimization objective functions. We conjecture that the nonconvexity associated with more general cost functions requires higher-order moment constraints at too many buses for computational tractability.

To apply the moment relaxations to large OPF problems with active power generation cost objective functions, we augment the moment relaxations with a reactive power penalty. Specifically, we add to the objective the total reactive power produced by all generators multiplied by a penalization parameter ϵ_b (which is a positive scalar). That is, rather than apply an apparent power loss penalization to the objective function, we apply higher-order moment constraints to specific buses [81]. As will be demonstrated in Section 6.4, higher-order moment constraints are only needed at a few buses in typical OPF problems after augmenting the objective function with a reactive power penalization term.

Similar to the existing penalization, when the rank condition is satisfied, the proposed “moment+penalization” approach yields the global solution to the *modified* OPF problem, but not necessarily to the original OPF problem. However, since the penalization does not change the constraint equations, the solution to the moment+penalization approach is *feasible* for the original OPF problem. The first-order moment relaxation without penalization (i.e., $\epsilon_b = 0$)

³Reference [72] uses the same penalization parameter for each “problematic” line. Generally, each line could have a different penalty parameter.

gives a lower bound on the globally optimal objective value for the original OPF problem. This lower bound provides an optimality metric for the feasible solution obtained from the moment+penalization approach. As will be shown in Section 6.4, the feasible solutions for a variety of problems are within at least 1% of the global optimum.

The moment+penalization approach inherits a mix of the advantages and disadvantages of the moment relaxation and penalization methods. First, the moment+penalization approach requires selection of a single scalar parameter (one more than needed for the moment relaxations, but one less than generally needed for the penalization in [72]). This parameter must be large enough to result in a near rank-one solution, but small enough to avoid large changes to the OPF problem.

Second, the penalization eliminates the moment relaxations' guarantees: the moment+penalization approach may yield a feasible solution that is far from the global optimum or not give any solution. However, the moment+penalization approach finds global or near-global solutions to a broader class of small OPF problems than penalization approach of [72] (e.g., case9mod and case39mod1 with $\epsilon_b = 0$, and case39mod3 with $\epsilon_b = \$0.10/\text{MVar}$ [22]). This suggests that the moment+penalization approach inherits the ability of the moment relaxations to solve a broad class of OPF problems.

Finally, the penalization in the moment+penalization approach enables calculation of feasible solutions that are at least nearly globally optimal for a variety of large OPF problems with objective functions that minimize active power generation cost rather than just active power losses.

Note that it is not straightforward to compare the computational costs of the moment+penalization approach and the penalization approach in [72]. A single solution of a penalized first-order moment relaxation, as in [72], is faster than a relaxation with higher-order moment constraints. Thus, if one knows appropriate penalty parameters, the method in [72] is faster. Although a relatively wide range of penalty parameters tends to work well for typical OPF problems, there are problems for which no known penalty parameters yield feasible solutions. For these problems, the moment+penalization approach has a clear advantage.

The moment+penalization approach has the advantage of systematically tightening the relaxation rather than requiring the choice of penalty parameters. However, the higher-order constraints can significantly increase solver times. Thus, there is a potential trade-off between finding appropriate penalization parameters for the approach in [72] and increased solver time from the moment+penalization approach. The speed of the moment+penalization approach may be improved using the mixed SDP/SOCP relaxation from [80].

5.6 Numerical results

This section first globally solves several large, active-power-loss minimizing OPF problems using moment relaxations without penalization ($\epsilon_b = 0$). Next, this section applies the mo-

ment+penalization approach to find feasible points that are at least nearly globally optimal for several test cases which minimize active power generation cost. Unless otherwise stated, the preprocessing method from Section 5.2 with `thrshz` set to 1×10^{-3} per unit is applied to all examples. No example enforces a minimum line resistance.

The results are generated using the iterative algorithm from [81] which selectively applies the higher-order moment relaxation constraints. The algorithm terminates when all power injection mismatches are less than 1 MVA.

The implementation uses MATLAB 2013a, YALMIP 2015.06.26 [69], and Mosek 7.1.0.28, and was solved using a computer with a quad-core 2.70 GHz processor and 16 GB of RAM. The test cases are the Polish system models in MATPOWER [132] and several PEGASE systems [42, 55] representing portions of the European power system.

Active Power Loss Minimization Results

Table 5.1 shows the results of applying the moment relaxations to several large OPF problems that minimize active power losses. The solutions to the preprocessed problems are guaranteed to be globally optimal since there is no penalization. The columns in Table 5.1 list the case name, the number of iterations of the algorithm from [81], the maximum power injection mismatch, the globally optimal objective value, and the solver time summed over all iterations. The abbreviation “PL” stands for “Poland”. Table 5.1 excludes several cases (the 89-bus PEGASE system and the Polish 2736sp, 2737sop, and 2746wp systems) which only require the first-order relaxation and thus do not illustrate the capabilities of the higher-order relaxations. PEGASE-1354 and PEGASE-2869 use a `thrshz` of 3×10^{-3} per unit. All other systems use 1×10^{-3} per unit.

Table 5.1: Active Power Loss Minimization Results

Case Name	Num. Iter.	Global Obj. Val. (\$/hr)	Max S^{mis} (MVA)	Solver Time (sec)
PL-2383wp	3	24990	0.25	583
PL-2746wop	2	19210	0.39	2662
PL-3012wp	5	27642	1.00	319
PL-3120sp	7	21512	0.77	387
PEGASE-1354	5	74043	0.85	407
PEGASE-2869	6	133944	0.63	921

Each iteration of the algorithm in [81] after the first adds second-order constraints at two buses. Thus, a small number of second-order buses (between 0.1% and 0.7% of the number of buses in the systems in Table 5.1 after the low-impedance line preprocessing) are applied to all examples in Table 5.1. This results in computational tractability for the moment relaxations.

Note that PL-2746wop has a much greater solver time than the other systems even though it only has second-order constraints at two buses. This slow solution time is due to the fact

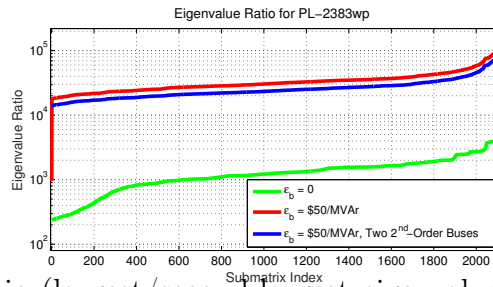


Figure 5.1: Eigenvalue ratio (largest/second-largest eigenvalue) for each submatrix for PL-2383wp. Large values ($>10^4$) indicate satisfaction of the rank-1 condition. For $\epsilon_b = 0$ (green), most of the submatrices are not rank one. For $\epsilon_b = \$50/\text{MVar}$ (red), most submatrices satisfy the rank condition with the exception of those on the far left of the figure. Applying second-order moment constraints to two of the buses that are in these submatrices (blue) results in all submatrices satisfying the rank condition.

that the two second-order buses are contained in submatrices corresponding to cliques with 10 and 11 buses. The second-order constraints for these large submatrices dominate the solver time. The mixed SDP/SOCP relaxation in [80] may be particularly useful beneficial for such cases.

Since the low-impedance line preprocessing has been applied to these systems, the solutions do not exactly match the original OPF problems. MATPOWER [132] solutions of the original problems have objective values that are slightly larger than the values in Table 5.1 due to losses associated with the line resistances removed by the preprocessing.

After the low-impedance line preprocessing, local solutions from MATPOWER match the solutions from the moment relaxations and are therefore, in fact, globally optimal. This is not the case for all OPF problems [22, 81].

Moment+Penalization for More General Cost Functions

As discussed in Section 5.4, minimization of active power generation cost often yields a nonconvex objective function in terms of the voltage components. Despite this nonconvexity, low-order moment relaxations typically yield global solutions to small problems, including problems without known penalty parameters for obtaining a feasible points (e.g., case9mod and case39mod1 from [22]).

However, the moment relaxations are thus far intractable for some large OPF problems with nonconvex objective functions. A reactive power penalty often results in the first-order moment relaxation yielding a solution that is nearly globally optimal (i.e., most of the submatrices in the clique decomposition satisfy the rank-one condition). Enforcing higher-order constraints at buses in the remaining submatrices yields a feasible solution to the OPF problem. This is illustrated in Fig. 5.1, which shows the ratio between the largest and second-largest eigenvalues of the submatrices of the moment matrix, arranged in increasing order, for the 2383-bus Polish system. If the submatrices were all rank one, then this eigenvalue ratio

would be infinite. Thus, large numeric values (i.e., greater than 1×10^4) indicate satisfaction of the rank condition within numerical precision. Without the reactive power penalty, the rank condition is not satisfied for most submatrices. With the reactive power penalty, the rank condition is satisfied for many but not all submatrices. Enforcing higher-order moment constraints at two buses which are in the high-rank submatrices results in a feasible (rank-one) operating point for the OPF problem which is within at least 0.74% of the global optimum.

To further illustrate the effectiveness of the moment+penalization approach, Table 5.2 shows the results of applying the moment+penalization approach to several large OPF problems with active power generation cost functions. The optimality gap column gives the percent difference between a lower bound on the optimal objective value from the first-order moment relaxation and the feasible solution obtained from the moment+penalization approach for the system after low-impedance line preprocessing.

Table 5.2: Generation Cost Minimization Results

Case Name	ϵ_b (\$/MVA _r)	Num. Iter.	Opt. Gap	Max S^{mis} (MVA)	Solver Time (sec)
PL-2383wp	50	2	0.74%	0.13	152.2
PL-3012wp	50	7	0.49%	0.20	1056.3
PL-3120sp	100	6	0.92%	0.08	1164.4

The penalized first-order relaxation requires 74.6, 88.9, and 97.0 seconds for PL-2383wp, PL-3012wp, and PL-3120sp, respectively. Attributing the rest of the solver time to the higher-order constraints implies that these constraints accounted for 3.1, 433.7, and 582.4 seconds beyond the time required to repeatedly solve the first-order relaxations.

The moment+penalization approach can yield feasible points that are at least nearly globally optimal for cases where both the penalization method of [72] and low-order moment relaxations fail individually. For instance, the moment+penalization approach with a reactive power penalty of $\epsilon_b = \$0.10/\text{MVA}_r$ gives a feasible point within 0.28% of the global optimum for case39mod3 from [22], but both second- and third-order moment relaxations and the penalization method in [72] fail to yield global solutions.

5.7 Conclusion

“Moment” relaxations from the Lasserre hierarchy for polynomial optimization globally solve a broad class of OPF problems. By exploiting sparsity and selectively applying the computationally intensive higher-order moment relaxations, previous literature demonstrated the moment relaxations’ capability to globally solve moderate-size OPF problems. This chapter presented a preprocessing method that removes low-impedance lines to improve the numerical conditioning of the moment relaxations. After applying the preprocessing method, the moment relaxations globally solve a variety of OPF problems that minimize active power losses for systems with thousands of buses. A proposed “moment+penalization” method is

capable of finding feasible points that are at least nearly globally optimal for large OPF problems with more general cost functions. This method has several advantages over previous penalization approaches, including requiring fewer parameter choices and solving a broader class of OPF problems. In the next chapter, we devise a method that provides nearly global solutions to large-scale OPF problems without the need to specify any parameter.

Chapter 6

Laplacian matrix approach

A semidefinite optimization (SDP) relaxation globally solves many optimal power flow (OPF) problems. For other OPF problems where the SDP relaxation only provides a lower bound on the objective value rather than the globally optimal decision variables, recent literature has proposed a penalization approach to find feasible points that are often nearly globally optimal. A disadvantage of this penalization approach is the need to specify penalty parameters. This paper presents an alternative approach that algorithmically determines a penalization appropriate for many OPF problems. The proposed approach constrains the generation cost to be close to the lower bound from the SDP relaxation. The objective function is specified using iteratively determined weights for a Laplacian matrix. This approach yields feasible points to the OPF problem that are guaranteed to be near the global optimum due to the constraint on generation cost. The proposed approach is demonstrated on both small OPF problems and a variety of large test cases representing portions of European power systems. The material presented in this chapter is based on the submitted manuscript:

D.K. MOLZAHN, C. JOSZ, I.A. HISKENS, AND P. PANCIATICI, *A Laplacian-Based Approach for Finding Near Globally Optimal Solutions to OPF Problems*, submitted to Institute of Electrical and Electronics Engineers, Transactions on Power Systems. [preprint]

6.1 Introduction

Literature has proposed an objective function penalization approach for finding feasible points that are near the global optimum for the OPF problem [72, 74]. The penalization approach has the advantage of not using potentially computationally expensive higher-order moment constraints, but has the disadvantage of requiring the choice of appropriate penalization parameters. This choice involves a compromise, as the parameters must induce a feasible solution to the original problem while avoiding large modifications to the problem that would cause unacceptable deviation from the global optimum.

The penalization formulation in the existing literature [72] generally requires specifying penalty parameters for both the total reactive power injection and apparent power flows on certain lines. Penalty parameters in the literature range over several orders of magnitude for various test cases, and existing literature largely lacks systematic algorithms for determining appropriate parameter values. Recent work [84] proposes a “moment+penalization” approach that eliminates the need to choose apparent power flow penalization parameters, but still requires selection of a penalty parameter associated with the total reactive power injection.

This chapter presents an iterative algorithm that builds an objective function intended to yield near-globally-optimal solutions to OPF problems for which the SDP relaxation is not exact. The proposed algorithm first solves the SDP relaxation to obtain a lower bound on the optimal objective value. This lower bound is often very close to the global optimum of many practical OPF problems. The proposed approach modifies the SDP relaxation by adding a constraint that the generation cost must be within a small percentage (e.g., 0.5%) of this lower bound. This percentage is the single externally specified parameter in the proposed approach.

This constraint on the generation cost provides freedom to specify an objective function that aims to obtain a *feasible* rather than *minimum-cost* solution for the OPF problem. In other words, we desire an objective function such that the SDP relaxation yields a feasible solution to the original nonconvex OPF problem, with near-global optimality ensured by the constraint on generation cost.

This chapter proposes an algorithm for calculating an appropriate objective function defined using a weighted Laplacian matrix. The weights are determined iteratively based on the mismatch between the solution to the relaxation and the power flows resulting from a related set of voltages. The paper will formalize these concepts and demonstrate that this approach results in near global solutions to many OPF problems, including large test cases. Like many penalization/regularization techniques, the proposed approach is not guaranteed to yield a feasible solution. As supported by the results for several large-scale, realistic test cases, the proposed algorithm broadens the applicability of the SDP relaxation to achieve operating points for many OPF problems that are within specified tolerances for both constraint feasibility and global optimality.

There is related work that chooses the objective function of a relaxation for the purpose of obtaining a feasible solution for the original nonconvex problem. For instance, [67] specifies objective functions that are linear combinations of squared voltage magnitudes in order to find multiple solutions to the power flow equations. Additionally, [73] proposes a method for determining an objective function that yields solutions to the power flow equations for a variety of parameter choices. The objective function in [73] is defined by a matrix with three properties: positive semidefiniteness, a simple eigenvalue of 0, and null space containing the all-ones vector. We note that the weighted Laplacian objective function developed in this paper is a special case of an objective function that also has these three properties.

This chapter is organized as follows. Section 6.2 describes the Laplacian objective function

approach that is the main contribution of this chapter. Section 6.3 describes an algorithm for determining the Laplacian weights. Section 6.4 demonstrates the effectiveness of the proposed approach through application to a variety of small OPF problems as well as several large test cases representing portions of European power systems. Section 6.5 concludes the chapter.

6.2 Laplacian objective function

The proposed approach exploits the empirical observation that the SDP relaxation provides a very close lower bound on the optimal objective value of many typical OPF problems (i.e., there is a very small *relaxation gap*). For instance, the SDP relaxation gaps for the large-scale Polish [132] and PEGASE [42, 55] systems, which represent portions of European power systems, are all less than 0.3%.¹² Further, the SDP relaxation is *exact* (i.e., zero relaxation gap) for the IEEE 14-, 30-, 39-, 57-bus systems, the 118-bus system modified to enforce a small minimum line resistance, and several of the large-scale Polish test cases.

We assume that the lower bound c^* provided by the SDP relaxation is within a given percentage δ of the global optimum to the OPF problem. (Most of the examples in Section 6.4 specify $\delta = 0.5\%$.) We constrain the generation cost using this assumption by requiring it to be less than or equal to $c^*(1 + \delta)$. Note that the feasible set thus defined is non-empty (i.e., the corresponding SDP problem is feasible) for any choice of $\delta \geq 0$. However, if δ is too small, there may not exist a rank-one matrix \mathbf{W} (i.e., a feasible point for the original OPF problem) in the feasible space.

The lack of a priori guarantees on the size of the relaxation gap is a challenge that the proposed approach shares with many related approaches for convex relaxations of the optimal power flow problem. Existing sufficient conditions that guarantee zero relaxation gap generally require satisfaction of non-trivial technical conditions and a limited set of network topologies [70, 74]. The SDP relaxation is exact for a significantly broader class of OPF problems than those that have a priori exactness guarantees, and has a small relaxation gap for an even broader class of OPF problems.

There are test cases that are specifically constructed to exhibit somewhat anomalous behavior in order to test the limits of the convex relaxations. The SDP relaxation gap is not small for some of these test cases. For instance, the 3-bus system in [78], the 5-bus system in [85], and the 9-bus system in [22] have relaxation gaps of 20.6%, 8.9%, and 10.8%, respectively, and the test cases in [58] have relaxation gaps as large as 52.7%. The approach proposed in this paper is not appropriate for such problems. Future progress in convex relaxation theory is required to develop broader conditions that provide a priori certification

¹To obtain satisfactory convergence of the SDP solver, these systems are pre-processed to remove low-impedance lines (i.e., lines whose impedance values have magnitudes less than 1×10^{-3} per unit) as in [84].

²These relaxation gaps are calculated using the objective values from the Shor SDP relaxation and solutions obtained either from the second-order moment relaxation [81] (where possible) or from MATPOWER [132].

that the SDP relaxation is exact or has a small relaxation gap. We also await the development of more extensive sets of OPF test cases to further explore the observation that many typical existing practical test cases have small SDP relaxation gaps.

By inserting the original cost function as a constraint, this effectively frees the choice of the objection function (which we will denote $f(\mathbf{W})$) to obtain a *feasible* rather than *minimum-cost* solution to the OPF. Here, \mathbf{W} denotes the real SDP matrix variable of $2n \times 2n$ where n is the number of buses. Any solution with $\text{rank}(\mathbf{W}) = 1$ yields a feasible solution to the OPF problem within δ of the globally optimal objective value. We therefore seek an objective function $f(\mathbf{W})$ which maximizes the likelihood of obtaining $\text{rank}(\mathbf{W}) = 1$. This section describes a Laplacian form for the function $f(\mathbf{W})$. Specifically, we consider a $n_l \times n_l$ diagonal matrix \mathbf{D} containing weights for the network Laplacian matrix $\mathbf{L} = \mathbf{A}_{inc}^\top \mathbf{D} \mathbf{A}_{inc}$, where \mathbf{A}_{inc} is the $n_l \times n$ incidence matrix for the network. The off-diagonal term \mathbf{L}_{ij} is equal to the negative of the sum of the weights for the lines connecting buses i and j , and the diagonal term \mathbf{L}_{ii} is equal to the sum of the weights of the lines connected to bus i . The objective function is

$$f(\mathbf{W}) = \text{tr} \left(\begin{bmatrix} \mathbf{L} & \mathbf{0}_{n \times n} \\ \mathbf{0}_{n \times n} & \mathbf{L} \end{bmatrix} \mathbf{W} \right). \quad (6.1)$$

The choice of an objective function based on a Laplacian matrix is motivated by previous literature. An existing penalization approach [74] augments the objective function by adding a term that minimizes the total reactive power injection. This reactive power penalty can be implemented by adding the term

$$\epsilon_b \text{tr} \left(\begin{bmatrix} \Re \left(\frac{\mathbf{Y}^H - \mathbf{Y}}{2j} \right) & \Im \left(\frac{\mathbf{Y}^H - \mathbf{Y}}{2j} \right) \\ -\Im \left(\frac{\mathbf{Y}^H - \mathbf{Y}}{2j} \right) & \Re \left(\frac{\mathbf{Y}^H - \mathbf{Y}}{2j} \right) \end{bmatrix} \mathbf{W} \right) \quad (6.2)$$

to the objective function of the Shor relaxation of the OPF, where ϵ_b is a specified penalty parameter, $(\cdot)^H$ indicates conjugate transpose, and \mathbf{Y} is the admittance matrix. In the absence of phase-shifting transformers (i.e., $\theta_{lm} = 0 \quad \forall (l, m) \in \mathcal{L}$), the matrix $\frac{\mathbf{Y}^H - \mathbf{Y}}{2j}$ is equivalent to $-\Im(\mathbf{Y}) = -\mathbf{B}$, which is a *weighted Laplacian matrix* (with weights determined by the branch susceptance parameters $b_{lm} = \frac{-X_{lm}}{R_{lm}^2 + X_{lm}^2}$) plus a diagonal matrix composed of shunt susceptances.

Early work on SDP relaxations of OPF problems [66] advocates enforcing a minimum resistance of ϵ_r for all lines in the network. For instance, the SDP relaxation fails to be exact for the IEEE 118-bus system [1], but the relaxation is exact after enforcing a minimum line resistance of $\epsilon_r = 1 \times 10^{-4}$ per unit. After enforcing a minimum line resistance, the active power losses are given by

$$\text{tr} \left(\begin{bmatrix} \Re \left(\frac{\mathbf{Y}_r + \mathbf{Y}_r^H}{2} \right) & \Im \left(\frac{\mathbf{Y}_r + \mathbf{Y}_r^H}{2} \right) \\ -\Im \left(\frac{\mathbf{Y}_r + \mathbf{Y}_r^H}{2} \right) & \Re \left(\frac{\mathbf{Y}_r + \mathbf{Y}_r^H}{2} \right) \end{bmatrix} \mathbf{W} \right) \quad (6.3)$$

where \mathbf{Y}_r is the network admittance matrix after enforcing a minimum branch resistance of ϵ_r . In the absence of phase-shifting transformers, $\frac{\mathbf{Y}_r + \mathbf{Y}_r^H}{2}$ is equivalent to $\Re(\mathbf{Y}_r)$, which is a *weighted Laplacian matrix* (with weights determined by the branch conductance parameters $g_{lm} = \frac{R_{lm}}{R_{lm}^2 + X_{lm}^2}$) plus a diagonal matrix composed of shunt conductances. Since typical OPF problems have objective functions that increase with active power losses, enforcing minimum line resistances is similar to a weighted Laplacian penalization.³

The proposed objective function (6.1) is equivalent to a linear combination of certain components of \mathbf{W} :

$$f(\mathbf{W}) = \sum_{(l,m) \in \mathcal{L}} \mathbf{D}_{(l,m)} (\mathbf{W}_{ll} - 2\mathbf{W}_{lm} + \mathbf{W}_{mm} + \mathbf{W}_{l+n,l+n} - 2\mathbf{W}_{l+n,m+n} + \mathbf{W}_{m+n,m+n}) \quad (6.4)$$

where $\mathbf{D}_{(l,m)}$ is the diagonal element of \mathbf{D} corresponding to the line from bus l to bus m .

From a physical perspective, the Laplacian objective's tendency to reduce voltage differences is similar to both the reactive power penalization proposed in [74] and the minimum branch resistance advocated in [66]. For typical operating conditions, reactive power injections are closely related to voltage magnitude differences, so penalizing reactive power injections tends to result in solutions with similar voltages. Likewise, the active power losses associated with line resistances increase with the square of the current flow through the line, which is determined by the voltage difference across the line. Thus, enforcing a minimum line resistance tends to result in solutions with smaller voltage differences in order to reduce losses.

In addition, a Laplacian regularizing term has been used to obtain desirable solution characteristics for a variety of other optimization problems (e.g., machine learning problems [77, 109], sensor network localization problems [128], and analysis of flow networks [120]). The Laplacian matrix is also used for image reconstruction problems [31].

6.3 Determining Laplacian weights

Having established a weighted Laplacian form for the objective function, we introduce an iterative algorithm for determining appropriate weights \mathbf{D} for obtaining a solution with $\text{rank}(\mathbf{W}) = 1$. We note that the proposed algorithm is similar in spirit to the method in [24, Section 2.4], which iteratively updates weighting parameters to promote low-rank solutions of SDPs related to image reconstruction problems.

The proposed algorithm is inspired by the apparent power line flow penalty used in [72] and the iterative approach to determining appropriate buses for enforcing higher-order moment constraints in [81]. The approach in [72] penalizes the apparent power flows on lines

³Note that since enforcing minimum line resistances also affects the power injections and the line flows, the minimum line resistance cannot be solely represented as a Laplacian penalization of the objective function.

associated with certain submatrices of \mathbf{W} that are not rank one.⁴ Similar to the approach in [72], the proposed algorithm adds terms to the objective function that are associated with certain “problematic lines.”

The heuristic for identifying problematic lines is inspired by the approach used in [81] to detect “problematic buses” for application of higher-order moment constraints. Denote the SDP solution as \mathbf{W}^* and the closest rank-one matrix to \mathbf{W}^* as $\mathbf{W}^{(1)}$. Previous work [81] compares the power injections associated with \mathbf{W}^* and $\mathbf{W}^{(1)}$ to calculate power injection mismatches $S_k^{inj\ mis}$ for each bus $k \in \mathcal{N}$:

$$S_k^{inj\ mis} = \left| \text{tr} \{ \mathbf{Y}_k (\mathbf{W}^* - \mathbf{W}^{(1)}) \} + \text{jtr} \{ \bar{\mathbf{Y}}_k (\mathbf{W}^* - \mathbf{W}^{(1)}) \} \right| \quad (6.5)$$

where $|\cdot|$ denotes the magnitude of the complex argument. In the parlance of [81], problematic buses are those with large power injection mismatches $S_k^{inj\ mis}$.

To identify problematic lines rather than buses, we modify (6.5) to calculate apparent power flow mismatches $S_{(l,m)}^{flow\ mis}$ for each line $(l, m) \in \mathcal{L}$:

$$S_{(l,m)}^{flow\ mis} = \left| \text{tr} \{ \mathbf{Z}_{lm} (\mathbf{W}^* - \mathbf{W}^{(1)}) \} + \text{jtr} \{ \bar{\mathbf{Z}}_{lm} (\mathbf{W}^* - \mathbf{W}^{(1)}) \} \right| + \left| \text{tr} \{ \mathbf{Z}_{ml} (\mathbf{W}^* - \mathbf{W}^{(1)}) \} + \text{jtr} \{ \bar{\mathbf{Z}}_{ml} (\mathbf{W}^* - \mathbf{W}^{(1)}) \} \right|. \quad (6.6)$$

Observe that $S_{(l,m)}^{flow\ mis}$ sums the magnitude of the apparent power flow mismatches at both ends of each line.

The condition $\text{rank}(\mathbf{W}^*) = 1$ (i.e., “feasibility” in this context) is considered satisfied for practical purposes using the criterion that the maximum line flow and power injection mismatches (i.e., $\max_{(l,m) \in \mathcal{L}} S_{(l,m)}^{flow\ mis}$ and $\max_{k \in \mathcal{N}} S_k^{inj\ mis}$) are less than specified tolerances ϵ_{flow} and ϵ_{inj} , respectively, and the voltage magnitude limits are satisfied to within a specified tolerance ϵ_V .⁵

As described in Algorithm 1, the weights on the diagonal of \mathbf{D} are determined from the line flow mismatches $S_{(l,m)}^{flow\ mis}$. Specifically, the proposed algorithm first solves the Shor relaxation to obtain both the lower bound c^* on the optimal objective value and the initial line flow and power injection mismatches $S_{(l,m)}^{flow\ mis}$, $\forall (l, m) \in \mathcal{L}$ and $S_k^{inj\ mis}$, $\forall k \in \mathcal{N}$.

While the termination criteria ($\max_{(l,m) \in \mathcal{L}} \{ S_{(l,m)}^{flow\ mis} \} < \epsilon_{flow}$, $\max_{k \in \mathcal{N}} \{ S_k^{inj\ mis} \} < \epsilon_{inj}$, and no voltage limits violated by more than ϵ_V) are not satisfied, the algorithm solves the SDP relaxation with the constraint ensuring that the generation cost is within δ of the lower bound. The objective function is defined using the weighting matrix $\mathbf{D} = \text{diag}(S^{flow\ mis})$,

⁴The submatrices are determined by the maximal cliques of a chordal supergraph of the network; see [52, 72, 82] for further details.

⁵For all test cases, the voltage magnitude limits were satisfied whenever the power injection and line flow mismatch tolerances were achieved.

Algorithm 1 Iterative Algorithm for Determining Weights

-
- 1: **Input:** tolerances ϵ_{flow} and ϵ_{inj} , max relaxation gap δ
 - 2: Set $\mathbf{D} = \mathbf{0}_{n_l \times n_l}$
 - 3: Solve the Shor SDP relaxation to obtain c^*
 - 4: Calculate $S^{flow\ mis}$ and $S^{inj\ mis}$ using (6.6) and (6.5)
 - 5: **while** termination criteria not satisfied
 - 6: Update weights: $\mathbf{D} \leftarrow \mathbf{D} + \text{diag}(S^{flow\ mis})$
 - 7: Solve the generation-cost-constrained relaxation
 - 8: Calculate $S^{flow\ mis}$ and $S^{inj\ mis}$ using (6.6) and (6.5)
 - 9: Calculate the voltage phasors and terminate
-

where $\text{diag}(\cdot)$ denotes the matrix with the vector argument on the diagonal and other entries equal to zero. Each iteration adds the line flow mismatch vector $S^{flow\ mis}$ to the previous weights (i.e., $\mathbf{D} \leftarrow \mathbf{D} + \text{diag}(S^{flow\ mis})$).

Note that Algorithm 1 is not guaranteed to converge. Non-convergence may be due to the value of δ being too small (i.e., there does not exist a rank-one solution) or failure to find a rank-one solution that does exist. To address the former case, Algorithm 1 could be modified to include an “outer loop” that increments δ by a specified amount (e.g., 0.5%) if convergence is not achieved in a certain number of iterations. We note that, like other convex relaxation methods, the proposed approach would benefit from further theoretical work regarding the development of a priori guarantees on the size of the relaxation gap for various classes of OPF problems.

For some problems with large relaxation gaps (e.g., the 3-bus system in [78], the 5-bus system in [85], and the 9-bus system in [22]), no purely penalization-based methods have so far successfully addressed the latter case where the proposed algorithm fails to find a rank-one solution that satisfies the generation cost constraint with sufficiently large δ (i.e., no known penalty parameters yield feasible solutions using the methods in [72, 74] for these test cases). One possible approach for addressing this latter case is the combination of penalization techniques with Lasserre’s moment relaxation hierarchy [44, 57, 79, 81]. The combination of the moment relaxations with the penalization methods enables the computation of near-globally-optimal solutions for a broader class of OPF problems than either method achieves individually. See [84] for further details on this approach.

We note that despite the lack of a convergence guarantee, the examples in Section 6.4 demonstrate that Algorithm 1 is capable of finding feasible points that are near the global optimum for many OPF problems, including large test cases.

6.4 Numerical results

This section demonstrates the effectiveness of the proposed approach using several small example problems as well as large test cases representing portions of European power systems. The SDP relaxation yields a small but non-zero relaxation gap for the test cases selected in this section, and Algorithm 1 yields points that are feasible for the OPF (to within the specified termination criteria) and that are near the global optimum for these test cases. For other test cases with a large SDP relaxation gap, such as those mentioned earlier in [22, 58, 78, 85], the proposed algorithm does not converge when tested with a variety of values for δ .

The results in this section use line flow and power injection mismatch tolerances ϵ_{flow} and ϵ_{inj} that are both equal to 1 MVA and $\epsilon_V = 5 \times 10^{-4}$ per unit. The implementation of Algorithm 1 uses MATLAB 2013a, YALMIP 2015.02.04 [69], and Mosek 7.1.0.28 [2], and was solved using a computer with a quad-core 2.70 GHz processor and 16 GB of RAM.

Applying Algorithm 1 to several small- to medium-size test cases from [1, 58, 67, 81, 85] yields the results shown in Table 6.1. Tables 6.2 and 6.3 show the results from applying Algorithm 1 to large test cases which minimize generation cost and active power losses, respectively. These test cases, which are from [132] and [42, 55], represent portions of European power systems. The Shor relaxation has a small but non-zero relaxation gap for all test cases considered in this section. The columns of Tables 6.1–6.3 show the case name and reference, the number of iterations of Algorithm 1, the final maximum apparent power flow mismatch $\max_{(l,m) \in \mathcal{L}} \{S_{(l,m)}^{flow\ mis}\}$ in MVA, the final maximum power injection mismatch $\max_{k \in \mathcal{N}} \{S_k^{inj\ mis}\}$ in MVA, the specified value of δ , an upper bound on the relaxation gap from the solution to the Shor relaxation, and the total solver time in seconds.

Note that the large test cases in Tables 6.2 and 6.3 were preprocessed to remove low-impedance lines as described in [84] in order to improve the numerical convergence of the SDP relaxation. Lines which have impedance magnitudes less than a threshold (`thrshz` in [84]) of 1×10^{-3} per unit are eliminated by merging the terminal buses. Table 6.4 describes the number of buses and lines before and after this preprocessing. Low-impedance line preprocessing was not needed for the test cases in Table 6.1. After preprocessing, MOSEK's SDP solver converged with sufficient accuracy to yield solutions that satisfied the voltage magnitude limits to within $\epsilon_V = 1 \times 10^{-4}$ per unit and the power injection and line flow constraints to within the corresponding mismatches shown in Tables 6.1–6.3.

These results show that Algorithm 1 finds feasible points (within the specified tolerances) that have objective values near the global optimum for a variety of test cases. Further, Algorithm 1 globally solves all OPF problems for which the Shor relaxation is exact (e.g., many of the IEEE test cases [66], several of the Polish test systems [82], and the 89-bus PEGASE system [42, 55]). Thus, the algorithm is a practical approach for addressing a broad class of OPF problems.

We note, however, that Algorithm 1 does not yield a feasible point for all OPF problems. For instance, the test case WB39mod from [22] has line flow and power injection mismatches

Table 6.1: Results for Small and Medium Size Test Cases

Case Name	Num. Iter.	Max Flow Mis. (MVA)	Max Inj. Mis. (MVA)	δ (%)	Max Relax. Gap (%)	Solver Time (sec)
LMBD3 [67]	1	1.3E-5	1.6E-5	0.5	0.50	0.7
MLD3 [85]	1	7.3E-6	7.2E-5	0.5	0.50	0.5
MH14Q [81]	2	1.8E-5	9.9E-6	0.5	0.02	1.2
MH14L [81]	2	8.1E-5	7.8E-5	0.5	0.33	1.2
KDS14Lin [58]	1	1.2E-3	9.2E-4	1.0	1.00	0.7
KDS14Quad [58]	1	1.4E-4	8.4E-5	1.0	1.00	0.6
KDS30Lin [58]	7	9.3E-1	9.2E-1	2.5	2.50	4.6
KDS30Quad [58]	6	8.1E-1	8.0E-1	2.0	2.00	3.6
KDS30IEEEQuad [58]	100	9.5E-1	7.2E-1	2.5	2.50	129.8
MH39L [81]	1	1.3E-2	9.8E-3	0.5	0.27	0.7
MH57Q [81]	1	1.2E-3	6.9E-4	0.5	0.03	0.7
MH57L [81]	1	3.2E-4	5.2E-4	0.5	0.16	0.9
MH118Q [81]	2	3.3E-3	2.7E-3	0.5	0.50	2.6
MH118L [81]	2	3.1E-3	3.1E-3	1.0	1.00	3.3
IEEE 300 [1]	1	1.3E-1	1.2E-1	0.5	0.01	3.0

of 18.22 MVA and 12.99 MVA, respectively, after 1000 iterations of Algorithm 1. The challenge associated with this case seems to result from light loading with limited ability to absorb a surplus of reactive power injections, yielding at least two local solutions. In addition to challenging the method proposed in this paper, no known penalty parameters yield feasible solutions to this problem. Generalizations of the SDP relaxation using the Lasserre hierarchy have successfully calculated the global solution to this case [81, 84]. Further, while Algorithm 1 converges for five of the seven test cases in [58] which have small relaxation gaps (less than 2.5%), the algorithm fails for two other such test cases as well as several other test cases in [58] which have large relaxation gaps. We note that the tree topologies used in the test cases in [58] are a significant departure from the mesh networks used in the standard test cases from which they were derived; the proposed algorithm succeeds for several test cases that share the original network topologies.

We note that the interior point solver in MATPOWER obtained superior relaxation gaps for the test cases considered in this paper. Within approximately five seconds for the large test cases in Tables 6.2 and 6.3, MATPOWER obtained relaxation gaps that ranged from 0.14%

Table 6.2: Results for Large Test Cases that Minimize Generation Cost

Case Name	Num. Iter.	Max Flow Mis. (MVA)	Max Inj. Mis. (MVA)	δ (%)	Max Relax. Gap (%)	Solver Time (sec)
PL-2383wp [132]	2	0.54	0.50	0.5	0.50	78.6
PL-3012wp [132]	2	0.36	0.27	0.5	0.50	107.6
PL-3120sp [132]	2	0.56	0.33	0.5	0.50	84.2

Table 6.3: Results for Large Test Cases that Minimize Active Power Loss

Case Name	Num. Iter.	Max Flow Mis. (MVA)	Max Inj. Mis. (MVA)	δ (%)	Max Relax. Gap (%)	Solver Time (sec)
PL-2383wp [132]	5	0.21	0.16	0.5	0.26	154.0
PL-3012wp [132]	5	0.08	0.04	0.5	0.18	232.2
PL-3120sp [132]	5	0.25	0.19	0.5	0.38	232.6
PEGASE-1354 [42, 55]	12	0.27	0.18	0.5	0.15	199.2
PEGASE-2869 [42, 55]	38	0.91	0.69	0.5	0.15	2378.4

Table 6.4: Descriptions of Large Test Cases Before and After Low-Impedance Line Preprocessing

Case Name	Before Preprocessing		After Preprocessing	
	Num. Buses	Num. Lines	Num. Buses	Num. Lines
PL-2383wp	2,383	2,869	2,177	2,690
PL-3012wp	3,012	3,572	2,292	2,851
PL-3120sp	3,120	3,693	2,314	2,886
PEGASE-1354	1,354	1,991	1,179	1,803
PEGASE-2869	2,869	4,582	2,120	4,164

to 0.32% smaller than those obtained with Algorithm 1.⁶ This suggests that smaller values of δ could be used in Algorithm 1. Indeed, additional numerical experiments demonstrated that Algorithm 1 converged with $\delta = 0.25\%$ (half the value used in previous numerical experiments) for all test cases for which the MATPOWER solution indicated that a value of

⁶Of course, MATPOWER cannot provide any measure of the quality of its solution in terms of a lower bound on the globally optimal objective value whereas Algorithm 1 provides such guarantees.

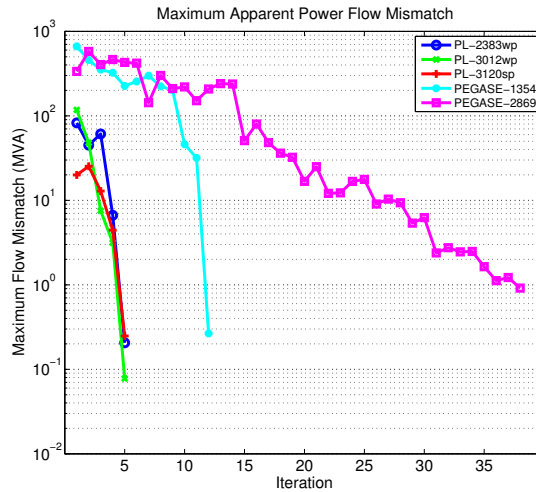


Figure 6.1: Maximum Apparent Power Flow Mismatches versus Iteration of Algorithm 1 for Active Power Loss Minimizing Test Cases

$\delta = 0.25\%$ was achievable.

We select termination parameter values of ϵ_{flow} and ϵ_{inj} of 1 MVA. This tolerance is typically numerically achievable with MOSEK’s SDP solver, which experience suggests is often a limiting factor to obtaining smaller mismatches.

Note that the maximum mismatches do not necessarily decrease monotonically with each iteration of Algorithm 1. Figs. 6.1 and 6.2 show the maximum flow mismatches (on a logarithmic scale) for the test cases that minimize active power losses (cf. Table 6.3). Likewise, Figs. 6.3 and 6.4 show the maximum power injection mismatches for the same test cases. Although the mismatches do not always decrease monotonically, there is a generally decreasing trend which results in satisfaction of the termination criteria for each test case. At each iteration, Algorithm 1 yields larger reactive power mismatches than active power mismatches for these test cases.

Note that it is not straightforward to compare the computational costs of the Laplacian objective approach and other penalization approaches in the literature [72, 74]. A single solution of the penalized SDP relaxations in [72] requires approximately the same computational effort as one iteration of Algorithm 1. Thus, if one knows appropriate penalty parameters, the method in [72] is faster for problems where the SDP relaxation is not exact. However, the key advantage of the proposed approach is that there is no need to specify any parameters other than the value of δ used in the generation cost constraint. In contrast, the literature largely lacks systematic approaches for identifying appropriate parameter values for the penalization methods in [72, 74].

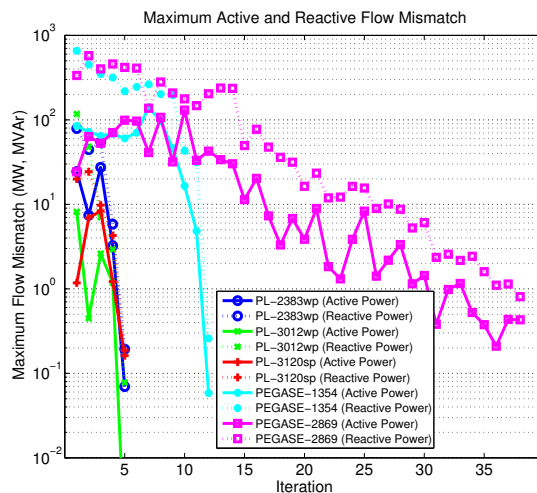


Figure 6.2: Maximum Active and Reactive Power Flow Mismatches versus Iteration of Algorithm 1 for Active Power Loss Minimizing Test Cases

6.5 Conclusion

The SDP relaxation of [66] is capable of globally solving a variety of OPF problems. To address a broader class of OPF problems, this paper has described an approach that finds feasible points with objective values that are within a specified percentage of the global optimum. Specifically, the approach in this paper adds a constraint to ensure that the generation cost is within a small specified percentage of the lower bound obtained from the SDP relaxation. This constraint frees the objective function to be chosen to yield a *feasible* (i.e., rank-one) solution rather than a *minimum-cost* solution. Inspired by previous penalization approaches and results in the optimization literature, an objective function based on a weighted Laplacian matrix is selected. The weights for this matrix are iteratively determined using “line flow mismatches.” The proposed approach is validated through successful application to a variety of both small and large test cases, including several OPF problems representing large portions of European power systems. There are, however, test cases for which the approach takes many iterations to converge or does not converge at all.

Future work includes modifying the algorithm for choosing the weights in order to more consistently require fewer iterations. Also, future work includes testing alternative SDP solution approaches with “hot start” capabilities to improve computational efficiency by leveraging knowledge of the solution to a “nearby” problem from the previous iteration of the algorithm. Future work also includes extension of the algorithm to a broader class of OPF problems, such as the test case WB39mod from [22] and several examples in [58].

In this chapter and the one before that, two methods were proposed to find nearly global solutions to large-scale OPF problems with generation cost minimization. In the case of active

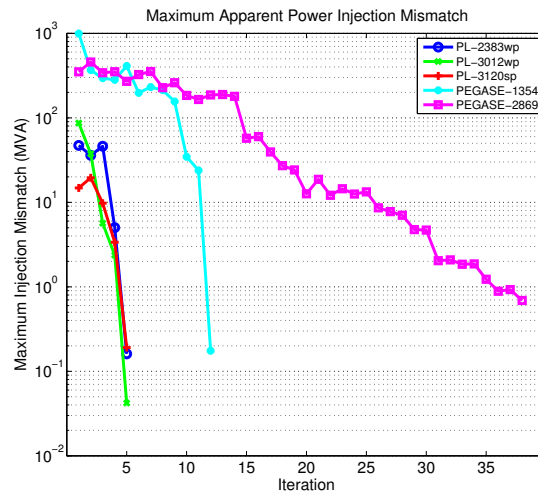


Figure 6.3: Maximum Apparent Power Injection Mismatches versus Iteration of Algorithm 1 for Active Power Loss Minimizing Test Cases

power loss minimization, Lasserre’s hierarchy finds the global solution to large-scale problems, provided the data is preprocessed to avoid bad conditioning. The difference between the two cases is that active power loss is a convex function of voltages, whereas generation cost is not. In the next chapter, Lasserre’s hierarchy is transposed to complex numbers in order to reduce its computational cost for OPF problems. Moreover, the sparsity exploiting algorithm from [81] designed for OPF problems is formalized for general polynomial optimization problems with either real or complex variables.

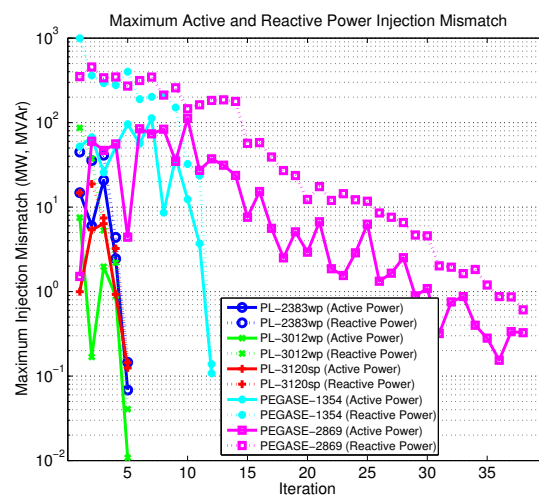


Figure 6.4: Maximum Active and Reactive Power Injection Mismatches versus Iteration of Algorithm 1 for Active Power Loss Minimizing Test Cases

Chapter 7

Complex hierarchy for enhanced tractability

We consider the problem of finding the global optimum of a real-valued complex polynomial $f : z \in \mathbb{C}^n \mapsto \sum_{\alpha, \beta} f_{\alpha, \beta} \bar{z}^\alpha z^\beta \in \mathbb{R}$ ($z^\alpha := z_1^{\alpha_1} \dots z_n^{\alpha_n}$, $\overline{f_{\alpha, \beta}} = f_{\beta, \alpha}$) on a compact set defined by real-valued complex polynomial inequalities. It reduces to solving a sequence of complex semidefinite optimization relaxations that grow tighter and tighter thanks to D'Angelo's and Putinar's Positivstellensatz discovered in 2008. In other words, the Lasserre hierarchy may be transposed to complex numbers. We propose a method for exploiting sparsity and apply the complex hierarchy to problems with several thousand complex variables. These problems consist of computing optimal power flows in the European high-voltage transmission network. The material presented in this chapter is based on the submitted manuscript:

C. JOSZ, D. K. MOLZAHN, *Moment/Sum-of-Squares Hierarchy for Complex Polynomial Optimization*, submitted to Society for Industrial and Applied Mathematics, Journal on Optimization. [preprint]

7.1 Introduction

Multivariate polynomial optimization where variables and data are complex numbers is a non-deterministic polynomial-time hard problem that arises in various applications such as electric power systems (Section 7.4), imaging science [14, 24, 43, 108], signal processing [3, 12, 29, 68, 71, 75], automatic control [121], and quantum mechanics [49]. Complex numbers are typically used to model oscillatory phenomena which are omnipresent in physical systems. Although complex polynomial optimization problems can readily be converted into real polynomial optimization problems where variables and data are real numbers, efforts have been made to find *ad hoc* solutions to complex problems [53, 54, 112]. The observation that relaxing nonconvex constraints and converting from complex to real numbers are two non-commutative

operations motivates our work. This leads us to transpose to complex numbers Lasserre's moment/sum-of-squares hierarchy [62] for real polynomial optimization.

The moment/sum-of-squares hierarchy succeeds to the vast development of real algebraic geometry during the twentieth century [91]. In 1900, Hilbert's seventeenth problem [103] raised the question of whether a non-negative polynomial in multiple real variables can be decomposed as a sum of squares of fractions of polynomials, to which Artin [10] answered in the affirmative in 1927. Later, positive polynomials on sets defined by a finite number of polynomial inequality constraints were investigated by Krivine [59], Stengle [113], Schmüdgen [102], and Putinar [92]. A theorem concerning such polynomials is referred to as *Positivstellensatz* [101]. For instance, Putinar proved under an assumption slightly stronger than compactness that they can be decomposed as a weighted sum of the constraints where the weights are sums of squares of polynomials. Lasserre [61, 62, 64] used this result in 2001 to develop a hierarchy of semidefinite programs to solve real polynomial optimization problems with compact feasible sets, with Parrilo [89, 90] making a similar contribution independently. In order to satisfy the assumption made by Putinar, Lasserre proposed to add a redundant ball constraint $x_1^2 + \dots + x_n^2 \leq R^2$ to the description of the feasible set when it is included in a ball of radius R . Subsequent work on the hierarchy includes its comparison with lift-and-project methods [65], a new proof of Putinar's Positivstellensatz via a 1928 theorem of Pólya [104], and a proof of generically finite convergence [87].

In 1968, Quillen [98] showed that a real-valued bihomogenous complex polynomial that is positive away from the origin can be decomposed as a sum of squared moduli of holomorphic polynomials when it is multiplied by $(|z_1|^2 + \dots + |z_n|^2)^r$ for some $r \in \mathbb{N}$. The result was rediscovered years later by Catlin and D'Angelo [28] and ignited a search for complex analogues of Hilbert's seventeenth problem [35, 36] and the ensuing Positivstellensätze [38, 94–96]. Notably, D'Angelo and Putinar [37] proved in 2008 that a positive complex polynomial on a sphere intersected by a finite number of polynomial inequality constraints can be decomposed as a weighted sum of the constraints where the weights are sums of squared moduli of holomorphic polynomials. Similar to Lasserre, we use D'Angelo's and Putinar's Positivstellensatz to construct a complex moment/sum-of-squares hierarchy of semidefinite programs to solve complex polynomial optimization problems with compact feasible sets. To satisfy the assumption in the Positivstellensatz, we propose to add a slack variable $z_{n+1} \in \mathbb{C}$ and a redundant constraint $|z_1|^2 + \dots + |z_{n+1}|^2 = R^2$ to the description of the feasible set when it is in a ball of radius R . The complex hierarchy is more tractable than the real hierarchy yet produces potentially weaker bounds. Computational advantages are shown using the optimal power flow problem in electrical engineering.

Below, Section 7.2 uses Shor and second-order conic relaxations to motivate the construction of a complex moment/sum-of-squares hierarchy in Section 7.3. Using a sparsity-exploiting method, numerical experiments on the optimal power flow problem are presented in Section 7.4. Section 7.5 concludes our work.

7.2 Motivation

Let \mathbb{N} , \mathbb{N}^* , \mathbb{R} , \mathbb{R}_+ and \mathbb{C} denote the set of natural, positive natural, real, non-negative real, and complex numbers respectively. Also, let “ \mathbf{i} ” denote the imaginary unit and \mathbb{H}_n denote the set of Hermitian matrices of order $n \in \mathbb{N}^*$. Let’s begin with the subclass of complex polynomial optimization composed of quadratically-constrained quadratic programs

$$\text{QCQP-}\mathbb{C} : \inf_{z \in \mathbb{C}^n} z^H H_0 z, \quad (7.1a)$$

$$\text{s.t. } z^H H_i z \leq h_i, \quad i = 1, \dots, m, \quad (7.1b)$$

where $m \in \mathbb{N}^*$, $H_0, \dots, H_m \in \mathbb{H}_n$, $h_0, \dots, h_m \in \mathbb{R}$, and $(\cdot)^H$ denotes the conjugate transpose. The feasible set is not assumed to contain a point (i.e. it may be empty). The Shor [106] and second-order conic relaxations of QCQP- \mathbb{C} share the following property: it is better to relax nonconvex constraints before converting from complex to real numbers rather than to do the two operations in the opposite order.

Shor relaxation

For $H \in \mathbb{H}_n$ and $z \in \mathbb{C}^n$, the relationship $z^H H z = \text{Tr}(H z z^H)$ holds where $\text{Tr}(\cdot)$ denotes the trace¹ of a complex square matrix. Relaxing the rank of $Z = z z^H$ in (7.1) yields

$$\text{SDP-}\mathbb{C} : \inf_{Z \in \mathbb{H}_n} \text{Tr}(H_0 Z), \quad (7.2a)$$

$$\text{s.t. } \text{Tr}(H_i Z) \leq h_i, \quad i = 1, \dots, m, \quad (7.2b)$$

$$Z \succcurlyeq 0, \quad (7.2c)$$

where $\succcurlyeq 0$ indicates positive semidefiniteness.

Let $\text{Re}Z$ and $\text{Im}Z$ denote the real and imaginary parts of the matrix $Z \in \mathbb{C}^{n \times n}$ respectively. Consider the ring homomorphism $\Lambda : (\mathbb{C}^{n \times n}, +, \times) \rightarrow (\mathbb{R}^{2n \times 2n}, +, \times)$ defined by

$$\Lambda(Z) := \begin{pmatrix} \text{Re}Z & -\text{Im}Z \\ \text{Im}Z & \text{Re}Z \end{pmatrix}, \quad (7.3)$$

whose relevant properties are proven in Appendix A. To convert SDP- \mathbb{C} into real numbers, real and imaginary parts of the complex matrix variable are identified using two properties: (1) a complex matrix Z is positive semidefinite if and only if the real matrix $\Lambda(Z)$ is positive semidefinite, and (2) if $Z_1, Z_2 \in \mathbb{H}_n$, then $\text{Tr}[\Lambda(Z_1)\Lambda(Z_2)] = \text{Tr}[\Lambda(Z_1 Z_2)] = 2\text{Tr}(Z_1 Z_2)$. This yields the converted problem

$$\text{CSDP-}\mathbb{R} : \inf_{X \in \mathbb{S}_{2n}} \text{Tr}(\Lambda(H_0)X), \quad (7.4a)$$

¹For all matrices $A, B \in \mathbb{C}^{n \times n}$, $\text{Tr}(AB) = \sum_{1 \leq i, j \leq n} A_{ij} B_{ji}$.

$$\text{s.t. } \text{Tr}(\Lambda(H_i)X) \leq h_i, \quad i = 1, \dots, m, \quad (7.4b)$$

$$X \succcurlyeq 0, \quad (7.4c)$$

$$X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \quad \& \quad \begin{matrix} A & = & C, \\ B^T & = & -B, \end{matrix} \quad (7.4d)$$

where \mathbb{S}_{2n} denotes the set of real symmetric matrices of order $2n$ and $(\cdot)^T$ indicates the transpose. Note that the set of matrices satisfying (7.4d) is isomorphic to $\mathbb{C}^{n \times n}$. A global solution to QCQP-C can be retrieved from CSDP- \mathbb{R} if and only if $\text{rank}(X) \in \{0, 2\}$ at optimality (proof in Appendix B).

In order to convert QCQP-C into real numbers, real and imaginary parts of the complex vector variable are identified. This is done by considering a new variable $x = ((\text{Re}z)^T (\text{Im}z)^T)^T$ and observing that if $H \in \mathbb{H}_n$, then $z^H H z = x^T \Lambda(H) x = \text{Tr}(\Lambda(H) x x^T)$. This gives rise to a problem which we will call QCQP- \mathbb{R} . Relaxing the rank of $X = x x^T$ yields

$$\text{SDP-}\mathbb{R} : \quad \inf_{X \in \mathbb{S}_{2n}} \text{Tr}(\Lambda(H_0)X), \quad (7.5a)$$

$$\text{s.t. } \text{Tr}(\Lambda(H_i)X) \leq h_i, \quad i = 1, \dots, m, \quad (7.5b)$$

$$X \succcurlyeq 0. \quad (7.5c)$$

A global solution to QCQP-C can be retrieved from SDP- \mathbb{R} if and only if $\text{rank}(X) \in \{0, 1\}$ or $\text{rank}(X) = 2$ and (7.4d) holds at optimality.

We have $\text{val}(\text{SDP-C}) = \text{val}(\text{CSDP-}\mathbb{R}) = \text{val}(\text{SDP-}\mathbb{R})$ where ‘‘val’’ is the optimal value of a problem (proof in Appendix C). The number of scalar variables of CSDP- \mathbb{R} is half that of SDP- \mathbb{R} due to constraint (7.4d). This constraint also halves the possible ranks of the matrix variable, which must be an even integer in CSDP- \mathbb{R} whereas it can be any integer between 0 and $2n$ in SDP- \mathbb{R} . The number of variables in SDP- \mathbb{R} can be reduced by a small fraction ($\frac{2}{2n+1}$ to be exact) by setting a diagonal element of X to 0. This does not affect the optimal value (proof in Appendix D). Figure 7.1 summarizes this section.

Second-order conic relaxation

In SDP-C, assume that the semidefinite constraint (7.2c) is relaxed to the second-order cones

$$\begin{pmatrix} Z_{ii} & Z_{ij} \\ Z_{ij}^H & Z_{jj} \end{pmatrix} \succcurlyeq 0, \quad 1 \leq i \neq j \leq n. \quad (7.6)$$

Equation (7.6) is equivalent to constraining the determinant $Z_{ii}Z_{jj} - Z_{ij}Z_{ij}^H$ and diagonal elements Z_{ii} to be non-negative. This yields

$$\text{SOCP-C} : \quad \inf_{Z \in \mathbb{H}_n} \text{Tr}(H_0 Z), \quad (7.7a)$$

$$\text{s.t. } \text{Tr}(H_i Z) \leq h_i, \quad i = 1, \dots, m, \quad (7.7b)$$

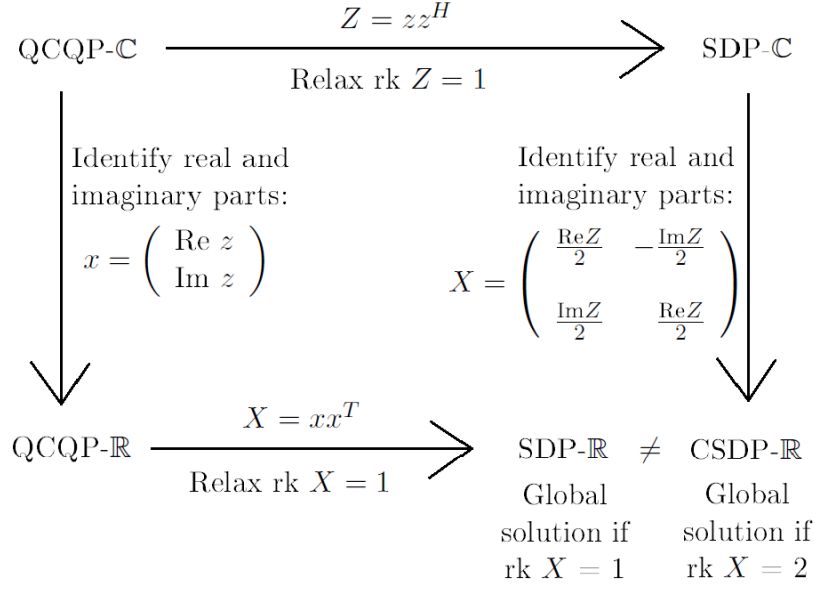


Figure 7.1: Non-Commutativity of Relaxation and Complex-to-Real Conversion

$$|Z_{ij}|^2 \leq Z_{ii}Z_{jj}, \quad 1 \leq i \neq j \leq n, \quad (7.7c)$$

$$Z_{ii} \geq 0, \quad i = 1, \dots, n, \quad (7.7d)$$

where $|\cdot|$ denotes the complex modulus. Identifying real and imaginary parts of the matrix variable Z leads to

$$\text{CSOCP-}\mathbb{R} : \inf_{X \in \mathbb{S}_{2n}} \text{Tr}(\Lambda(H_0)X), \quad (7.8a)$$

$$\text{s.t. } \text{Tr}(\Lambda(H_i)X) \leq h_i, \quad i = 1, \dots, m, \quad (7.8b)$$

$$X_{ij}^2 + X_{n+i,j}^2 \leq X_{ii}X_{jj}, \quad 1 \leq i \neq j \leq n, \quad (7.8c)$$

$$X_{ii} + X_{n+i,n+i} \geq 0, \quad i = 1, \dots, n, \quad (7.8d)$$

$$X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \quad \& \quad \begin{matrix} A & = & C, \\ B^T & = & -B. \end{matrix} \quad (7.8e)$$

In SDP- \mathbb{R} of Section, assume that the semidefinite constraint (7.5c) is relaxed to the second-order cones

$$\begin{pmatrix} X_{ii} & X_{ij} \\ X_{ij} & X_{jj} \end{pmatrix} \succcurlyeq 0, \quad 1 \leq i \neq j \leq 2n. \quad (7.9)$$

This leads to

$$\text{SOCP-}\mathbb{R} : \inf_{X \in \mathbb{S}_{2n}} \text{Tr}(\Lambda(H_0)X), \quad (7.10a)$$

$$\text{s.t. } \text{Tr}(\Lambda(H_i)X) \leq h_i, \quad i = 1, \dots, m, \quad (7.10b)$$

$$X_{ij}^2 \leq X_{ii}X_{jj}, \quad 1 \leq i \neq j \leq 2n, \quad (7.10c)$$

$$X_{ii} \geq 0, \quad i = 1, \dots, 2n. \quad (7.10d)$$

Unlike in the Shor relaxation, we have $\text{val}(\text{SOCP-C}) = \text{val}(\text{CSOCP-}\mathbb{R}) \geq \text{val}(\text{SOCP-}\mathbb{R})$ (proof in Appendix E). The number of scalar variables of CSOCP- \mathbb{R} is half that of SOCP- \mathbb{R} due to constraint (7.8e). The number of second-order conic constraints in CSOCP- \mathbb{R} , equal to $\frac{n(n-1)}{2}$, is roughly a fourth of that in SOCP- \mathbb{R} , equal to $\frac{2n(2n-1)}{2}$.

Exploiting sparsity

Given an undirected graph $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} \subset \{1, \dots, n\}$ and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, define for all $Z \in \mathbb{H}_n$

$$\Psi_{(\mathcal{V}, \mathcal{E})}(Z)_{ij} := \begin{cases} Z_{ij} & \text{if } (i, j) \in \mathcal{E} \text{ or } i = j \in \mathcal{V}, \\ 0 & \text{else.} \end{cases} \quad (7.11)$$

We associate an undirected graph \mathcal{G} to QCQP-C whose nodes are $\{1, \dots, n\}$ and that satisfies $H_i = \Psi_{\mathcal{G}}(H_i)$ for $i = 0, \dots, m$. Let \mathbb{H}_n^+ denote the set of positive semidefinite Hermitian matrices of size n and let “Ker” denote the kernel of a linear application. Given the definition of \mathcal{G} , constraint (7.2c) of SDP-C can be relaxed to $Z \in \mathbb{H}_n^+ + \text{Ker } \Psi_{\tilde{\mathcal{G}}}$ without changing its optimal value for any graph $\tilde{\mathcal{G}}$ whose nodes are $\{1, \dots, n\}$ and where $\mathcal{G} \subset \tilde{\mathcal{G}}$. Consider a chordal extension $\mathcal{G} \subset \mathcal{G}^{\text{ch}}$, that is to say that all cycles of length four or more have a chord (edge between two non-consecutive nodes of the cycle). Let $\mathcal{C}_1, \dots, \mathcal{C}_p \subset \mathcal{G}^{\text{ch}}$ denote the maximal cliques of \mathcal{G}^{ch} . (A clique is a subgraph where all nodes are linked to one another. The set of maximally sized cliques of a given graph can be computed in linear time [117]). A chordal extension has a useful property for exploiting sparsity [46]: for all $Z \in \mathbb{H}_n$, we have that $Z \in \mathbb{H}_n^+ + \text{Ker } \Psi_{\mathcal{G}^{\text{ch}}}$ if and only if $\Psi_{\mathcal{C}_i}(Z) \succeq 0$ for $i = 1, \dots, p$. Note that $\Psi_{\mathcal{C}_i}(Z) \succeq 0$ if and only if $\Lambda \circ \Psi_{\mathcal{C}_i}(Z) \succeq 0$, where “ \circ ” is the composition of functions. Given a graph $(\mathcal{V}, \mathcal{E})$, define for $X \in \mathbb{S}_{2n}$

$$\tilde{\Psi}_{(\mathcal{V}, \mathcal{E})}(X) := \begin{pmatrix} \Psi_{(\mathcal{V}, \mathcal{E})}(A) & \Psi_{(\mathcal{V}, \mathcal{E})}(B^T) \\ \Psi_{(\mathcal{V}, \mathcal{E})}(B) & \Psi_{(\mathcal{V}, \mathcal{E})}(C) \end{pmatrix}, \quad (7.12)$$

using the block decomposition in the left hand part of (7.4d). Notice that $\Lambda \circ \Psi_{(\mathcal{V}, \mathcal{E})} = \tilde{\Psi}_{(\mathcal{V}, \mathcal{E})} \circ \Lambda$. As a result, (7.4c) can be replaced by $\tilde{\Psi}_{\mathcal{C}_i}(X) \succeq 0$ for $i = 1, \dots, p$ without changing the optimal value of CSDP- \mathbb{R} , with an analogous replacement for constraint (7.5c) in SDP- \mathbb{R} . If in SDP- \mathbb{R} we exploit the sparsity of matrices $\Lambda(H_i)$ instead of that of H_i , the resulting graph has twice as many nodes. Computing a chordal extension and maximal cliques is hence more costly.

Sparsity in the second-order conic relaxations is exploited using the fact that applying (7.8c) and (7.10c) only for (i, j) that are edges of \mathcal{G} does not change the optimal values of CSOCP- \mathbb{R} and SOCP- \mathbb{R} .

7.3 Complex moment/sum-of-squares hierarchy

We now transpose the work of Lasserre [62] from real to complex numbers. Let z^α denote the monomial $z_1^{\alpha_1} \cdots z_n^{\alpha_n}$ where $z \in \mathbb{C}^n$ and $\alpha \in \mathbb{N}^n$ for some integer $n \in \mathbb{N}^*$. Define $|\alpha| := \alpha_1 + \dots + \alpha_n$ and \bar{w} as the conjugate of $w \in \mathbb{C}$. Define $\bar{z} := (\bar{z}_1, \dots, \bar{z}_n)^T$ where $z \in \mathbb{C}^n$. Consider the sets

$$\begin{aligned} \mathbb{C}[z] &:= \{ p : \mathbb{C}^n \rightarrow \mathbb{C} \mid p(z) = \sum_{|\alpha| \leq l} p_\alpha z^\alpha, l \in \mathbb{N}, p_\alpha \in \mathbb{C} \}, \\ \mathbb{C}[\bar{z}, z] &:= \{ f : \mathbb{C}^n \rightarrow \mathbb{C} \mid f(z) = \sum_{|\alpha|, |\beta| \leq l} f_{\alpha, \beta} \bar{z}^\alpha z^\beta, l \in \mathbb{N}, f_{\alpha, \beta} \in \mathbb{C} \}, \\ \mathbb{R}[\bar{z}, z] &:= \{ f \in \mathbb{C}[\bar{z}, z] \mid \overline{f(z)} = f(z), \forall z \in \mathbb{C}^n \}, \\ \Sigma[z] &:= \{ \sigma : \mathbb{C}^n \rightarrow \mathbb{C} \mid \sigma = \sum_{j=1}^r |p_j|^2, r \in \mathbb{N}^*, p_j \in \mathbb{C}[z] \}, \end{aligned} \quad (7.13)$$

and for all $d \in \mathbb{N}$

$$\begin{aligned} \mathbb{C}_d[z] &:= \{ p : \mathbb{C}^n \rightarrow \mathbb{C} \mid p(z) = \sum_{|\alpha| \leq d} p_\alpha z^\alpha, p_\alpha \in \mathbb{C} \}, \\ \mathbb{C}_d[\bar{z}, z] &:= \{ f : \mathbb{C}^n \rightarrow \mathbb{C} \mid f(z) = \sum_{|\alpha|, |\beta| \leq d} f_{\alpha, \beta} \bar{z}^\alpha z^\beta, f_{\alpha, \beta} \in \mathbb{C} \}, \\ \mathbb{R}_d[\bar{z}, z] &:= \{ f \in \mathbb{C}_d[\bar{z}, z] \mid \overline{f(z)} = f(z), \forall z \in \mathbb{C}^n \}, \\ \Sigma_d[z] &:= \{ \sigma : \mathbb{C}^n \rightarrow \mathbb{C} \mid \sigma = \sum_{j=1}^r |p_j|^2, r \in \mathbb{N}^*, p_j \in \mathbb{C}_d[z] \}. \end{aligned} \quad (7.14)$$

Note that the coefficients of a function $f \in \mathbb{R}[\bar{z}, z]$ satisfy $\overline{f_{\alpha, \beta}} = f_{\beta, \alpha}$ for all $|\alpha|, |\beta| \leq l$ for some $l \in \mathbb{N}$. The set of complex polynomials $\mathbb{C}[\bar{z}, z]$ is a \mathbb{C} -algebra (i.e. commutative ring and vector space over \mathbb{C}) and the set of holomorphic polynomials $\mathbb{C}[z]$ is a subalgebra of it (i.e. subspace closed under sum and product). The set of real-valued complex polynomials $\mathbb{R}[\bar{z}, z]$ is an \mathbb{R} -algebra. The set of sums of squared moduli of holomorphic polynomials $\Sigma[z]$ and the set $\Sigma_d[z] \subset \mathbb{R}_d[z]$ are pointed cones (i.e. closed under multiplication by elements of \mathbb{R}_+) that are convex (i.e. $tu + (1-t)v$ with $0 \leq t \leq 1$ belongs to them if u and v do). Let $C(K, \mathbb{C})$ denote the Banach (i.e. complete) \mathbb{C} -algebra of continuous functions from a compact set $K \subset \mathbb{C}^n$ to \mathbb{C} equipped with the norm $\|\varphi\|_\infty := \sup_{z \in K} |\varphi(z)|$. Consider $R_K : \mathbb{C}[\bar{z}, z] \rightarrow C(K, \mathbb{C})$ defined by $f \mapsto f|_K$ where $f|_K$ denotes the restriction of f to K . $R_K(\mathbb{C}[\bar{z}, z])$ is a unital subalgebra of $C(K, \mathbb{C})$ (i.e. contains multiplicative unit) that separates points of K (i.e. $u \neq v \in K \implies \exists \varphi \in R_K(\mathbb{C}[\bar{z}, z]) : \varphi(u) \neq \varphi(v)$) and that is closed under complex conjugation. It is hence a dense subalgebra due to the Complex Stone-Weierstrass Theorem. Likewise, $C(K, \mathbb{R}) := \{\varphi \in C(K, \mathbb{C}) \mid \overline{\varphi(z)} = \varphi(z), \forall z \in \mathbb{C}^n\}$ is a Banach \mathbb{R} -algebra of which $R_K(\mathbb{R}[\bar{z}, z])$ is a dense subalgebra. In other words, a continuous real-valued function of multiple complex variables can be approximated as close as desired by real-valued complex polynomials when restricted to a compact set. They are hence a powerful modeling tool in optimization. Speaking of which, let $m \in \mathbb{N}^*$ and $k, k_1, \dots, k_m \in \mathbb{N}$. Consider $(f, g_1, \dots, g_m) \in \mathbb{R}_k[\bar{z}, z] \times \mathbb{R}_{k_1}[\bar{z}, z] \times \dots \times \mathbb{R}_{k_m}[\bar{z}, z]$ where there exists $|\alpha| = k$ and $|\beta| \leq k$ such that $f_{\alpha, \beta} \neq 0$. In addition, for $i = 1, \dots, m$, there exists $|\alpha| = k_i$ and $|\beta| \leq k_i$ such that $g_{i, \alpha, \beta} \neq 0$. Consider the complex multivariate polynomial optimization problem

$$\boxed{f^{\text{opt}} := \inf_{z \in \mathbb{C}^n} f(z) \quad \text{s.t.} \quad g_i(z) \geq 0, \quad i = 1, \dots, m,} \quad (7.15)$$

where by convention $f^{\text{opt}} := +\infty$ if the feasible set is empty. The feasible set is a closed semi-algebraic set on which we make the following assumption from now on:

$$\boxed{K := \{ z \in \mathbb{C}^n \mid g_i(z) \geq 0, i = 1, \dots, m \} \text{ is compact.}} \quad (7.16)$$

Let K^{opt} denote the set of optimal solutions to (7.15). It may be empty because we do not assume K to be non-empty. (Note that in practice, it is often hard to know whether there exists a feasible solution, as for the application of Section 7.4.)

Let $\mathcal{M}(K)$ denote the Banach space over \mathbb{R} of Radon measures on K . Bear in mind that since K is compact, $\mathcal{M}(K)$ may be identified with the topological dual of $C(K, \mathbb{R})$ i.e. the Banach space over \mathbb{R} of linear continuous applications from $C(K, \mathbb{R})$ to \mathbb{R} equipped with the operator norm. (This is due to the Riesz-Markov-Kakutani Representation Theorem.) For $\varphi \in C(K, \mathbb{C})$, define $\int_K \varphi d\mu := \int_K \text{Re}(\varphi) d\mu + \mathbf{i} \int_K \text{Im}(\varphi) d\mu$ [100, 1.31 Definition]². Next, consider the convex pointed cone $\mathcal{P}(K) := \{ \varphi \in C(K, \mathbb{R}) \mid \varphi(z) \geq 0, \forall z \in K \}$. A Radon measure μ is positive (denoted $\mu \geq 0$) if $\varphi \in \mathcal{P}(K)$ implies that $\int_K \varphi d\mu \geq 0$. Let $\mathcal{M}_+(K)$ denote the set of positive Radon measures. With these definitions, we have

$$f^{\text{opt}} = \inf_{\mu \in \mathcal{M}(K)} \int_K f d\mu \quad \text{s.t.} \quad \int_K d\mu = 1 \quad \& \quad \mu \geq 0. \quad (7.17)$$

Indeed, if $z \in K$, then the Dirac³ measure δ_z is a feasible point of (7.17) for which the objective value is equal to $f(z)$. Hence the optimal value of (7.17) is less than or equal to f^{opt} . Conversely, if μ is a feasible point of (7.17), then $\int_K (f - f^{\text{opt}}) d\mu \geq 0$ and hence $\int_K f d\mu \geq \int_K f^{\text{opt}} d\mu = f^{\text{opt}} \int_K d\mu = f^{\text{opt}}$.

Proposition 7.1. *The set of optimal solutions to (7.17) is*

$$\{ \mu \in \mathcal{M}_+(K) \mid \mu(K^{\text{opt}}) = 1 \quad \& \quad \mu(K \setminus K^{\text{opt}}) = 0 \}. \quad (7.18)$$

As a consequence, if K^{opt} is a finite set of $S \in \mathbb{N}^$ points $z(1), \dots, z(S) \in \mathbb{C}^n$, then the set of optimal solutions to (7.17) is $\{ \sum_{j=1}^S \lambda_j \delta_{z(j)} \mid \sum_{j=1}^S \lambda_j = 1 \quad \& \quad \lambda_1, \dots, \lambda_S \in \mathbb{R}_+ \}$.*

Proof. Consider μ an optimal solution to (7.17). It must be that $\int_K (f - f^{\text{opt}}) d\mu = 0$. Thus $\int_{K \setminus K^{\text{opt}}} (f - f^{\text{opt}}) d\mu = 0$ and $\mu(K \setminus K^{\text{opt}}) = \int_{K \setminus K^{\text{opt}}} d\mu = 0$. Therefore $\mu(K^{\text{opt}}) = \int_{K^{\text{opt}}} d\mu = \mu(K) - \mu(K \setminus K^{\text{opt}}) = 1$. Conversely, if μ belongs to the set in (7.18), then it is feasible for (7.17) and $\int_K (f - f^{\text{opt}}) d\mu = \int_{K \setminus K^{\text{opt}}} (f - f^{\text{opt}}) d\mu = 0$. Hence $\int_K f d\mu = \int_K f^{\text{opt}} d\mu = f^{\text{opt}} \int_K d\mu = f^{\text{opt}}$. \square

In order to dualize the equality constraint in (7.17), consider the Lagrange function $\mathcal{L} : \mathcal{M}_+(K) \times \mathbb{R} \rightarrow \mathbb{R}$ defined by $(\mu, \lambda) \mapsto \int_K f d\mu + \lambda (1 - \int_K d\mu)$. We have $\mathcal{L}(\mu, \lambda) =$

²We wish to thank Bruno Nazaret for bringing this reference to our attention.

³The Dirac measure δ_z with $z \in K$ may be identified with the continuous linear application from $C(K, \mathbb{R})$ to \mathbb{R} defined by $\varphi \mapsto \varphi(z)$. This is one way to interpret the fact that $\int_K f d\delta_z = f(z)$.

$\lambda + \int_K (f - \lambda) d\mu$ and

$$\inf_{\mu \in \mathcal{M}_+(K)} \int_K (f - \lambda) d\mu = \begin{cases} 0 & \text{if } f(z) - \lambda \geq 0, \quad \forall z \in K, \\ -\infty & \text{else,} \end{cases} \quad (7.19)$$

since, in the second case, we may consider $t\delta_z$ for a $z \in K$ such that $f(z) - \lambda < 0$ and $t \rightarrow +\infty$. This leads to the dual problem

$$f^{\text{opt}} = \sup_{\lambda \in \mathbb{R}} \lambda \quad \text{s.t.} \quad f(z) - \lambda \geq 0, \quad \forall z \in K. \quad (7.20)$$

Primal problem (7.17) gives rise to the complex moment hierarchy below. Dual problem (7.20) gives rise to the complex sum-of-squares hierarchy below.

Complex moment hierarchy

Let \mathcal{H} (respectively \mathcal{H}_d) denote the set of sequences of complex numbers $(y_{\alpha,\beta})_{\alpha,\beta \in \mathbb{N}^n}$ (respectively $(y_{\alpha,\beta})_{|\alpha|,|\beta| \leq d}$) such that $\overline{y_{\alpha,\beta}} = y_{\beta,\alpha}$ for all $\alpha, \beta \in \mathbb{N}^n$ (respectively $|\alpha|, |\beta| \leq d$).

Definition 7.2. *An element $y \in \mathcal{H}$ is said to have a representing measure μ on K if $\mu \in \mathcal{M}_+(K)$ and $y_{\alpha,\beta} = \int_K \bar{z}^\alpha z^\beta d\mu$ for all $\alpha, \beta \in \mathbb{N}^n$. In that case, $y_{\alpha,\beta}$ is called the (α, β) -moment of μ .*

When $y \in \mathcal{H}$ has a representing measure on K , the measure is unique because $R_K(\mathbb{C}[\bar{z}, z])$ is dense in $C(K, \mathbb{C})$. The complex moment problem consists in characterizing the sequences that are representable by a measure on K and is connected to other branches of mathematics such as functional analysis and spectral theory of operators [4]. It has been studied by Atzmon [11], Schmüdgen [102], Putinar [93], Curto and Fialkow [32–34], Stochel [114], and Vasilescu [124]. For example, Atzmon [11, Theorem 2.1] proved that the solutions to the complex moment problem where $K = \{z \in \mathbb{C} \mid |z| = 1\}$ are the sequences $y \in \mathcal{H}$ such that $\sum_{m,n,j,k \in \mathbb{N}} c_{n,j} \bar{c}_{m,k} y_{m+j,n+k} \geq 0$ and $\sum_{m,n \in \mathbb{N}} w_m \bar{w}_n (y_{m,n} - y_{m+1,n+1}) \geq 0$ for all complex numbers $(c_{j,k})_{j,k \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$ with only finitely many non-zero terms. A generalization to the multidimensional case is considered in Section 7.3. We conclude our presentation of the complex moment problem by noting that the case where K is not compact is an open problem.

Consider a feasible point μ of (7.17) and the sequence $y \in \mathcal{H}$ that has representation measure μ on K . Notice that $\int_K f d\mu = \int_K \sum_{|\alpha|,|\beta| \leq k} f_{\alpha,\beta} \bar{z}^\alpha z^\beta d\mu = \sum_{|\alpha|,|\beta| \leq k} f_{\alpha,\beta} \int_K \bar{z}^\alpha z^\beta d\mu = \sum_{|\alpha|,|\beta| \leq k} f_{\alpha,\beta} y_{\alpha,\beta} =: L_y(f)$ and $\int_K d\mu = \int_K \bar{z}^0 z^0 d\mu = y_{0,0} = 1$. For all $p \in \mathbb{C}[z]$, we have $|p|^2 g_i \geq 0$ on K . Since $\mu \geq 0$, this implies that $\int_K |p|^2 g_i d\mu \geq 0$. Naturally, we also have $\int_K |p|^2 g_0 d\mu \geq 0$ if we define $g_0 := 1$. Define $k_0 := 0$ and $d^{\min} := \max\{k, k_1, \dots, k_m\}$. Consider

$d \geq d^{\min}$, $0 \leq i \leq m$, and $p \in \mathbb{C}_{d-k_i}[z]$. We have $\int_K |p|^2 g_i d\mu = \dots$

$$\begin{aligned}
&= \int_K \left| \sum_{|\alpha| \leq d-k_i} p_\alpha z^\alpha \right|^2 \left(\sum_{|\gamma|, |\delta| \leq k_i} g_{i,\gamma,\delta} \bar{z}^\gamma z^\delta \right) d\mu \\
&= \int_K \left(\sum_{|\alpha|, |\beta| \leq d-k_i} \bar{p}_\alpha p_\beta \bar{z}^\alpha z^\beta \right) \left(\sum_{|\gamma|, |\delta| \leq k_i} g_{i,\gamma,\delta} \bar{z}^\gamma z^\delta \right) d\mu \\
&= \int_K \sum_{|\alpha|, |\beta| \leq d-k_i} \bar{p}_\alpha p_\beta \sum_{|\gamma|, |\delta| \leq k_i} g_{i,\gamma,\delta} \bar{z}^{\alpha+\gamma} z^{\beta+\delta} d\mu \\
&= \sum_{|\alpha|, |\beta| \leq d-k_i} \bar{p}_\alpha p_\beta \sum_{|\gamma|, |\delta| \leq k_i} g_{i,\gamma,\delta} \int_K \bar{z}^{\alpha+\gamma} z^{\beta+\delta} d\mu \\
&= \sum_{|\alpha|, |\beta| \leq d-k_i} \bar{p}_\alpha p_\beta \left(\sum_{|\gamma|, |\delta| \leq k_i} g_{i,\gamma,\delta} y_{\alpha+\gamma, \beta+\delta} \right) =: M_{d-k_i}(g_i y)(\alpha, \beta) \\
&= \sum_{|\alpha|, |\beta| \leq d-k_i} \bar{p}_\alpha p_\beta M_{d-k_i}(g_i y)(\alpha, \beta) \\
&= \vec{p}^H M_{d-k_i}(g_i y) \vec{p},
\end{aligned} \tag{7.21}$$

where $\vec{p} := (p_\alpha)_{|\alpha| \leq d-k_i}$ and $M_{d-k_i}(g_i y)$ is a Hermitian matrix indexed by $|\alpha|, |\beta| \leq d-k_i$. As a result

$$M_{d-k_i}(g_i y) \succcurlyeq 0, \quad i = 0, \dots, m, \quad \forall d \geq d^{\min}. \tag{7.22}$$

To sum up, y is a feasible point of

$$\begin{aligned}
\rho &:= \inf_{y \in \mathcal{H}} L_y(f), \\
&\text{s.t.} \quad y_{0,0} = 1, \\
&\quad M_{d-k_i}(g_i y) \succcurlyeq 0, \quad i = 0, \dots, m, \quad \forall d \geq d^{\min},
\end{aligned} \tag{7.23}$$

with same objective value as μ in (7.17). Automatically, $\rho \leq f^{\text{opt}}$. Consider the relaxation of (7.23) defined by

$$\boxed{
\begin{aligned}
\rho_d &:= \inf_{y \in \mathcal{H}_d} L_y(f), \\
&\text{s.t.} \quad y_{0,0} = 1, \\
&\quad M_{d-k_i}(g_i y) \succcurlyeq 0, \quad i = 0, \dots, m,
\end{aligned}
} \tag{7.24}$$

which we name the *complex moment relaxation of order d* for reasons that will become clear with Theorem 7.11. In Section 7.3, we will introduce its dual counterpart.

Remark 7.3. Given $y \in \mathcal{H}$, the function L_y in this section can be formally be defined by the \mathbb{C} -linear operator $L_y : \mathbb{C}[\bar{z}, z] \rightarrow \mathbb{C}$ such that $L_y(\bar{z}^\alpha z^\beta) = y_{\alpha,\beta}$ for all $\alpha, \beta \in \mathbb{N}$. If $\varphi \in \mathbb{C}[\bar{z}, z]$ and $\bar{\varphi} = \varphi$, then $\overline{L_y(\varphi)} = L_y(\varphi)$. Given $l, d \in \mathbb{N}$ and $\varphi \in \mathbb{R}_l[\bar{z}, z]$, the matrix M_d in (7.21) can be formally be defined as the Hermitian matrix indexed by $|\alpha|, |\beta| \leq d$ such that $M_d(\varphi y)(\alpha, \beta) := L_y(\varphi(z) \bar{z}^\alpha z^\beta) = \sum_{|\gamma|, |\delta| \leq l} \varphi_{\gamma,\delta} y_{\alpha+\gamma, \beta+\delta}$. Notice that $M_d(\varphi y)(0, 0) = L_y(\varphi)$. Lastly, define $M_d(y) := M_d(g_0 y)$ which we refer to as *complex moment matrix of order d* .

Complex sum-of-squares hierarchy

We introduced notation \vec{p} for $p \in \mathbb{C}_d[z]$ where $d \in \mathbb{N}$ and will now extend it to $\sigma \in \Sigma_d[z]$. For such an element, there exists $r \in \mathbb{N}^*$ and $p_j \in \mathbb{C}_d[z]$ such that $\sigma = \sum_{j=1}^r |p_j|^2$. Let $\vec{\sigma} := \sum_{j=1}^r \vec{p}_j \vec{p}_j^H$. Also, define $\langle A, B \rangle_{\mathcal{H}_d} := \text{Tr}(AB)$ where $A, B \in \mathcal{H}_d$. Given $d \geq$

d^{\min} , consider the Lagrange function $\mathcal{L}_d : \mathcal{H}_d \times \mathbb{R} \times \Sigma_{d-k_0}[z] \times \dots \times \Sigma_{d-k_m}[z] \rightarrow \mathbb{R}$ defined by $(y, \lambda, \sigma_0, \dots, \sigma_m) \mapsto L_y(f) + \lambda(1 - y_{0,0}) - \sum_{i=0}^m \langle M_{d-k_i}(g_i y), \vec{\sigma}_i \rangle_{\mathcal{H}_{d-k_i}}$. Compute $\mathcal{L}_d(y, \lambda, \sigma_0, \dots, \sigma_m) = \lambda + L_y(f - \lambda) - \sum_{i=0}^m \sum_{j=0}^{r_i} (\vec{p}_j^i)^H M_{d-k_i}(g_i y) \vec{p}_j^i = \lambda + L_y(f - \lambda) - \sum_{i=0}^m \sum_{j=0}^{r_i} L_y(|p_j^i|^2 g_i) = \lambda + L_y(f - \lambda - \sum_{i=0}^m \sigma_i g_i)$. Observe that

$$\inf_{y \in \mathcal{H}} L_y \left(f - \lambda - \sum_{i=0}^m \sigma_i g_i \right) = \begin{cases} 0 & \text{if } f(z) - \lambda - \sum_{i=0}^m \sigma_i(z) g_i(z) = 0, \\ & \text{for all } z \in \mathbb{C}^n, \\ -\infty & \text{else.} \end{cases} \quad (7.25)$$

Indeed, in the second case, there exists $z \in \mathbb{C}^n$ such that $f(z) - \lambda - \sum_{i=0}^m \sigma_i(z) g_i(z) \neq 0$. With $(y_{\alpha, \beta})_{\alpha, \beta \in \mathbb{N}} := (\bar{z}^\alpha z^\beta)_{\alpha, \beta \in \mathbb{N}}$, $L_{ty}(f - \lambda - \sum_{i=0}^m \sigma_i g_i) \rightarrow -\infty$ for either $t \rightarrow -\infty$ or $t \rightarrow +\infty$. The associated dual problem of (7.24) is thus

$$\boxed{\begin{aligned} \rho_d^* &:= \sup_{\lambda, \sigma} \lambda, \\ \text{s.t.} & \quad f - \lambda = \sum_{i=0}^m \sigma_i g_i, \\ & \quad \lambda \in \mathbb{R}, \sigma_i \in \Sigma_{d-k_i}[z], i = 0, \dots, m, \end{aligned}} \quad (7.26)$$

which we name the *complex sum-of-squares relaxation of order d* . Consider

$$\begin{aligned} \rho^* &:= \sup_{\lambda, \sigma} \lambda, \\ \text{s.t.} & \quad f - \lambda = \sum_{i=0}^m \sigma_i g_i, \\ & \quad \lambda \in \mathbb{R}, \sigma_i \in \Sigma[z], i = 0, \dots, m, \end{aligned} \quad (7.27)$$

whose relationship with (7.23) is touched upon in Proposition 7.4 below.

Proposition 7.4. *We have $\rho_d^* \leq \rho_d$ for all $d \geq d^{\min}$ and $\rho_d^* \rightarrow \rho^* \leq \rho \leq f^{\text{opt}}$.*

Proof. The sequence $(\rho_d^*)_{d \geq d^{\min}}$ is non-decreasing and upper bounded by $\rho^* \in \mathbb{R} \cup \{\pm\infty\}$. Thus it converges towards some limit $\rho_{\text{lim}}^* \in \mathbb{R} \cup \{\pm\infty\}$ such that $\rho_{\text{lim}}^* \leq \rho^*$. If $\rho^* = -\infty$, then $\rho_d^* = -\infty$ for all $d \geq d^{\min}$ and $\rho_d^* \rightarrow \rho^*$. If not, by definition of the optimum ρ^* , there exists a sequence $(\lambda^l, \sigma_0^l, \dots, \sigma_m^l)$ of feasible points such that $\lambda^l \leq \rho^*$ and $\lambda^l \rightarrow \rho^*$. To each $l \in \mathbb{N}$, we may associate an integer $d(l) \in \mathbb{N}$ such that $(\lambda^l, \sigma_0^l, \dots, \sigma_m^l)$ is a feasible point of the complex sum-of-squares relaxation of order $d(l)$. Thus $\lambda^l \leq \rho_{d(l)}^* \leq \rho^*$. As a result, $\rho_{\text{limit}}^* = \rho^*$. Moreover, $(\rho_d)_{d \geq d^{\min}}$ is non-decreasing and upper bounded by $\rho \in \mathbb{R} \cup \{\pm\infty\}$. Thus it converges towards some limit $\rho_{\text{lim}} \in \mathbb{R} \cup \{\pm\infty\}$ such that $\rho_{\text{lim}} \leq \rho$. Moreover, weak duality implies that $\rho_d^* \leq \rho_d (\leq \rho)$. Thus $\rho^* \leq \rho_{\text{lim}} \leq \rho$. It was already shown that $\rho \leq f^{\text{opt}}$. \square

Remark 7.5. Problems (7.27) and (7.23) may be interpreted as a pair of primal-dual linear programs in infinite-dimensional spaces [8]. Indeed, consider the duality bracket $\langle \cdot, \cdot \rangle$ defined from $\mathbb{R}[\bar{z}, z] \times \mathcal{H}$ to \mathbb{R} by $\langle \varphi, y \rangle := L_y(\varphi)$. A sequence $(\varphi^n)_{n \in \mathbb{N}}$ in $\mathbb{R}[\bar{z}, z]$ is said to converge weakly towards $\varphi \in \mathbb{R}[\bar{z}, z]$ if for all $y \in \mathcal{H}$, we have $\langle \varphi^n, y \rangle \rightarrow \langle \varphi, y \rangle$. Consider the weakly

continuous \mathbb{R} -linear operator $A : \mathbb{R}[\bar{z}, z] \rightarrow \mathbb{R}[\bar{z}, z]$ defined by $\varphi \mapsto \varphi - \varphi_{0,0}$. Its dual $A^* : \mathcal{H} \rightarrow \mathcal{H}$ is defined by $y \mapsto y - y_{0,0}\delta_{0,0}$ where $(\delta_{0,0})_{0,0} = 1$ and $(\delta_{0,0})_{\alpha,\beta} = 0$ if $(\alpha, \beta) \neq (0, 0)$. Indeed, $\langle A\varphi, y \rangle = \langle \varphi, A^*y \rangle$ for all $(\varphi, y) \in \mathbb{R}[\bar{z}, z] \times \mathcal{H}$. Consider the convex pointed cone defined by $C := \Sigma[z]g_0 + \dots + \Sigma[z]g_m$ and its dual cone $C^* := \{y \in \mathcal{H} \mid \forall \varphi \in C, \langle \varphi, y \rangle \geq 0\}$. With $b := Af$, notice that

$$\begin{aligned} f_{0,0} - \rho^* &= \inf_{\varphi \in \mathbb{R}[\bar{z}, z]} \langle \varphi, \delta_{0,0} \rangle \quad \text{s.t.} \quad A\varphi = b \ \& \ \varphi \in C, \\ f_{0,0} - \rho &= \sup_{y \in \mathcal{H}} \langle b, y \rangle \quad \text{s.t.} \quad \delta_{0,0} - A^*y \in C^*. \end{aligned} \quad (7.28)$$

Let $\text{cl}(C)$ denote the weak closure of C in $\mathbb{R}[\bar{z}, z]$. According to [5, 5.91 Bipolar Theorem]⁴, we have $\text{cl}(C) = C^{**}$. In the next section, Theorem 7.6 and Theorem 7.11 provide a sufficient condition ensuring no duality gap in (7.28) and $\text{cl}(C) = \{\varphi \in \mathbb{R}[\bar{z}, z] \mid \varphi|_K \geq 0\}$ respectively.

Convergence of the complex hierarchy

We now turn our attention to a result from algebraic geometry discovered in 2008.

Theorem 7.6 (D'Angelo's and Putinar's Positivstellensatz [37]). *If one of the constraints that define K in (7.16) is a sphere constraint $|z_1|^2 + \dots + |z_n|^2 = 1$, then*

$$f|_K > 0 \quad \implies \quad \exists \sigma_0, \dots, \sigma_m \in \Sigma[z] : \quad f = \sum_{i=0}^m \sigma_i g_i. \quad (7.29)$$

Proof. D'Angelo and Putinar wrote the theorem slightly differently so we provide an explanation. Say that constraints g_{m-1} and g_m are such that $g_{m-1} = s$ and $g_m = -s$ where $s(z) := 1 - |z_1|^2 - \dots - |z_n|^2$. With the assumptions of Theorem 7.6, the authors of [37, Theorem 3.1] show that there exists $\sigma_0, \dots, \sigma_{m-2} \in \Sigma[z]$ and $r \in \mathbb{R}[\bar{z}, z]$ such that $f(z) = \sum_{i=0}^{m-2} \sigma_i(z)g_i(z) + r(z)s(z)$ for all $z \in \mathbb{C}^n$. Thanks to [36, Proposition 1.2], there exists $\sigma_{m-1}, \sigma_m \in \Sigma[z]$ such that $r = \sigma_{m-1} - \sigma_m$ hence the desired result. \square

Theorem 7.6 can easily be generalized to any sphere $|z_1|^2 + \dots + |z_n|^2 = R^2$ of radius $R > 0$. With scaled variable $w = \frac{z}{R} \in \mathbb{C}^n$, the sphere constraint has radius 1 and a monomial of (7.15) with coefficient $c_{\alpha,\beta} \in \mathbb{C}$ reads $c_{\alpha,\beta} \bar{z}^\alpha z^\beta = c_{\alpha,\beta} (R\bar{w})^\alpha (Rw)^\beta = R^{|\alpha|+|\beta|} c_{\alpha,\beta} \bar{w}^\alpha w^\beta$. With the scaled coefficients $R^{|\alpha|+|\beta|} c_{\alpha,\beta}$, Theorem 7.6 can then be applied. Reverting back to the old scale $z = Rw$ in (7.29) leads to the desired result. Accordingly, we define the following statement which we will consider true only when explicitly stated:

Sphere Assumption:

One of the constraints of polynomial optimization problem (7.15) is a sphere constraint $|z_1|^2 + \dots + |z_n|^2 = R^2$ for some radius $R > 0$. (7.30)

Under the sphere assumption, K is compact so assumption (7.16) holds.

⁴We wish to thank Jean-Bernard Baillon for bringing this reference to our attention.

Corollary 7.7. Under the sphere assumption (7.30), $\rho_d^* \rightarrow f^{\text{opt}}$ and $\rho_d \rightarrow f^{\text{opt}}$.

Proof. Theorem 7.6 implies that $\rho^* = f^{\text{opt}}$ because for all $\epsilon > 0$, function $f - (f^{\text{opt}} - \epsilon)$ is positive on K . The sequences $(\rho_d^*)_{d \geq d^{\text{min}}}$ and $(\rho_d)_{d \geq d^{\text{min}}}$ converge towards f^{opt} due to Proposition 7.4. \square

To require a sphere constraint in a complex polynomial optimization problem seems very restrictive and irrelevant for many problems. But in fact, a sphere constraint can be applied to any complex polynomial optimization problem (7.15) with a feasible set contained in a ball $|z_1|^2 + \dots + |z_n|^2 \leq R^2$ of known radius $R > 0$. Indeed, simply add a slack variable $z_{n+1} \in \mathbb{C}$ and the constraint

$$\boxed{|z_1|^2 + \dots + |z_{n+1}|^2 = R^2.} \quad (7.31)$$

Let \hat{K} denote the feasible set of the problem in $n + 1$ variables. If $(z_1, \dots, z_{n+1}) \in \hat{K}$, then $(z_1, \dots, z_n) \in K$ and has the same objective value. Conversely, if $(z_1, \dots, z_n) \in K$, then $(z_1, \dots, z_{n+1}) \in \hat{K}$ for all $z_{n+1} \in \mathbb{C}$ such that $|z_{n+1}|^2 = R^2 - |z_1|^2 - \dots - |z_n|^2$. Again, the objective value is unchanged. To ensure a bijection between K and \hat{K} , add yet two more constraints $\mathbf{i}z_{n+1} - \mathbf{i}\bar{z}_{n+1} = 0$ and $z_{n+1} + \bar{z}_{n+1} \geq 0$, thereby preserving the number of global solutions. In that case, the application from K to \hat{K} defined by $(z_1, \dots, z_n) \mapsto (z_1, \dots, z_n, \sqrt{R^2 - |z_1|^2 - \dots - |z_n|^2})$ is a bijection. Adding the two extra constraints is optional and not required for convergence of optimal values.

As seen in Theorem 7.6, an equality constraint may be enforced via two opposite inequality constraints. Let h_1, \dots, h_e denote $e \in \mathbb{N}^*$ equality constraints in polynomial optimization problem (7.15). Putinar and Scheiderer [95, Propositions 6.6 and 3.2 (iii)] show that the sphere assumption in D'Angelo's and Putinar's Positivstellensatz may be weakened to the existence of $r_1, \dots, r_e \in \mathbb{R}[\bar{z}, z]$, $\sigma \in \Sigma[z]$, and $a \in \mathbb{R}$ such that

$$\sum_{j=1}^e r_j(z) h_j(z) = \sum_{i=1}^n |z_i|^2 + \sigma(z) + a, \quad \forall z \in \mathbb{C}^n. \quad (7.32)$$

If a problem contains the constraints $|z_1|^2 - 1 = \dots = |z_n|^2 - 1 = 0$, then the assumption is satisfied by $r_1 = \dots = r_n = 1$, $\sigma = 0$ and $a = -n$. In particular, there is no need to add a slack variable in the non-bipartite Grothendieck problem over the complex numbers [14].

Example 7.8. D'Angelo and Putinar [37] consider $\frac{1}{3} < a < \frac{4}{9}$ and problem

$$\begin{aligned} \inf_{z \in \mathbb{C}} f(z) &:= 1 - \frac{4}{3}|z|^2 + a|z|^4, \\ \text{s.t. } g(z) &:= 1 - |z|^2 \geq 0, \end{aligned} \quad (7.33)$$

whose set of global solutions is $K^{\text{opt}} = \{z \in \mathbb{C} \mid |z| = 1\}$ and $f^{\text{opt}} = a - \frac{1}{3} > 0$. They prove that the decomposition $f = \sigma_0 + \sigma_1 g$ ($\sigma_0, \sigma_1 \in \Sigma[z]$) of Theorem 7.6 does not hold. As a result, the optimal values of the complex sum-of-squares relaxations cannot exceed 0 even though $f^{\text{opt}} > 0$. Indeed, if $\rho_d^* > 0$ for some order $d \geq d^{\text{min}}$, then there exists $\lambda \geq \frac{\rho_d^*}{2}$ and

$\sigma_0, \sigma_1 \in \Sigma_d[z]$ such that $f - \lambda = \sigma_0 + \sigma_1 g$. Thus $f = \lambda + \sigma_0 + \sigma_1 g$ where $\lambda + \sigma_0 \in \Sigma_d[z]$, which is a contradiction. We suggest solving

$$\begin{aligned} \inf_{z_1, z_2 \in \mathbb{C}} \quad & \hat{f}(z_1, z_2) := 1 - \frac{4}{3}|z_1|^2 + a|z_1|^4, \\ \text{s.t.} \quad & \hat{g}(z_1, z_2) := 1 - |z_1|^2 - |z_2|^2 = 0, \end{aligned} \quad (7.34)$$

for which the decomposition of Theorem 7.6 holds thereby ensuring convergence of the complex moment/sum-of-squares hierarchy (Corollary 7.7). In other words, for all $\lambda < f^{\text{opt}}$ there exists $\hat{\sigma}_0 \in \Sigma[z_1, z_2]$ and $\hat{r} \in \mathbb{R}[\bar{z}_1, \bar{z}_2, z_1, z_2]$ such that

$$\hat{f}(z_1, z_2) - \lambda = \hat{\sigma}_0(z_1, z_2) + \hat{r}(z_1, z_2)\hat{g}(z_1, z_2), \quad \forall z_1, z_2 \in \mathbb{C}. \quad (7.35)$$

Plug in $z_1 = z$ and $z_2 = 0$ and obtain $f(z) - \lambda = \hat{\sigma}_0(z, 0) + \hat{r}(z, 0)g(z)$ for all $z \in \mathbb{C}$. While function $z \mapsto \hat{\sigma}_0(z, 0)$ belongs to $\Sigma[z]$, function $z \mapsto \hat{r}(z, 0)$ does not! Hence we do not contradict the fact that $f = \sigma_0 + \sigma_1 g$ ($\sigma_0, \sigma_1 \in \Sigma[z]$) is impossible. Consider $a = \frac{1}{2}(\frac{1}{3} + \frac{4}{9}) = \frac{7}{18}$ so that $f^{\text{opt}} = \frac{1}{18}$. Notice that $d^{\text{min}} = 2$ for (7.33) and (7.34). The complex relaxations of orders $2 \leq d \leq 3$ of (7.33) both yield⁵ the value -0.3333 . The complex relaxation of order 2 of (7.34) yields the value 0.0556 ($\approx f^{\text{opt}}$) and optimal polynomials $\hat{\sigma}_0(z_1, z_2) = 0.2780|z_2|^2 + 0.2776|z_1 z_2|^2 + 0.6667|z_2|^4$ and $\hat{r}(z_1, z_2) = 0.9444 - 0.3889|z_1|^2 + 0.6665|z_2|^2$, all of which satisfy (7.35).

Example 7.9. Putinar and Scheiderer [96] consider parameters $0 < a < \frac{1}{2}$ and $C > \frac{1}{1-2a}$, and problem

$$\begin{aligned} \inf_{z \in \mathbb{C}} \quad & f(z) := C - |z|^2, \\ \text{s.t.} \quad & g(z) := |z|^2 - az^2 - a\bar{z}^2 - 1 = 0, \end{aligned} \quad (7.36)$$

whose set of global solutions is $K^{\text{opt}} = \left\{ \pm \frac{1}{\sqrt{1-2a}} \right\}$ and $f^{\text{opt}} = C - \frac{1}{1-2a} > 0$. They prove that the decomposition of Theorem 7.6 does not hold. Since the feasible set is included in the Euclidean ball of radius \sqrt{C} , we suggest solving

$$\begin{aligned} \inf_{z_1, z_2 \in \mathbb{C}} \quad & \hat{f}(z_1, z_2) := C - |z_1|^2, \\ \text{s.t.} \quad & \hat{g}_1(z_1, z_2) := |z_1|^2 - az_1^2 - a\bar{z}_1^2 - 1 = 0, \\ & \hat{g}_2(z_1, z_2) := C - |z_1|^2 - |z_2|^2 = 0, \\ & \hat{g}_3(z_1, z_2) := \mathbf{i}z_2 - \mathbf{i}\bar{z}_2 = 0, \\ & \hat{g}_4(z_1, z_2) := z_2 + \bar{z}_2 \geq 0. \end{aligned} \quad (7.37)$$

Consider $a = \frac{1}{4}$ and $C = \frac{1}{1-2a} + 1 = 3$ so that $f^{\text{opt}} = 1$. Notice that $d^{\text{min}} = 2$ for (7.36) and (7.37). The complex relaxations of orders $2 \leq d \leq 3$ of (7.36) are unbounded. The complex relaxation of order 2 of (7.37) yields the value 0.6813 . That of order 3 yields the value 1.0000

⁵MATLAB 2013a, YALMIP 2015.06.26 [69], and MOSEK are used for the numerical experiments.

and the complex moment matrix⁶ with 10^{-4} precision

$$M_3(y) = \begin{array}{c|ccccccccccc} & \begin{matrix} (0,0) \\ (1,0) \\ (0,1) \\ (2,0) \\ (1,1) \\ (0,2) \\ (3,0) \\ (2,1) \\ (1,2) \\ (0,3) \end{matrix} & & & & & & & & & & \\ \begin{matrix} (0,0) \\ (1,0) \\ (0,1) \\ (2,0) \\ (1,1) \\ (0,2) \\ (3,0) \\ (2,1) \\ (1,2) \\ (0,3) \end{matrix} & & & & & & & & & & & \\ \hline & 1 & 0 & 1 & 2 & 0 & 1 & 0 & 2 & 0 & 1 \\ & 0 & 2 & 0 & 0 & 2 & 0 & 4 & 0 & 2 & 0 \\ & 1 & 0 & 1 & 2 & 0 & 1 & 0 & 2 & 0 & 1 \\ & 2 & 0 & 2 & 4 & 0 & 2 & 0 & 4 & 0 & 2 \\ & 0 & 2 & 0 & 0 & 2 & 0 & 4 & 0 & 2 & 0 \\ & 1 & 0 & 1 & 2 & 0 & 1 & 0 & 2 & 0 & 1 \\ & 0 & 4 & 0 & 0 & 4 & 0 & 8 & 0 & 4 & 0 \\ & 2 & 0 & 2 & 4 & 0 & 2 & 0 & 4 & 0 & 2 \\ & 0 & 2 & 0 & 0 & 2 & 0 & 4 & 0 & 2 & 0 \\ & 1 & 0 & 1 & 2 & 0 & 1 & 0 & 2 & 0 & 1 \end{array} \quad (7.38)$$

which satisfies $\text{rank } M_3(y) = \text{rank } M_1(y) = 2$.

Examples 7.8 and 7.9 show the importance of the modeling of the feasible set of the optimization problem. Depending on what equations are used to define the feasible set, the complex moment/sum-of-squares hierarchy may or may not converge towards the globally optimal value. If one of the constraints is a sphere, convergence is guaranteed. The real moment/sum-of-squares hierarchy also depends on how the feasible set is modeled. In that case, convergence is guaranteed if one of the constraints is a ball.

As a by-product of Corollary 7.7, we propose a solution to the complex moment problem in Theorem 7.11 below. To that end, consider Lemma 7.10 below where we transpose [56, Lemma 3] from real to complex numbers.

Lemma 7.10. Let $s : \mathbb{C}^n \rightarrow \mathbb{R}$ be defined by $s(z) := R^2 - |z_1|^2 - \dots - |z_n|^2$. Given $d \in \mathbb{N}^*$ and $y \in \mathcal{H}_d$, we have

$$(M_d(g_0y) \succcurlyeq 0 \quad \& \quad M_{d-1}(sy) = 0) \quad \implies \quad \text{Tr}(M_d(g_0y)) \leq y_{0,0} \sum_{l=0}^d R^{2l}. \quad (7.39)$$

Proof. Given $1 \leq l \leq d$, we have $\text{Tr}(M_{l-1}(sy)) = \dots$

$$\begin{aligned} &= \sum_{|\alpha| \leq l-1} M_{l-1}(sy)(\alpha, \alpha) \\ &= \sum_{|\alpha| \leq l-1} L_y(s(z) \bar{z}^\alpha z^\alpha) \\ &= \sum_{|\alpha| \leq l-1} \sum_{|\gamma| \leq 1} s_{\gamma, \gamma} y_{\gamma+\alpha, \gamma+\alpha} \\ &= \sum_{|\alpha| \leq l-1, |\gamma|=0} s_{\gamma, \gamma} y_{\gamma+\alpha, \gamma+\alpha} + \sum_{|\alpha| \leq l-1, |\gamma|=1} s_{\gamma, \gamma} y_{\gamma+\alpha, \gamma+\alpha} \\ &= \sum_{|\alpha| \leq l-1} R^2 y_{\alpha, \alpha} - \sum_{|\alpha| \leq l-1, |\gamma|=1} y_{\gamma+\alpha, \gamma+\alpha}. \end{aligned} \quad (7.40)$$

⁶It so happens that the Hermitian matrix $M_3(y)$, indexed by $(\alpha, \beta) \in \mathbb{N}^2 \times \mathbb{N}^2$ with $|\alpha|, |\beta| \leq 3$, is real-valued in this example.

$M_{d-1}(sy) = 0$ implies that $M_{l-1}(sy) = 0$ for all $1 \leq l \leq d$ and hence $\text{Tr}(M_{l-1}(sy)) = 0$. In addition, $\sum_{0 < |\alpha| \leq l} y_{\alpha,\alpha} \leq \sum_{|\alpha| \leq l-1, |\gamma|=1} y_{\gamma+\alpha, \gamma+\alpha}$. Thus

$$\sum_{|\alpha| \leq l} y_{\alpha,\alpha} \leq y_{0,0} + R^2 \sum_{|\alpha| \leq l-1} y_{\alpha,\alpha}, \quad l = 1, \dots, d, \quad (7.41)$$

which proves the lemma. \square

The next theorem can be deduced from [95] but we provide a different proof.

Theorem 7.11. *Under the sphere assumption (7.30), a sequence $y \in \mathcal{H}$ has a representing measure on K if and only if*

$$M_d(g_i y) \succcurlyeq 0, \quad i = 0, \dots, m, \quad \forall d \in \mathbb{N}. \quad (7.42)$$

Proof. It was already shown that if $y \in \mathcal{H}$ has a representing measure on K , then (7.22) holds. Notice that (7.22) and (7.42) are equivalent, hence the “only if” part. Concerning the “if” part, assume that $y \in \mathcal{H}$ satisfies (7.42). If $y_{0,0} = 0$, then Lemma 7.10 implies that $y = 0$ which can be represented by $\mu = 0$ on K . Otherwise $y_{0,0} > 0$ and $y/y_{0,0}$ is a feasible point of problem (7.23) whose optimal value is f^{opt} for all $f \in \mathbb{R}[\bar{z}, z]$ according to Corollary 7.7. If moreover $f|_K \geq 0$, then $L_{y/y_{0,0}}(f) \geq f^{\text{opt}} \geq 0$. In particular, if $f|_K = 0$, then $L_{y/y_{0,0}}(f) = 0$. We may therefore define $\tilde{L}_{y/y_{0,0}} : R_K(\mathbb{C}[\bar{z}, z]) \rightarrow \mathbb{C}$ such that $\tilde{L}_{y/y_{0,0}}(\varphi|_K) := L_{y/y_{0,0}}(\varphi)$ (similarly to Schweighofer [104, Proof of Theorem 2]). If $\varphi \in R_K(\mathbb{R}[\bar{z}, z])$, then $\tilde{L}_{y/y_{0,0}}(\|\varphi\|_\infty - \varphi) \geq 0$ and $\tilde{L}_{y/y_{0,0}}(\varphi) \leq \|\varphi\|_\infty$. Linearity implies that $|\tilde{L}_{y/y_{0,0}}(\varphi)| \leq \|\varphi\|_\infty$. As a result, for all $\varphi \in R_K(\mathbb{C}[\bar{z}, z])$, we have $|\tilde{L}_{y/y_{0,0}}(\varphi)| = |\tilde{L}_{y/y_{0,0}}(\text{Re}(\varphi) + \mathbf{i}\text{Im}(\varphi))| = |\tilde{L}_{y/y_{0,0}}(\text{Re}(\varphi)) + \mathbf{i}\tilde{L}_{y/y_{0,0}}(\text{Im}(\varphi))| \leq |\tilde{L}_{y/y_{0,0}}(\text{Re}(\varphi))| + |\tilde{L}_{y/y_{0,0}}(\text{Im}(\varphi))| \leq \|\text{Re}(\varphi)\|_\infty + \|\text{Im}(\varphi)\|_\infty \leq 2\|\varphi\|_\infty$. Moreover, $R_K(\mathbb{C}[\bar{z}, z])$ is dense in $C(K, \mathbb{C})$. Therefore $\tilde{L}_{y/y_{0,0}}$ may be extended to a continuous linear functional on $C(K, \mathbb{C})$ (we preserve the same name for the extension). K is compact thus the Riesz-Markov-Kakutani Representation Theorem implies that there exists a unique Radon measure μ such that $\tilde{L}_{y/y_{0,0}}(\varphi) = \int_K \varphi d\mu$ for all $\varphi \in C(K, \mathbb{C})$. It is positive because $\varphi \in \mathcal{P}(K)$ implies that $\tilde{L}_{y/y_{0,0}}(\varphi) \geq 0$ (density argument). Finally, if $\alpha, \beta \in \mathbb{N}^n$, $y_{\alpha,\beta}/y_{0,0} = L_{y/y_{0,0}}(\bar{z}^\alpha z^\beta)$ (c.f. Remark 7.3) hence y has representing measure $y_{0,0}\mu$ on K . \square

Vasilescu [124, Theorem I.2.17] has already proposed a different solution to the complex moment problem on K . We now transpose the proof of [56, Theorem 1] from real to complex numbers.

Proposition 7.12. *Under the sphere assumption (7.30), $\rho_d^* = \rho_d \in \mathbb{R} \cup \{+\infty\}$ for all $d \geq d^{\min}$.*

Proof. Given $A \in \mathcal{H}_d$, consider the operator norm $\|A\|$, the greatest eigenvalue of A in absolute value, and the Frobenius norm $\|A\|_{\mathcal{H}_d} := \sqrt{\langle A, A \rangle_{\mathcal{H}_d}}$. Consider $d \geq d^{\min}$. Two

cases can occur. The first is that the feasible set of the complex moment relaxation of order d is non-empty, in which case we consider a feasible point $(y_{\alpha,\beta})_{|\alpha|,|\beta|\leq d}$. All norms are equivalent in finite dimension so there exists a constant $C_d \in \mathbb{R}$ such that $\sqrt{\sum_{|\alpha|,|\beta|\leq d} |y_{\alpha,\beta}|^2} = \|M_d(g_0y)\|_{\mathcal{H}_d} \leq C_d \|M_d(g_0y)\| \leq C_d \sum_{l=0}^d R^{2l}$, according to Lemma 7.10. As a result, the feasible set of the complex moment relaxation of order d is a non-empty compact set and so is its image by Λ (defined in (7.3)). We can thus apply Trnovská's result [122] which states that in a semidefinite program in real numbers, if the primal feasible set is non-empty and compact, then there exists a dual interior point and there is no duality gap.

The second case is that the feasible set of the complex moment relaxation of order d is empty, i.e. $\rho_d = +\infty$. It must be strongly infeasible because it cannot be weakly infeasible (see [40, Section 5.2] for definitions). Indeed, if it is weakly infeasible, then there exists a sequence $(y^j)_{j \in \mathbb{N}}$ of elements of \mathcal{H} such that for all $j \in \mathbb{N}$, we have $|y_{0,0}^j - 1| \leq \frac{1}{j+1}$ and $\lambda_{\min}(M_{d-k_i}(g_i y^j)) \geq -\frac{1}{j+1}$ where $i = 0, \dots, m$. Define $c := (n+d)!/(n!d!)$. We now mimic the computations in Lemma 7.10 using $y_{0,0}^j \leq 1 + \frac{1}{j+1} \leq 2$ and $|\text{Tr}(M_{l-1}(s y^j))| \leq \frac{c}{j+1} \leq c$ if $1 \leq l \leq d$. Consider $j_0 \in \mathbb{N}$ such that for all $j \geq j_0$ and $1 \leq l \leq d$, we have $\sum_{|\alpha| \leq l-1, |\gamma|=1} y_{\gamma+\alpha, \gamma+\alpha}^j - \sum_{0 < |\alpha| \leq l} y_{\alpha, \alpha}^j \geq -1$. Equation (7.41) then becomes $\sum_{|\alpha| \leq l} y_{\alpha, \alpha}^j \leq 2 + R^2 \left(\sum_{|\alpha| \leq l-1} y_{\alpha, \alpha}^j \right) + c + 1$. As a result, $\text{Tr}(M_d(g_0 y^j)) = \sum_{|\alpha| \leq d} y_{\alpha, \alpha}^j \leq (3+c) \sum_{l=0}^d R^{2l}$, which, together with $\lambda_{\min}(M_d(g_0 y^j)) \geq -\frac{1}{j+1} \geq -1$, yields $\lambda_{\max}(M_d(g_0 y^j)) \leq (3+c) \sum_{l=0}^d R^{2l} + c - 1$. Hence for all $j \geq j_0$, the spectrum of $M_d(g_0 y^j)$ is lower bounded by -1 and upper bounded by $B_d := (3+c) \sum_{l=0}^d R^{2l} + c - 1 \geq 1$. We therefore have $\sqrt{\sum_{|\alpha|,|\beta|\leq d} |y_{\alpha,\beta}^j|^2} \leq C_d \|M_d(g_0 y)\| \leq C_d \times B_d$. The sequence $(y^j)_{j \geq j_0}$ is thus included in a compact set. Hence there exists a subsequence that converges towards a limit y^{lim} which satisfies $y_{0,0}^{\text{lim}} = 1$ and the constraints $\lambda_{\min}(M_{d-k_i}(g_i y^{\text{lim}})) \geq 0$, $i = 0, \dots, m$. Therefore y^{lim} is a feasible point of the complex moment relaxation of order d , which is a contradiction. Strong infeasibility means that the dual feasible set contains an improving ray [40, Definition 5.2.2]. Moreover, $\inf_{y \in \mathcal{H}_d} L_y(f)$ subject to $y_{0,0} = 1$, $M_d(g_0 y) \succcurlyeq 0$, and $M_{d-1}(s y) = 0$ is a semidefinite program with a non-empty compact feasible set hence the dual feasible set contains a point $(\lambda, \sigma_0, \sigma_1)$. As result $(\lambda, \sigma_0, \sigma_1, 0, \dots, 0)$ is a feasible point of the complex sum-of-squares relaxation of order d . Together with the improving ray, this means that $\rho_d^* = +\infty$. To conclude, $\rho_d^* = \rho_d$ in both cases. \square

Proposition 7.13. *Assume that complex polynomial optimization problem (7.15) satisfies (7.32) and has a global solution $z^{\text{opt}} \in K^{\text{opt}}$. In addition, assume that $(\sigma_0^{\text{opt}}, \dots, \sigma_m^{\text{opt}}) \in \Sigma[z]^{m+1}$ is an optimal solution to the sum-of-squares problem (7.27). Then $(z^{\text{opt}}, \sigma_1^{\text{opt}}, \dots, \sigma_m^{\text{opt}})$ is a saddle point of $\phi : \mathbb{C}^n \times \Sigma[z]^m \rightarrow \mathbb{R}$ defined by $(z, \sigma) \mapsto f(z) - \sum_{i=1}^m \sigma_i(z) g_i(z)$.*

Proof. The optimality of $(\sigma_0^{\text{opt}}, \dots, \sigma_m^{\text{opt}})$ means that $f - f^{\text{opt}} = \sum_{i=0}^m \sigma_i^{\text{opt}} g_i$. With $f(z^{\text{opt}}) - f^{\text{opt}} = \sum_{i=0}^m \sigma_i^{\text{opt}}(z^{\text{opt}}) g_i(z^{\text{opt}}) = 0$, $\sigma_i^{\text{opt}}(z^{\text{opt}}) \geq 0$, and $g_i(z^{\text{opt}}) \geq 0$, we have $\sigma_i^{\text{opt}}(z^{\text{opt}}) g_i(z^{\text{opt}}) =$

0 for $i = 0, \dots, m$. It follows that $\phi(z^{\text{opt}}, \sigma) \leq \phi(z^{\text{opt}}, \sigma^{\text{opt}})$ for all $\sigma \in \Sigma[z]$. For all $z \in \mathbb{C}^n$, $\phi(z^{\text{opt}}, \sigma^{\text{opt}}) \leq \phi(z, \sigma^{\text{opt}})$ because $f(z) - f^{\text{opt}} - \sum_{i=1}^m \sigma_i^{\text{opt}}(z)g_i(z) = \sigma_0^{\text{opt}}(z) \geq 0$. \square

Given an application $\varphi : \mathbb{C}^n \rightarrow \mathbb{R}$, define $\tilde{\varphi} : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ by $(x, y) \mapsto \varphi(x + \mathbf{i}y)$. If $\tilde{\varphi}$ is \mathbb{R} -differentiable at point $(x, y) \in \mathbb{R}^{2n}$, consider the Wirtinger derivative [129] defined by $\nabla \varphi(x + \mathbf{i}y) := \frac{1}{2}(\nabla_x \tilde{\varphi}(x, y) - \mathbf{i} \nabla_y \tilde{\varphi}(x, y)) \in \mathbb{C}^n$.

Corollary 7.14. *With the same assumptions as in Proposition 7.13, we have*

$$\begin{aligned} \nabla f(z^{\text{opt}}) &= \sum_{i=1}^m \sigma_i^{\text{opt}}(z^{\text{opt}}) \nabla g_i(z^{\text{opt}}), \\ \sigma_i^{\text{opt}}(z^{\text{opt}}), g_i(z^{\text{opt}}) &\geq 0, \quad i = 1, \dots, m, \\ \sigma_i^{\text{opt}}(z^{\text{opt}}) g_i(z^{\text{opt}}) &= 0, \quad i = 1, \dots, m. \end{aligned} \tag{7.43}$$

Proof. z^{opt} is a minimizer of $z \in \mathbb{C}^n \mapsto \phi(z, \sigma^{\text{opt}})$ thus $\nabla_z \phi(z^{\text{opt}}, \sigma^{\text{opt}}) = \nabla f(z^{\text{opt}}) - \sum_{i=1}^m \nabla \sigma_i^{\text{opt}}(z^{\text{opt}}) g_i(z^{\text{opt}}) - \sum_{i=1}^m \sigma_i^{\text{opt}}(z^{\text{opt}}) \nabla g_i(z^{\text{opt}}) = 0$. Consider $1 \leq i \leq m$. Since $\sigma_i^{\text{opt}}(z^{\text{opt}}) = 0$ and $\sigma_i^{\text{opt}} \in \Sigma[z]$, it must be that $|z_k - z_k^{\text{opt}}|^2$ divides $\sigma_{i,k}^{\text{opt}} : z_k \in \mathbb{C} \mapsto \sigma_i^{\text{opt}}(z_1^{\text{opt}}, \dots, z_{k-1}^{\text{opt}}, z_k, z_{k+1}^{\text{opt}}, \dots, z_n^{\text{opt}})$. With $z_k^{\text{opt}} =: x_k^{\text{opt}} + \mathbf{i}y_k^{\text{opt}}$, the real number x_k^{opt} is a root of multiplicity 2 of $x_k \in \mathbb{R} \mapsto \sigma_{i,k}^{\text{opt}}(x_k + \mathbf{i}y_k^{\text{opt}})$, with an analogous remark for y_k^{opt} . Thus $\nabla \sigma_i^{\text{opt}}(z^{\text{opt}}) = 0$ which leads to the desired result. \square

Comparison of real and complex hierarchies

Similar to Shor relaxation and the second-order conic relaxation, the following notations will be used: POP- \mathbb{C} denotes the complex polynomial optimization problem (7.15); POP- \mathbb{R} denotes the real polynomial optimization problem after conversion of POP- \mathbb{C} into real numbers; MSOS- d - \mathbb{C} denotes the complex moment/sum-of-squares relaxation of order d applied to POP- \mathbb{C} ; CMSOS- d - \mathbb{R} denotes the conversion of MSOS- d - \mathbb{C} into real numbers; and MSOS- d - \mathbb{R} denotes the real moment/sum-of-squares relaxation of order d applied to POP- \mathbb{R} . Let $d^{\text{min}}\text{-}\mathbb{R}$ and $d^{\text{min}}\text{-}\mathbb{C}$ respectively denote the minimum orders of the real and complex hierarchies. Consider the sets

$$\begin{aligned} \mathbb{R}[x, y] &:= \{ q : \mathbb{R}^{2n} \rightarrow \mathbb{R} \mid q(x, y) = \sum_{|\kappa| \leq j} q_\kappa(x, y)^\kappa, j \in \mathbb{N}, q_\kappa \in \mathbb{R} \}, \\ \mathbb{R}_d[x, y] &:= \{ q : \mathbb{R}^{2n} \rightarrow \mathbb{R} \mid q(x, y) = \sum_{|\kappa| \leq d} q_\kappa(x, y)^\kappa, q_\kappa \in \mathbb{R} \}, \\ \Sigma_d[x, y] &:= \{ \sigma : \mathbb{R}^{2n} \rightarrow \mathbb{R} \mid \sigma = \sum_{i=1}^r q_i^2, \text{ with } r \in \mathbb{N}^*, q_i \in \mathbb{R}_d[x, y] \}, \end{aligned} \tag{7.44}$$

where $\kappa \in \mathbb{N}^{2n}$ and $(x, y)^\kappa := x_1^{\kappa_1} \dots x_n^{\kappa_n} y_1^{\kappa_{n+1}} \dots y_n^{\kappa_{2n}}$.

Proposition 7.15. *Under the sphere assumption (7.30), for all integer d greater than or equal to $\max\{d^{\text{min}}\text{-}\mathbb{R}, d^{\text{min}}\text{-}\mathbb{C}\}$, we have*

$$\text{val}(\text{MSOS}_d\text{-}\mathbb{C}) = \text{val}(\text{CMSOS}_d\text{-}\mathbb{R}) \leq \text{val}(\text{MSOS}_d\text{-}\mathbb{R}). \tag{7.45}$$

Proof. It suffices to compare the optimal values of the real and complex sum-of-squares relaxations. This is due to Proposition 7.12 and [56] where the ball constraint can be replaced by the sphere constraint to ensure no duality gap. We have

$$\begin{aligned}
\text{val}(\text{POP-C}) &= \sup_{\lambda \in \mathbb{R}} \lambda, \\
&\text{s.t.} \quad f(z) - \lambda \geq 0, \quad \forall z \in K, \\
\text{val}(\text{MSOS}_d\text{-C}) &= \sup_{\lambda, \sigma} \lambda, \\
&\text{s.t.} \quad f - \lambda = \sum_{i=0}^m \sigma_i g_i, \\
&\quad \lambda \in \mathbb{R}, \quad \sigma_i \in \Sigma_{d-k_i}[z], \quad i = 0, \dots, m, \\
\text{val}(\text{CMSOS}_d\text{-}\mathbb{R}) &= \sup_{\lambda, \sigma} \lambda, \\
&\text{s.t.} \quad \tilde{f} - \lambda = \sum_{i=0}^m \sigma_i \tilde{g}_i, \\
&\quad \lambda \in \mathbb{R}, \quad \sigma_i \in \Sigma_{d-k_i}[x + \mathbf{i}y], \quad i = 0, \dots, m, \\
\text{val}(\text{POP-}\mathbb{R}) &= \sup_{\lambda \in \mathbb{R}} \lambda, \\
&\text{s.t.} \quad \tilde{f}(x, y) - \lambda \geq 0, \quad \forall (x + \mathbf{i}y) \in K, \\
\text{val}(\text{MSOS}_d\text{-}\mathbb{R}) &= \sup_{\lambda, \sigma} \lambda, \\
&\text{s.t.} \quad \tilde{f} - \lambda = \sum_{i=0}^m \sigma_i \tilde{g}_i, \\
&\quad \lambda \in \mathbb{R}, \quad \sigma_i \in \Sigma_{d-k_i}[x, y], \quad i = 0, \dots, m.
\end{aligned} \tag{7.46}$$

We now conclude because for all $d \in \mathbb{N}$, $\Sigma_d[x + \mathbf{i}y] \subset \Sigma_d[x, y]$. Indeed, if $\sigma = \sum_{j=1}^r |p_j|^2$ with $r \in \mathbb{N}^*$ and $p_1, \dots, p_r \in \mathbb{C}_d[z]$, then $\tilde{\sigma}(x, y) = \sum_{j=1}^r |\tilde{p}_j(x, y)|^2 = \frac{1}{4} \sum_{j=1}^r \left(\tilde{p}_j(x, y) + \overline{\tilde{p}_j(x, y)} \right)^2 + \left(\mathbf{i}\tilde{p}_j(x, y) + \overline{\mathbf{i}\tilde{p}_j(x, y)} \right)^2 \in \Sigma_d[x, y]$. \square

We may suspect the inequality in (7.45) to be strict in some cases because $\Sigma_d[x + \mathbf{i}y]$ is a strict subset of $\Sigma_d[x, y]$ for all $d \in \mathbb{N}^*$. Indeed, for $i = 1, \dots, n$, we have $x_i^2 = \left(\frac{z_i + \bar{z}_i}{2} \right)^2 = \frac{1}{4}(z_i^2 + 2|z_i|^2 + \bar{z}_i^2) \in \Sigma_d[x, y] \setminus \Sigma_d[x + \mathbf{i}y]$. According to numerical experiments⁷, the inequality is strict for (7.37) in Example 7.9 ($\text{val}(\text{CMSOS}_2\text{-}\mathbb{R}) \approx 0.6813$ and $\text{val}(\text{MSOS}_2\text{-}\mathbb{R}) \approx 1.0000$).

Proposition 7.45 seems to imply that the real moment/sum-of-squares hierarchy is better than the complex one. However, the size of the largest semidefinite constraint of $\text{CMSOS}_d\text{-}\mathbb{R}$, equal to $2(n+d)!/(n!d!)$, is far inferior to that of $\text{MSOS}_d\text{-}\mathbb{R}$, equal to $(2n+d)!/((2n)!d!)$. For instance, if $n = 10$ and $d = 3$, the former is 572 and the latter is 1,771.

Proposition 7.16. *Given $l \in \mathbb{N}$ and $\varphi \in \mathbb{C}_l[\bar{z}, z]$, we have*

$$\forall z \in \mathbb{C}^n, \quad \forall \theta \in \mathbb{R}, \quad \varphi(e^{\mathbf{i}\theta} z) = \varphi(z) \quad \iff \quad \forall |\alpha|, |\beta| \leq l, \quad |\alpha - \beta| \varphi_{\alpha, \beta} = 0. \tag{7.47}$$

⁷We attempted a formal proof but it is difficult even on such a small example.

Proof. (\implies) Notice that $\varphi(e^{i\theta}z) = \sum_{|\alpha|, |\beta| \leq l} \varphi_{\alpha, \beta} \overline{(e^{i\theta}z)}^\alpha (e^{i\theta}z)^\beta = \dots$
 $\sum_{|\alpha|, |\beta| \leq l} \varphi_{\alpha, \beta} e^{i(|\alpha| - |\beta|)\theta} \bar{z}^\alpha z^\beta$. Polarization implies that for all $z, w \in \mathbb{C}^n$, we have
 $\sum_{|\alpha|, |\beta| \leq l} \varphi_{\alpha, \beta} e^{i(|\alpha| - |\beta|)\theta} \bar{z}^\alpha w^\beta = \sum_{|\alpha|, |\beta| \leq l} \varphi_{\alpha, \beta} \bar{z}^\alpha w^\beta$ and hence for all $|\alpha|, |\beta| \leq l$, $\varphi_{\alpha, \beta} e^{i(|\alpha| - |\beta|)\theta} = \varphi_{\alpha, \beta}$. If $\varphi_{\alpha, \beta} \neq 0$, then for all $\theta \in \mathbb{R}$, $|\alpha - \beta|\theta \equiv 0[2\pi]$ and thus $|\alpha - \beta| = 0$. (\impliedby) Simply compute $\varphi(e^{i\theta}z)$. \square

Definition 7.17. *Complex polynomial optimization problem (7.15) is said to be oscillatory if $f, g_1, \dots, \text{ and } g_m$ satisfy either of the two equivalent properties in (7.47).*

Proposition 7.18. *If complex polynomial optimization problem (7.15) is oscillatory, then $d^{\min}\text{-}\mathbb{R} = d^{\min}\text{-}\mathbb{C}$.*

Proof. Observe that $d^{\min}\text{-}\mathbb{C} = \max\{|\alpha|, |\beta| \mid f_{\alpha, \beta} g_{1, \alpha, \beta} \dots g_{m, \alpha, \beta} \neq 0\}$ and $d^{\min}\text{-}\mathbb{R} = \max\{\lceil (|\alpha| + |\beta|)/2 \rceil \mid f_{\alpha, \beta} g_{1, \alpha, \beta} \dots g_{m, \alpha, \beta} \neq 0\}$ where $\lceil \cdot \rceil$ denotes the ceiling of a real number. Both are equal if the problem is oscillatory. \square

Conjecture 7.19. *Under the sphere assumption (7.30), if complex polynomial optimization problem (7.15) is oscillatory, then for all $d \geq d^{\min}\text{-}\mathbb{R} = d^{\min}\text{-}\mathbb{C}$, we have*

$$\text{val}(\text{MSOS}_d\text{-}\mathbb{C}) = \text{val}(\text{CMSOS}_d\text{-}\mathbb{R}) = \text{val}(\text{MSOS}_d\text{-}\mathbb{R}). \quad (7.48)$$

In Section 7.4, we consider problems for which Conjecture 7.19 seems to hold numerically. This suggests that for oscillatory problems, the complex hierarchy is more tractable than the real hierarchy at no loss of bound quality.

Exploiting sparsity in real and complex hierarchies

The chordal sparsity technique described in Section 7.2 has been extended to the real hierarchy by Waki [126] and may readily be transposed to the complex hierarchy. Each positive semidefinite constraint in (7.24) is replaced by a set of positive semidefinite constraints on certain submatrices of $M_{d-k_i}(g_i y)$. These submatrices are defined by the maximal cliques of a chordal extension of the graph associated with the objective and constraint equations. Equivalently, the sum-of-squares variables σ_i in the dual formulation (7.26) are restricted to be functions of a subset (defined by the same maximal cliques) of the decision variables z_1, \dots, z_n . These sparse relaxation hierarchies provide potentially lower bounds than their dense counterparts yet retain convergence guarantees [64]. However, further size reduction is often necessary. We propose to selectively apply computationally intensive higher-order constraints in the sparse relaxations. In other words, rather than a single relaxation order applied to all constraints, each constraint has an associated relaxation order. This allows for solving many large-scale problems.

We now formalize our approach applied to the complex hierarchy.⁸ Objective function f and constraints $(g_i)_{1 \leq i \leq m}$ in (7.15) have an associated undirected sparsity graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$

⁸See [81] for the details of this approach as applied to $\text{MSOS}_d\text{-}\mathbb{R}$ in the context of the optimal power flow problem.

with nodes $\mathcal{N} = \{1, \dots, n\}$ corresponding to each variable and edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ for each pair of variables that appear together in any monomial that has a non-zero coefficient in the objective function or constraints.

Each constraint function g_i has an associated relaxation order d_i so that $d \in \mathbb{N}^m$. When $d_i > 1$, there must exist at least one clique of a chordal extension of \mathcal{G} that contains all variables with non-zero coefficients in g_i . To ensure this, define a supergraph $\hat{\mathcal{G}} = (\mathcal{N}, \hat{\mathcal{E}})$ where $\hat{\mathcal{E}}$ is composed of \mathcal{E} augmented with edges connecting all variables with non-zero coefficients in g_i , not necessarily in the same monomial. For example, $g_1(z) = \bar{z}_1 z_2 + \bar{z}_2 z_1 + \bar{z}_1 z_3 + \bar{z}_3 z_1 + \bar{z}_1 z_4 + \bar{z}_4 z_1$ with $d_1 > 1$ implies $\mathcal{E} \supset \{(1, 2), (1, 3), (1, 4)\}$ and $\hat{\mathcal{E}} \supset \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$.

To exploit sparsity, construct a chordal extension \mathcal{G}^{ch} of $\hat{\mathcal{G}}$.⁹ Denote the set of maximally sized cliques of the chordal extension by $\mathcal{C}_1, \dots, \mathcal{C}_p \subset \mathcal{G}^{\text{ch}}$. By construction of $\hat{\mathcal{G}}$, each constraint function g_i for which $d_i > 1$ has all associated variables contained in at least one clique. For each g_i for which $d_i > 1$, denote as $\mathcal{C}^{(i)}$ the *minimal covering clique* (i.e., the smallest clique in $\{\mathcal{C}_1, \dots, \mathcal{C}_p\}$ that contains all variables in g_i). (If not unique, a single clique $\mathcal{C}^{(i)}$ is chosen arbitrarily among the smallest cliques.) Associate an order $\tilde{d} \in \mathbb{N}^p$ with each clique \mathcal{C}_γ , $\gamma = 1, \dots, p$ defined such that \tilde{d}_γ is the maximum relaxation order d_i among all constraints for which the clique \mathcal{C}_γ is the minimal covering clique. If a clique \mathcal{C}_γ is not a minimal covering clique for any constraints, then $\tilde{d}_\gamma = 1$. See Appendix F for a small illustrative example.

For all $1 \leq i \leq m$ such that $d_i > 1$, the positive semidefinite constraints $M_{d-k_i}(g_i y) \succcurlyeq 0$ in the moment hierarchy (7.24) are replaced by $N_i(g_i y) \succcurlyeq 0$, $i = 1, \dots, m$, where $N_i(g_i y)(\alpha, \beta) := M_{d_i-k_i}(g_i y)(\alpha, \beta)$ such that all non-zero entries of α and β correspond to variables in $\mathcal{C}^{(i)}$. For $i = 0$, the positive semidefinite constraint $M_d(y) \succcurlyeq 0$ (recall that $g_0 = 1$ and $k_0 = 0$) is replaced by constraints defined by each maximal clique: $\tilde{N}_\gamma(y) \succcurlyeq 0$, $\gamma = 1, \dots, p$, where $\tilde{N}_\gamma(y)(\alpha, \beta) := M_{\tilde{d}_\gamma}(y)(\alpha, \beta)$ such that all non-zero entries of α and β correspond to variables in \mathcal{C}_γ .

For the sum-of-squares representation of the hierarchy, the polynomials $\sigma_i \in \Sigma_{d-k_i}[z]$, $i = 1, \dots, m$ in (7.26) are replaced by sums-of-squares polynomials $\omega_i \in \Sigma_{d_i-k_i}[z_{\mathcal{C}^{(i)}}]$, $i = 1, \dots, m$, where $z_{\mathcal{C}^{(i)}}$ denotes the subset of variables z that are in the clique $\mathcal{C}^{(i)}$. The polynomial $\sigma_0 \in \Sigma_{d-k_0}[z]$ is replaced by the polynomial $\sum_{\gamma=1}^p \tau_\gamma$ where $\tau_\gamma \in \Sigma_{\tilde{d}_\gamma}[z_{\mathcal{C}_\gamma}]$, $\gamma = 1, \dots, p$.

The sparse version of the real hierarchy $\text{MSOS}_d\text{-}\mathbb{R}$ converges to the global optimum of a polynomial optimization problem when the constraints include ball constraints on all decision variables x included in each clique: $\sum_{k \in \mathcal{C}_i} x_k^2 \leq (R_{\mathcal{C}_i})^2$, $i = 1, \dots, p$, where $R_{\mathcal{C}_i}$ is the radius of a ball enclosing all decision variables in clique \mathcal{C}_i [64]. A similar result holds for the complex hierarchy $\text{MSOS}_d\text{-}\mathbb{C}$ with sphere constraints enforced for the variables included in each clique. Due to (7.32), the sparse version of the complex hierarchy is guaranteed to converge to

⁹One approach to creating a chordal extension is to use the sparsity pattern of a Cholesky factorization (employing a minimum degree ordering to maintain sparsity) of the Laplacian matrix associated with $\hat{\mathcal{G}}$ plus an identity matrix.

the global optimum of (7.15) with increasing relaxation order when the constraints include $(\sum_{k \in \mathcal{C}_i} |z_k|^2) + |z_{n+i}|^2 = (R_{\mathcal{C}_i})^2$, $i = 1, \dots, p$, where z_{n+i} is a slack variable associated with clique \mathcal{C}_i .

Selectively applying the higher-order constraints requires a method for determining the relaxation order d_i for each constraint. We use a heuristic based on “mismatches” to the closest rank-one matrix [81]. The idea is to extract the largest eigenvalue λ_1 and its associated unit-length eigenvector η_1 from $(y_{\alpha,\beta})_{|\alpha|=|\beta|=1}$, hence defining an “approximate” solution $z^{\text{approx}} := \sqrt{\lambda_1} \eta_1$ to the polynomial optimization problem. Define “mismatches” $\zeta \in \mathbb{R}$ and $\Delta \in \mathbb{R}^m$ between the solution y to the relaxation and z^{approx} :

$$\zeta := |f(z^{\text{approx}}) - L_y(f)|, \quad (7.49a)$$

$$\Delta_i := |g_i(z^{\text{approx}}) - L_y(g_i)|, \quad i = 1, \dots, m. \quad (7.49b)$$

We use the iteration in Algorithm 2 to determine relaxation orders d_i , $i = 1, \dots, m$. Each iteration solves the moment/sum-of-squares relaxation after increasing the relaxation orders d_i in a manner that is dependent on the largest associated Δ_i values. Denote $d^{\text{max}} := \max_i \{d_i\}$.¹⁰ At each iteration of the algorithm, increment d_i at up to h constraints, where h is a specified parameter, that have the largest mismatches Δ_i among constraints satisfying two conditions: (1) $d_i < d^{\text{max}}$ and (2) $\Delta_i > \epsilon_g$, where ϵ_g is a specified mismatch tolerance. If no constraints satisfy both of these conditions, increment d_i at up to h constraints with the largest Δ_i greater than the specified tolerance and increment d^{max} . That is, in order to prevent unnecessarily increasing the size of the matrices, the heuristic avoids incrementing the maximum relaxation order d^{max} until $d_i = d^{\text{max}}$ at all constraints g_i with mismatch $\Delta_i > \epsilon_g$.

There is a computational trade-off in choosing the value of h . Larger values of h likely result in fewer iterations of the algorithm but each iteration is slower if more buses than necessary have high-order relaxations. Smaller values of h result in faster solution at each iteration, but may require more iterations.

The algorithm terminates upon satisfaction of two conditions: First, $|\Delta|_\infty \leq \epsilon_g$, where $|\cdot|_\infty$ denotes the infinity norm (maximum absolute value), which indicates that the iterate is a numerically feasible point of polynomial optimization problem (7.15). Second, $\zeta \leq \epsilon_f$, which indicates global optimality to within a relative tolerance ϵ_f . If the relaxation satisfies the former but not the latter termination condition (which was never observed in practice for the problem in Section 7.4), the algorithm increases d_i at the h constraints with largest mismatch Δ_i and continues iterating.

The moment/sum-of-squares hierarchy is successively tightened in a manner that preserves computational tractability. For sufficiently small tolerances ϵ_f and ϵ_g , Algorithm 2 eventually proceeds to build the complete moment/sum-of-squares hierarchies. Thus, Algorithm 2 inherits the theoretical convergence guarantees of MSOS $_d$ -C. The same can be said of the real version of Algorithm 2 applied to MSOS $_d$ -R.

¹⁰Note that d^{max} is not a specified maximum relaxation order but can change at each iteration.

Algorithm 2 Iterative Solution for Sparse Moment/Sum-of-Squares Relaxations

-
- 1: Set $d_i = 1$, $i = 1, \dots, m$.
 - 2: **repeat**
 - 3: Solve relaxation with order d_i for constraints $g_i(z) \geq 0$, $i = 1, \dots, m$.
 - 4: Calculate mismatches Δ_i , $i = 1, \dots, m$ using (7.49b).
 - 5: Increase entries of d according to the mismatch heuristic.
 - 6: **until** $|\Delta|_\infty < \epsilon_g$ and $\zeta < \epsilon_f$
 - 7: Extract solution z^{opt} .
-

7.4 Numerical results

The optimal power flow problem is an instance of complex polynomial optimization. Since 2006, the power systems literature has been studying the ability of the Shor and second-order conic relaxations to find global solutions [7, 13, 18, 20, 22, 30, 51, 66, 70, 72, 80, 82–85, 118, 119, 131]. Some relaxations are presented in real numbers [66, 82] and some in complex numbers [18, 20, 131]. Nevertheless, in all numerical applications, standard solvers such as SeDuMi, SDPT3, and MOSEK are used which currently handle only real numbers. Modeling languages such as YALMIP and CVX do handle inputs in complex numbers, but the data is transformed into real numbers before calling the solver [21, Example 4.42]. We use the European network to illustrate the fact that it is beneficial to relax nonconvex constraints before converting from complex to real numbers. The Shor relaxation, the second-order conic relaxation, and the moment/sum-of-squares hierarchy are considered.

We consider large test cases representing portions of European electric power systems. They represent Great Britain (GB) [123] and Poland (PL) [132] power systems as well as other European systems from the PEGASE project [42, 55]. The test cases were preprocessed to remove low-impedance lines as described in [84] in order to improve the solver’s numerical convergence, which is a typical procedure in power system analyses.¹¹ A 1×10^{-3} per unit low-impedance line threshold was used for all test cases except for PEGASE-1354 and PEGASE-2869 which use a 3×10^{-3} per unit threshold. The processed data is described in Table 7.1. This table also includes the at-least-locally-optimal objective values obtained from the interior point solver in MATPOWER [132] for the problems after preprocessing. Note that the PEGASE systems specify generation costs that minimize active power losses, so the objective values in both columns are the same.

Implementations use YALMIP 2015.06.26 [69], Mosek 7.1.0.28, and MATLAB 2013a on a computer with a quad-core 2.70 GHz processor and 16 GB of RAM. The results do not include the typically small formulation times.

¹¹Low-impedance lines often model connections between buses in the same physical location.

¹²PEGASE-9241 contains negative resistances to account for generators at lower voltage levels. In PEGASE-9241R these are set to 0.

Table 7.1: Size of Data (After Low-Impedance Line Preprocessing)

Test Case Name	Number of Complex Variables	Number of Edges in Graph	MATPOWER Solution [132]	
			Gen. Cost (\$/hr)	Loss Min. (MW)
GB-2224	2,053	2,581	1,942,260	60,614
PL-2383wp	2,177	2,651	1,868,350	24,991
PL-2736sp	2,182	2,675	1,307,859	18,336
PL-2737sop	2,183	2,675	777,617	11,397
PL-2746wop	2,189	2,708	1,208,257	19,212
PL-2746wp	2,192	2,686	1,631,737	25,269
PL-3012wp	2,292	2,805	2,592,462	27,646
PL-3120sp	2,314	2,835	2,142,720	21,513
PEGASE-89	70	185	5,819	5,819
PEGASE-1354	983	1,526	74,043	74,043
PEGASE-2869	2,120	3,487	133,945	133,945
PEGASE-9241	7,154	12,292	315,749	315,749
PEGASE-9241R ¹²	7,154	12,292	315,785	315,785

Shor relaxation

Table 7.2 shows the results of applying SDP- \mathbb{R} and SDP- \mathbb{C} to the test cases. For some problems, the Shor relaxation is *exact* and yields the globally optimal decision variables and objective values. To practically identify such problems, solutions for which all power injection mismatches $S_k^{\text{inj mis}}$ (see Section 7.4) are less than a tolerance of 1 MVA are considered exact. These problems are identified with an asterisk (*) in Table 7.2.

The lower bounds in Table 7.2 suggest that the corresponding MATPOWER solutions in Table 7.1 are at least very close to being globally optimal. The gap between the MATPOWER solutions and the lower bounds from SDP- \mathbb{C} for the generation cost minimizing problems are less than 0.72% for GB-2224, 0.29% for the Polish systems, and 0.02% for the PEGASE systems with the exception of PEGASE-9241. The non-physical negative resistances in PEGASE-9241 result in weaker lower bounds from the relaxations, yielding a gap of 1.64% for this test case.

As shown in Appendices C and D, the optimal objective values for SDP- \mathbb{R} and SDP- \mathbb{C} should be identical. With all objective values in Table 7.2 matching to within 0.037%, this is numerically validated.

For these test cases, SDP- \mathbb{C} is significantly faster (between a factor of 1.60 and 3.31) than SDP- \mathbb{R} . This suggests that exploiting the isomorphic structure of complex matrices in SDP- \mathbb{C} is better than eliminating a row and column in SDP- \mathbb{R} .

Table 7.2: Real and Complex SDP (Generation Cost Minimization)

Case Name	SDP- \mathbb{R}		SDP- \mathbb{C}	
	Val. (\$/hr)	Time (sec)	Val. (\$/hr)	Time (sec)
GB-2224	1,928,194	10.9	1,928,444	6.2
PL-2383wp	1,862,979	48.1	1,862,985	23.0
PL-2736sp*	1,307,749	35.7	1,307,764	22.0
PL-2737sop*	777,505	41.7	777,539	19.5
PL-2746wop*	1,208,168	51.1	1,208,182	22.8
PL-2746wp	1,631,589	43.8	1,631,655	20.0
PL-3012wp	2,588,249	52.8	2,588,259	24.3
PL-3120sp	2,140,568	64.4	2,140,605	25.5
PEGASE-89*	5,819	1.5	5,819	0.9
PEGASE-1354	74,035	11.2	74,035	5.6
PEGASE-2869	133,936	38.2	133,936	20.6
PEGASE-9241	310,658	369.7	310,662	136.1
PEGASE-9241R	315,848	317.2	315,731	95.9

Second-order conic relaxation

Table 7.3 shows the results of applying SOCP- \mathbb{R} and SOCP- \mathbb{C} to the test cases. Unlike the Shor relaxation, the second-order conic relaxation is not exact for any of the test cases. (SOCP- \mathbb{C} is generally not exact with the exception of radial systems for which the relaxation is provably exact when certain non-trivial technical conditions are satisfied [70].)

SOCP- \mathbb{C} provides better lower bounds and is computationally faster than SOCP- \mathbb{R} . Specifically, lower bounds from SOCP- \mathbb{C} are between 0.87% and 3.96% larger and solver times are faster by between a factor of 1.24 and 6.76 than those from SOCP- \mathbb{R} .

Moment/sum-of-squares hierarchy

Relaxations from the real moment/sum-of-squares hierarchy globally solve a broad class of optimal power flow problems [44, 57, 79, 81]. Previous work uses MSOS- d - \mathbb{R} by first converting the complex formulation of the optimal power flow problem to real numbers.

We next summarize computational aspects of both the real and complex hierarchies. The dense formulations of the hierarchies solve small problems (up to approximately ten buses). Without also selectively applying the higher-order relaxations' constraints (i.e., $d_1 = d_2 = \dots = d_m = d$), exploiting network sparsity enables solution of the second-order relaxations for problems with up to approximately 40 buses.

Scaling to larger problems is accomplished by both exploiting network sparsity and selectively applying the computationally intensive higher-order relaxation constraints to specific "problematic" buses. To better match the structure of the optimal power flow constraint

Table 7.3: Real and Complex SOCP (Generation Cost Minimization)

Case Name	SOCP- \mathbb{R}		SOCP- \mathbb{C}	
	Val. (\$/hr)	Time (sec)	Val. (\$/hr)	Time (sec)
GB-2224	1,855,393	3.5	1,925,723	1.4
PL-2383wp	1,776,726	8.5	1,849,906	2.4
PL-2736sp	1,278,926	4.8	1,303,958	1.7
PL-2737sop	765,184	5.5	775,672	1.6
PL-2746wop	1,180,352	5.1	1,203,821	1.7
PL-2746wp	1,586,226	5.5	1,626,418	1.7
PL-3012wp	2,499,097	5.9	2,571,422	2.0
PL-3120sp	2,080,418	6.2	2,131,258	2.2
PEGASE-89	5,744	0.5	5,810	0.4
PEGASE-1354	73,102	3.4	73,999	1.5
PEGASE-2869	132,520	9.0	133,869	2.7
PEGASE-9241	306,050	35.3	309,309	10.0
PEGASE-9241R	312,682	36.7	315,411	5.4

equations, we use the algorithm in [81], which is slightly different than that described in Section 7.3. Rather than consider each constraint individually, we use the mismatch in apparent power injections at each bus rather than the active and reactive power injection equations separately. The relaxation orders d_i associated with all constraints at a bus are changed together.

Specifically, mismatches for the active and reactive power injection constraints at bus i , denoted as $P_i^{\text{inj mis}}$ and $Q_i^{\text{inj mis}}$, are calculated using (7.49b). Problematic buses are identified as those with large apparent power injection mismatch $S_i^{\text{inj mis}} = |P_i^{\text{inj mis}} + \mathbf{i}Q_i^{\text{inj mis}}|$. Application of the higher-order relaxation's constraints to these problematic buses using the iterative algorithm described in [81] (cf Section 7.3) results in global solutions to many optimal power flow problems and enables computational scaling to systems with thousands of buses [81, 84]. This section extends this approach to the complex hierarchy.

Tables 7.4 and 7.5 show the results of applying the algorithm from [81] for both the real and complex hierarchies to several test cases with tolerances $\epsilon_g = 1$ MVA and $\epsilon_f = 0.05\%$.¹³ The optimal objective values in these tables match to at least 0.007%, which is within the expected solver tolerance. Further, the solutions for both the real and complex hierarchies match the optimal objective values for the loss minimizing problems obtained from MATPOWER shown in Table 7.1 to within 0.013%, providing an additional numerical proof that these solutions are globally optimal. Note, however, that local solvers do not

¹³The algorithm in [81] has a parameter h specifying the maximum number of buses to increase the relaxation order d_i at each iteration. This parameter is set to two for these examples. Additionally, bounds on the lifted variables y derived from the voltage magnitude limits are enforced to improve numeric convergence.

always globally solve optimal power flow problems [22, 26, 81].

The test cases considered in Tables 7.4 and 7.5 minimize active power losses rather than generation costs. Although the moment/sum-of-squares hierarchy solves many small- and medium-size test cases which minimize generation cost, application of the algorithm in [81] to larger generation-cost-minimizing test cases often requires too many higher-order constraints for tractability. See [72, 83, 84] for related algorithms which often find feasible points that are nearly globally optimal for such problems.

The feasible set of the optimal power flow problem is included in the ball of radius $\sum_{k \in \mathcal{B}} (v_k^{\max})^2$ so a slack variable and a sphere constraint may be added as suggested in Section 7.3. In order to preserve sparsity, a slack variable and a sphere constraint may be added *for each maximal clique* of the chordal extension of the network graph. Global convergence is then guaranteed due to (7.32). However, the sphere constraint tends to introduce numerical convergence challenges in problems with several thousand buses, resulting in the need for higher-order constraints at more buses and correspondingly longer solver times.

Interestingly, the examples in Table 7.5 converged without the slack variables and sphere constraints, and the results therein correspond to relaxations without sphere constraints. A potential way to account for the success of the complex hierarchy without sphere constraints would be to compute the Hermitian complexity [38] of the ideal generated by the polynomials associated with equality constraints. A step in that direction would be to assess the greatest number of distinct points (possibly infinite) $v^i \in \mathbb{C}^n$, $1 \leq i \leq p$, such that $(v^i)^H (H_k + \mathbf{i}\tilde{H}_k) v^j = -p_k^{\text{dem}} - \mathbf{i}q_k^{\text{dem}}$ for all buses k not connected to a generator and for all $1 \leq i, j \leq p$. Note that the Hermitian complexity of the ideal generated by $\sum_{i=1}^n |z_i|^2 + \sigma(z) + a$ as defined in (7.32) with $a < 0$ is equal to 1.

Despite being unnecessary for convergence of the hierarchies in Table 7.5, the sphere constraint can tighten the relaxations of some optimal power flow problems. Consider, for instance, the 9-bus example in [22]. The dense second-order relaxations from the real and complex hierarchies (both with and without the sphere constraint) yield the global optimum of \$3088/hour. Likewise, with second-order constraints enforced at all buses, the sparse versions of the real hierarchy and the complex hierarchy with the sphere constraint yield the global optimum. However, the sparse version of the second-order complex hierarchy without the sphere constraint only provided a lower bound of \$2939/hour. Thus, the sphere constraint tightens the sparse version of the second-order complex hierarchy for this test case. Since the sparse version of the third-order complex hierarchy without the sphere constraint yields the global optimum, the sphere constraint is unnecessary for convergence in this example.

Similar to the second-order conic relaxation, the results in Tables 7.4 and 7.5 show that the complex hierarchy generally has computational advantages over the real hierarchy. For all the test cases except PEGASE-1354, $\text{MSOS}_d\text{-}\mathbb{C}$ solves between a factor of 1.31 and 21.42 faster than $\text{MSOS}_d\text{-}\mathbb{R}$. The most significant computational speed improvements for the complex hierarchy over the real hierarchy are seen for cases (e.g., PL-2383wp and PL-2746wop) where the higher-order constraints account for a large portion of the solver times. The complex

Table 7.4: Real Moment/Sum-of-Squares Hierarchy $\text{MSOS}_d\text{-}\mathbb{R}$ (Active Power Loss Minimization)

Case Name	Num. Iter.	Global Obj. Val. (MW)	Max S^{mis} (MVA)	Solver Time (sec)
PL-2383wp	3	24,990	0.25	583.4
PL-2736sp	1	18,334	0.39	44.0
PL-2737sop	1	11,397	0.45	52.4
PL-2746wop	2	19,210	0.28	2,662.4
PL-2746wp	1	25,267	0.40	45.9
PL-3012wp	5	27,642	1.00	318.7
PL-3120sp	7	21,512	0.77	386.6
PEGASE-1354	5	74,043	0.85	406.9
PEGASE-2869	6	133,944	0.63	921.3

hierarchy for these cases has significantly fewer terms in the higher-order constraints than the real hierarchy.

Observe that several of the test cases (PL-3012wp, PL-3120sp, PEGASE-1354, and PEGASE-2869) require more iterations of the algorithm from [81] for $\text{MSOS}_d\text{-}\mathbb{C}$ than for $\text{MSOS}_d\text{-}\mathbb{R}$. Nevertheless, the improved speed per iteration results in faster overall solution times for all of these test cases except for PEGASE-1354, for which six additional iterations result in a factor of 2.78 slower solver time.

Both hierarchies were also applied to a variety of small test cases (less than ten buses) from [22, 67, 78, 85] for which the first-order relaxations failed to yield the global optima. For all these test cases, the dense versions of both $\text{MSOS}_d\text{-}\mathbb{C}$ and $\text{MSOS}_d\text{-}\mathbb{R}$ converged at the same relaxation order. Section 7.3 demonstrates that the $\text{MSOS}_d\text{-}\mathbb{R}$ is at least as tight as $\text{MSOS}_d\text{-}\mathbb{C}$. The results for small problems suggest that the hierarchies have the same tightness for some class of polynomial optimization problems which includes the optimal power flow problem with the sphere constraint (cf Conjecture 7.19). The numerical results for some large test cases have different numbers of iterations between the real and complex hierarchies. Rather than differences in the theoretical tightness of the relaxation hierarchies, we attribute this discrepancy in the number of iterations to numerical convergence inaccuracies; not enforcing the sphere constraint for the sparse complex hierarchy; and, in some cases, the algorithm from [81] selecting different buses at which to enforce the higher-order constraints.

7.5 Conclusion

We construct a complex moment/sum-of-squares hierarchy for complex polynomial optimization and prove convergence toward the global optimum. Theoretical and experimental evidence suggest that relaxing nonconvex constraints before converting from complex to real

Table 7.5: Complex Moment/Sum-of-Squares Hierarchy MSOS_d-C (Active Power Loss Minimization)

Case Name	Num. Iter.	Global Obj. Val. (MW)	Max S^{mis} (MVA)	Solver Time (sec)
PL-2383wp	3	24,991	0.10	53.9
PL-2736sp	1	18,335	0.11	17.8
PL-2737sop	1	11,397	0.07	25.7
PL-2746wop	2	19,212	0.12	124.3
PL-2746wp	1	25,269	0.05	18.5
PL-3012wp	7	27,644	0.91	141.0
PL-3120sp	9	21,512	0.27	193.9
PEGASE-1354	11	74,042	1.00	1,132.6
PEGASE-2869	9	133,939	0.97	700.8

numbers is better than doing the operations in the opposite order. We conclude with the question: is it possible to gain efficiency by transposing convex optimization algorithms from real to complex numbers?

This chapter contains several appendices that may be found after Chapter 8.

Chapter 8

Conclusion and perspectives

The main challenge that prompted this doctoral project was to be able to provide global solutions to the optimal power flow problem using semidefinite programming when the Shor relaxation fails. Having realized that the Lasserre hierarchy offers a solution to this challenge for small networks, the goal of the dissertation became to apply the Lasserre hierarchy to solve large-scale networks. The main contribution was to adapt the Lasserre hierarchy to the complex structure of our problem to enhance its tractability. This yielded a new general approach, the complex moment/sum-of-squares hierarchy.

In Chapter 2, it was shown that the Lasserre hierarchy solves small-scale networks to global optimality. These networks could not be solved using the Shor relaxation. Surprisingly, the hierarchy solves them for low orders, generally the second or third order. However, the second order relaxation can only be applied to about a dozen of variables. With more variables, it becomes intractable.

In Chapter 3, it was proven that there is no duality gap at each order of the Lasserre hierarchy provided one of the constraints is a ball constraint. This result is relevant because Lasserre proposes to add a redundant ball constraint to bounded feasible sets in order to guarantee convergence of the hierarchy. As a corollary, we obtained that there is no duality gap at each order of the hierarchy applied to the optimal power flow problem, without having to add a redundant ball constraint. Note that the ball constraint is not needed in the case of our problem of interest due to upper bound constraints on the variables, that is to say upper voltage bounds. The property we've proven is necessary for interior-point solvers to converge to solutions of the semidefinite programs in the Lasserre hierarchy.

In Chapter 4, new large-scale test cases are presented. They correspond to sections of the European high-voltage transmission network and the entire network. They can be viewed as quadratically-constrained quadratic programs where the variables and data are complex numbers. They consist in sparse problems with several thousand complex variables, with 9,241 variables in the biggest test case. The new data are representative of the size and complexity of real world power systems. They can hence be used to validate new methods and tools, such as those developed in this dissertation.

In Chapter 5, the Lasserre hierarchy is applied to large-scale networks by combining it with a penalization approach. As a result, nearly global solutions are found to generation cost minimization problems, with a guarantee of how far the value is from the global value. For power loss minimization problems, the objective function is convex. In those cases, the Lasserre hierarchy finds the global solution. In all cases, the sparsity of the problem is exploited using the notions of chordal graph and maximal clique, as well as a technique to identify problematic constraints. Higher-orders of the Lasserre hierarchy are then only applied to those constraints, reducing computation time. Moreover, low impedance lines are removed to cope with the inherent bad conditioning of power systems data.

In Chapter 6, a method for finding nearly global solutions to the optimal power flow problem is proposed. It does so without having to specify a parameter, which is a major disadvantage of penalization approaches. It is inspired by successful penalizations of the optimal power flow, which we observed to be linked with Laplacian matrices of the graph of the power network. Minimizing a quadratic form defined by such a Laplacian matrix over the power flow equations promotes low rank solutions. In fact, by iterative update of the weights of the Laplacian matrix, the rank can be reduced to one for many large-scale test cases. To guarantee near global optimality, the original objective function is set as a constraint. It is constrained to be less than or equal to the lower bound obtained by the Shor relaxation, plus a small fraction of it. This is founded because in all practical test cases, the Shor relaxation computes a lower bound of very high quality.

In Chapter 7, the Lasserre hierarchy is transposed to complex numbers in order to reduce the computational burden when solving polynomial problems with complex data and variables. The motivation for this is that the optimal power flow problem is a special case of complex polynomial optimization. We introduce a complex hierarchy and prove its convergence to the global solution for any complex polynomial problem with a feasible set of known radius. The proof relies on recent developments in algebraic geometry. The global solution to problems with several thousand complex variables is retrieved with the complex hierarchy. Sparsity is exploited by using chordal graphs techniques and a mismatch procedure to identify problematic constraints.

There are various future research directions as a result of this dissertation. One direction is to enhance the tractability of the complex moment/sum-of-squares hierarchy. A way to accomplish this may be to develop a solver in complex numbers. Interior point solvers involve Cholesky factorizations, and Cholesky factorizations could be carried out on the Hermitian matrices. Another speed-up could come from developing a randomized complex hierarchy. This is based on an idea proposed by Lasserre. Two polynomials are equal to one another with high probability if they are equal on a randomly generated set of points. Thus far, the complex hierarchy is only able to solve optimal power flow problems which minimize active power loss, which is convex function of voltage. By enhancing the tractability of the complex hierarchy, it will hopefully be possible to tackle more general objective functions such as generation cost minimization or minimum deviation from a generation plan.

The complex hierarchy entails a trade-off. It is more tractable than the real hierarchy at a given order, but provides a potentially lower bound. It would be interesting to know when the bounds generated by both hierarchies are the same. That would correspond to the cases for which it is certainly advantageous to use the complex hierarchy. In the case of the optimal power flow problem, numerical results show it is advantageous. It would be desirable to better understand why this is so.

Another research direction is to answer the following question. Do the power flow equations possess the Quillen property? In other words, is the complex hierarchy guaranteed to converge without having to add a slack variable and a redundant sphere constraint? Numerical experiments seem to show that this is true, but it is not clear why.

Lastly, transmission system operators are interested in optimization tools that cope with discrete variables. Indeed, there are many decisions which must be made from a finite number of possibilities: unit commitment, tap of phase-shifting transformers, and changes in network topology. The framework of real and complex polynomial optimization encompasses such cases, so real and complex hierarchies are relevant from a theoretical perspective. The Lasserre hierarchy is known to provide the best bounds to hard combinatorial problems, so it makes sense to try to apply real and complex hierarchies to the optimal power flow problem with discrete variables.

Appendix A

Ring Homomorphism

It is shown here that the application Λ defined by (7.3) is a ring homomorphism.

Let I_p denote the identity matrix of order $p \in \mathbb{N}$. $\Lambda(I_n) = I_{2n}$ and if $Z_1, Z_2 \in \mathbb{C}^{n \times n}$, $\Lambda(Z_1 + Z_2) = \Lambda(Z_1) + \Lambda(Z_2)$ and

$$\begin{aligned}\Lambda(Z_1)\Lambda(Z_2) &= \begin{pmatrix} \operatorname{Re}Z_1 & -\operatorname{Im}Z_1 \\ \operatorname{Im}Z_1 & \operatorname{Re}Z_1 \end{pmatrix} \begin{pmatrix} \operatorname{Re}Z_2 & -\operatorname{Im}Z_2 \\ \operatorname{Im}Z_2 & \operatorname{Re}Z_2 \end{pmatrix} \\ &= \begin{pmatrix} \operatorname{Re}Z_1\operatorname{Re}Z_2 - \operatorname{Im}Z_1\operatorname{Im}Z_2 & -\operatorname{Re}Z_1\operatorname{Im}Z_2 - \operatorname{Im}Z_1\operatorname{Re}Z_2 \\ \operatorname{Im}Z_1\operatorname{Re}Z_2 + \operatorname{Re}Z_1\operatorname{Im}Z_2 & \operatorname{Re}Z_1\operatorname{Re}Z_2 - \operatorname{Im}Z_1\operatorname{Im}Z_2 \end{pmatrix} \\ &= \Lambda[\operatorname{Re}Z_1\operatorname{Re}Z_2 - \operatorname{Im}Z_1\operatorname{Im}Z_2 + \mathbf{i}(\operatorname{Im}Z_1\operatorname{Re}Z_2 + \operatorname{Re}Z_1\operatorname{Im}Z_2)] \\ &= \Lambda[(\operatorname{Re}Z_1 + \mathbf{i}\operatorname{Im}Z_1)(\operatorname{Re}Z_2 + \mathbf{i}\operatorname{Im}Z_2)] \\ &= \Lambda(Z_1Z_2).\end{aligned}$$

Appendix B

Rank-2 Condition

It is proven here that a Hermitian matrix Z is positive semidefinite and has rank 1 if and only if $\Lambda(Z)$ is positive semidefinite and has rank 2.

(\implies) Say $Z = zz^H$ where real and imaginary parts are defined by $z = x_1 + \mathbf{i}x_2$ and $(x_1, x_2) \neq (0, 0)$. Then

$$\Lambda(Z) = \begin{pmatrix} x_1x_1^T + x_2x_2^T & x_1x_2^T - x_2x_1^T \\ x_2x_1^T - x_1x_2^T & x_1x_1^T + x_2x_2^T \end{pmatrix} \quad (\text{B.1a})$$

$$= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T + \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}^T. \quad (\text{B.1b})$$

The rank of $\Lambda(Z)$ is equal to 2 since $(x_1^T \ x_2^T)^T$ and $(-x_2^T \ x_1^T)^T$ are non-zero orthogonal vectors.

(\impliedby) Say $\Lambda(Z) = xx^T + yy^T$ where x and y are non-zero real vectors. Consider the block structure $x = (x_1^T \ x_2^T)^T$ and $y = (y_1^T \ y_2^T)^T$. For $i = 1, \dots, n$, it must be that

$$x_{1i}^2 + y_{1i}^2 = x_{2i}^2 + y_{2i}^2, \quad (\text{B.2a})$$

$$x_{1i}x_{2i} + y_{1i}y_{2i} = 0. \quad (\text{B.2b})$$

Two cases can occur. The first is that $x_{1i}x_{2i} \neq 0$ in which case there exists a real number $\lambda_i \neq 0$ such that

$$\begin{cases} y_{1i} &= -\lambda_i x_{2i}, \\ y_{2i} &= \frac{1}{\lambda_i} x_{1i}. \end{cases} \quad (\text{B.3})$$

Equation (B.2a) implies that $(1 - \lambda_i^2)x_{1i}^2 = (1 - \frac{1}{\lambda_i^2})x_{2i}^2$ thus $(1 - \lambda_i^2)(1 - \frac{1}{\lambda_i^2}) \geq 0$ and $\lambda_i = \pm 1$. The second case is that $x_{1i}x_{2i} = 0$. Then, according to (B.2b), $y_{1i}y_{2i} = 0$. If either $x_{1i} = y_{1i} = 0$ or $x_{2i} = y_{2i} = 0$, then (B.2a) implies that $x_{1i} = x_{2i} = y_{1i} = y_{2i} = 0$. If $x_{1i} = y_{2i} = 0$, then (B.2a) implies that $y_{1i} = \pm x_{2i}$. If $x_{2i} = y_{1i} = 0$, then (B.2a) implies that $y_{2i} = \pm x_{1i}$.

In any case, there exists $\epsilon_i = \pm 1$ such that

$$\begin{cases} y_{1i} &= -\epsilon_i x_{2i}, \\ y_{2i} &= \epsilon_i x_{1i}. \end{cases} \quad (\text{B.4})$$

For $i, j = 1, \dots, n$ it must be that

$$(1 - \epsilon_i \epsilon_j)(x_{1i}x_{1j} - x_{2i}x_{2j}) = 0, \quad (\text{B.5a})$$

$$(1 - \epsilon_i \epsilon_j)(x_{1j}x_{2i} + x_{1i}x_{2j}) = 0. \quad (\text{B.5b})$$

Moreover

$$\begin{cases} x_{1i}x_{1j} + y_{1i}y_{1j} &= x_{1i}x_{1j} + \epsilon_i \epsilon_j x_{2i}x_{2j}, \\ x_{1i}x_{2j} + y_{1i}y_{2j} &= x_{1i}x_{2j} - \epsilon_i \epsilon_j x_{2i}x_{1j}. \end{cases} \quad (\text{B.6})$$

It will now be shown that

$$\begin{cases} x_{1i}x_{1j} + y_{1i}y_{1j} &= x_{1i}x_{1j} + x_{2i}x_{2j}, \\ x_{1i}x_{2j} + y_{1i}y_{2j} &= x_{1i}x_{2j} - x_{2i}x_{1j}. \end{cases} \quad (\text{B.7})$$

It is obvious if $\epsilon_i \epsilon_j = 1$. If $\epsilon_i \epsilon_j = -1$, then (B.5a)–(B.5b) imply

$$x_{1i}x_{1j} - x_{2i}x_{2j} = 0, \quad (\text{B.8a})$$

$$x_{1j}x_{2i} + x_{1i}x_{2j} = 0. \quad (\text{B.8b})$$

If $x_{1i}x_{1j}x_{2i}x_{2j} = 0$, it can be seen that (B.7) holds. If not, (B.8a) implies that there exists a real number $\mu_{ij} \neq 0$ such that

$$\begin{cases} x_{2i} &= \mu_{ij} x_{1i}, \\ x_{2j} &= \frac{1}{\mu_{ij}} x_{1j}. \end{cases} \quad (\text{B.9})$$

Further, (B.8b) implies that $(\mu_{ij} + \frac{1}{\mu_{ij}})x_{1j}x_{2i} = 0$. This is impossible ($\mu_{ij} + \frac{1}{\mu_{ij}} \neq 0$ and $x_{1j}x_{2i} \neq 0$). Thus, (B.7) holds.

With the left hand side corresponding to $\Lambda(Z) = xx^T + yy^T$ and the right hand side corresponding to (B.1b), equation (B.7) implies that $\Lambda(Z)$ is equal to (B.1b). Since the function Λ is injective, it must be that $Z = (x_1 + \mathbf{i}x_2)(x_1 + \mathbf{i}x_2)^H$.

Appendix C

Invariance of Shor Relaxation Bound

It is shown here that the Shor relaxation bound obtained by relaxing nonconvexities then converting from complex to real numbers is the same as that obtained by converting from complex to real number then relaxing nonconvexities.

We have $\text{val}(\text{CSDP-}\mathbb{R}) \geq \text{val}(\text{SDP-}\mathbb{R})$ since the feasible set is more tightly constrained due to (7.4d). To prove the opposite inequality, define $\tilde{\Lambda}(X) := (A + C)/2 + \mathbf{i}(B - B^T)/2$ for all $X \in \mathbb{S}_{2n}$ using the block decomposition in the left hand part of (7.4d). It is proven here that if X is a feasible point of $\text{SDP-}\mathbb{R}$, then $\Lambda \circ \tilde{\Lambda}(X)$ is a feasible point of $\text{CSDP-}\mathbb{R}$ with same objective value as X . Firstly, $\Lambda \circ \tilde{\Lambda}(X)$ satisfies (7.4d) because $\tilde{\Lambda}(X)$ is a Hermitian matrix. Secondly, in order to show that $\Lambda \circ \tilde{\Lambda}(X)$ satisfies (7.4c), notice that if $x = (x_1^T \ x_2^T)^T$ then

$$\begin{aligned} & \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} C & -B \\ -B^T & A \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ & \qquad \qquad \qquad = \\ & \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}^T \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}. \end{aligned} \tag{C.1}$$

Hence $\Lambda \circ \tilde{\Lambda}(X)$ is equal to the sum of two positive semidefinite matrices. Finally, to prove that $\Lambda \circ \tilde{\Lambda}(X)$ satisfies (7.4b) and has same objective value as X , notice that if $H \in \mathbb{H}_n$ and $Y \in \mathbb{S}_{2n}$, then $\text{Tr}[\Lambda(H)Y] = \sum_{1 \leq i, j \leq 2n} \Lambda(H)_{ij} Y_{ji} = \sum_{1 \leq i, j \leq 2n} \Lambda(H)_{ij} Y_{ij} = \sum_{1 \leq i, j \leq n} \text{Re}(H)_{ij} A_{ij} + \text{Im}(H)_{ij} B_{ij} + (-\text{Im}(H)_{ij})(B^T)_{ij} + \text{Re}(H)_{ij} C_{ij} = \sum_{1 \leq i, j \leq n} \text{Re}(H_{ij})(A + C)_{ij} + \text{Im}(H_{ij})(B - B^T)_{ij} = 2 \sum_{1 \leq i, j \leq n} \text{Re}[H_{ij}(\tilde{\Lambda}(Y)_{ij})^H] = 2 \sum_{1 \leq i, j \leq n} H_{ij}(\tilde{\Lambda}(Y)_{ij})^H = 2\text{Tr}[H\tilde{\Lambda}(Y)]$. Completing the proof, for all $H \in \mathbb{H}_n$, $\text{Tr}[\Lambda(H) \Lambda \circ \tilde{\Lambda}(X)] = 2\text{Tr}[H\tilde{\Lambda}(X)] = \text{Tr}[\Lambda(H)X]$.

Appendix D

Invariance of SDP- \mathbb{R} Relaxation Bound

We consider the semidefinite problem obtained by converting from complex to real numbers then relaxing nonconvexities. It is proven here that setting the phase of one of the variables to zero does not affect the relaxation bound.

We assume that X is a feasible point of SDP- \mathbb{R} and construct a feasible point of SDP- \mathbb{R} with same objective value and first diagonal entry equal to 0. Consider the eigenvalue decomposition $X = \sum_{k=1}^p x_k x_k^T$ for some $x_k \in \mathbb{R}^{2n}$ and $p \in \mathbb{N}$. For all $\theta \in \mathbb{R}$, define

$$R_\theta := \Lambda[\cos(\theta)I_n + \mathbf{i}\sin(\theta)I_n] = \begin{pmatrix} \cos(\theta)I_n & -\sin(\theta)I_n \\ \sin(\theta)I_n & \cos(\theta)I_n \end{pmatrix}. \quad (\text{D.1})$$

For $k = 1, \dots, p$, define $\theta_k \in \mathbb{R}$ such that $x_{k,n+1} + \mathbf{i}x_{k,1} =: \sqrt{x_{k,n+1}^2 + x_{k,1}^2} e^{\mathbf{i}\theta_k}$. Construct $\tilde{X} := \sum_{k=1}^p (R_{\theta_k} x_k)(R_{\theta_k} x_k)^T \succcurlyeq 0$ whose first diagonal entry is equal to 0. If $H \in \mathbb{H}_n$, then $\text{Tr}(\Lambda(H)\tilde{X}) = \sum_{k=1}^p \text{Tr}[\Lambda(H)R_{\theta_k} x_k x_k^T R_{\theta_k}^T] = \sum_{k=1}^p \text{Tr}[R_{\theta_k}^T \Lambda(H) R_{\theta_k} x_k x_k^T] = \sum_{k=1}^p \text{Tr}[\Lambda\{(\cos(\theta_k)I_n - \mathbf{i}\sin(\theta_k)I_n)H(\cos(\theta_k)I_n + \mathbf{i}\sin(\theta_k)I_n)\} x_k x_k^T] = \sum_{k=1}^p \text{Tr}[\Lambda(H)x_k x_k^T] = \text{Tr}(\Lambda(H)X)$.

Appendix E

Discrepancy Between Second-Order Conic Relaxation Bounds

It is shown here that the second-order conic relaxation bound obtained by relaxing nonconvexities then converting from complex to real numbers is different from that obtained by converting from complex to real number then relaxing nonconvexities.

We have $\text{val}(\text{CSOCP-}\mathbb{R}) \geq \text{val}(\text{SOCP-}\mathbb{R})$ since the feasible set is more tightly constrained. The opposite inequality between optimal values does not hold, and this can be proven by considering the example QCQP- \mathbb{C} defined by $\inf_{z_1, z_2 \in \mathbb{C}} (1 + \mathbf{i})\bar{z}_1 z_2 + (1 - \mathbf{i})\bar{z}_2 z_1$ s.t. $\bar{z}_1 z_1 \leq 1$, $\bar{z}_2 z_2 \leq 1$. CSOCP- \mathbb{R} yields the globally optimal value of $-2\sqrt{2}$, while SOCP- \mathbb{R} yields -4 , as can be seen below.

$$\begin{aligned}
 -2\sqrt{2} &= \inf_{X \in \mathbb{S}_4} 2X_{12} + 2X_{34} + 2X_{23} - 2X_{14}, \\
 \text{s.t.} & \quad X_{11} + X_{33} \leq 1, \quad X_{22} + X_{44} \leq 1, \\
 & \quad X_{12}^2 + X_{23}^2 \leq X_{11}X_{22}, \\
 & \quad X_{11} = X_{33}, \quad X_{12} = X_{34}, \quad X_{22} = X_{44}, \\
 & \quad X_{13} = X_{24} = 0, \quad X_{23} + X_{14} = 0,
 \end{aligned} \tag{E.1}$$

$$\begin{aligned}
 -4 &= \inf_{X \in \mathbb{S}_4} 2X_{12} + 2X_{34} + 2X_{23} - 2X_{14}, \\
 \text{s.t.} & \quad X_{11} + X_{33} \leq 1, \quad X_{22} + X_{44} \leq 1, \\
 & \quad X_{12}^2 \leq X_{11}X_{22}, \quad X_{13}^2 \leq X_{11}X_{33}, \\
 & \quad X_{14}^2 \leq X_{11}X_{44}, \quad X_{23}^2 \leq X_{22}X_{33}, \\
 & \quad X_{24}^2 \leq X_{22}X_{44}, \quad X_{34}^2 \leq X_{33}X_{44}.
 \end{aligned}$$

Appendix F

Five-Bus Illustrative Example for Exploiting Sparsity

To illustrate the selective application of second-order constraints, consider the five-bus optimal power flow problem in [22] which is an instance of QCQP-C. Let $\text{ind}(\cdot)$ denote the set of indices corresponding to monomials of either the objective f or constraint functions $(g_i)_{1 \leq 20}$. We have

$$\begin{aligned}
 \text{ind}(f) &= \{(1, 1), (1, 2), (1, 3), (3, 5), (4, 5), (5, 5)\}, \\
 \text{ind}(g_1) = \text{ind}(g_2) &= \{(1, 1), (1, 2), (1, 3)\} && [P_1^{\min}, Q_1^{\min}], \\
 \text{ind}(g_3) = \text{ind}(g_4) &= \{(1, 2), (2, 2), (2, 3), (2, 4)\} && [P_2, Q_2], \\
 \text{ind}(g_5) = \text{ind}(g_6) &= \{(1, 3), (2, 3), (3, 3), (3, 5)\} && [P_3, Q_3], \\
 \text{ind}(g_7) = \text{ind}(g_8) &= \{(2, 4), (4, 4), (4, 5)\} && [P_4, Q_4], \\
 \text{ind}(g_9) = \text{ind}(g_{10}) &= \{(3, 5), (4, 5), (5, 5)\} && [P_5^{\min}, Q_5^{\min}], \\
 \text{ind}(g_{11}) = \text{ind}(g_{12}) &= \{(1, 1)\} && [V_1^{\min}, V_1^{\max}], \\
 \text{ind}(g_{13}) = \text{ind}(g_{14}) &= \{(2, 2)\} && [V_2^{\min}, V_2^{\max}], \\
 \text{ind}(g_{15}) = \text{ind}(g_{16}) &= \{(3, 3)\} && [V_3^{\min}, V_3^{\max}], \\
 \text{ind}(g_{17}) = \text{ind}(g_{18}) &= \{(4, 4)\} && [V_4^{\min}, V_4^{\max}], \\
 \text{ind}(g_{19}) = \text{ind}(g_{20}) &= \{(5, 5)\} && [V_5^{\min}, V_5^{\max}],
 \end{aligned} \tag{F.1}$$

where the text in brackets indicates the origin of the constraint: P_i and Q_i for active and reactive power injection equality constraints, P_i^{\min} and Q_i^{\min} for lower limits on active and reactive power injections, and V_i^{\min} and V_i^{\max} for squared voltage magnitude limits at bus i .

For brevity, the sphere constraints discussed in Section 7.3 are not enforced in this example. Regardless, the complex hierarchy with $d_i = 1, \forall i \in \{1, 2, 3, 4, 5, 6, 11, 12, 13, 14, 15, 16\}$, $d_i = 2, \forall i \in \{7, 8, 9, 10, 17, 18, 19, 20\}$ converges to the global solution. The second-order constraints are identified using the maximum power injection mismatch heuristic in [81].

The graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ corresponding to (F.1) is shown in Fig. F.1. The nodes correspond to the complex variables $\mathcal{N} = \{1, \dots, 5\}$. Edges \mathcal{E} , which are denoted by solid lines in Fig. F.1, connect variables that appear in the same monomial in any of the constraint equations or objective function. The supergraph $\hat{\mathcal{G}} = (\mathcal{N}, \hat{\mathcal{E}})$ has edges $\hat{\mathcal{E}}$ comprised of \mathcal{E} (solid lines in Fig. F.1) augmented with edges connecting all variables within each constraint with $d_i > 1$ (dashed lines in Fig. F.1). In this case, $\hat{\mathcal{G}}$ is already chordal, so there is no need to form a chordal extension \mathcal{G}^{ch} .

The maximal cliques of $\hat{\mathcal{G}}$ are $\mathcal{C}_1 = \{1, 2, 3\}$ and $\mathcal{C}_2 = \{2, 3, 4, 5\}$. Clique \mathcal{C}_2 is the minimal covering clique for all second-order constraints $g_i(z)$, $\forall i \in \{7, 8, 9, 10, 17, 18, 19, 20\}$. The order associated with \mathcal{C}_2 is two ($\tilde{d}_2 = 2$) since the highest order d_i among all constraints for which \mathcal{C}_2 is the minimal covering clique is two. Clique \mathcal{C}_1 is not the minimal covering clique for any constraints with $d_i > 1$, so $\tilde{d}_1 = 1$.

The globally optimal objective value obtained from the complex hierarchy specified above is 946.8 with corresponding decision variable $z = (1.0467 + 0.0000\mathbf{i}, 0.9550 - 0.0578\mathbf{i}, 0.9485 - 0.0533\mathbf{i}, 0.7791 + 0.6011\mathbf{i}, 0.7362 + 0.7487\mathbf{i})^T$.

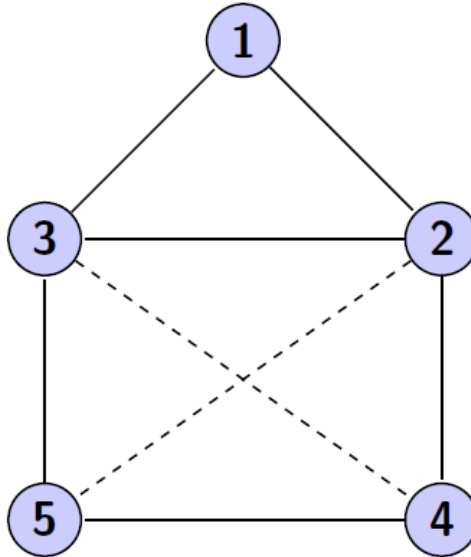


Figure F.1: Graph Corresponding to Equations (F.1) from Five-Bus System in [22]

Appendix G

Complex Hierarchy Applied to Optimal Power Flow

We consider an example of power loss minimization. The system of Figure G.1 links a generator to a load via a line of admittance $g + \mathbf{j}b$ while respecting upper voltage constraints.

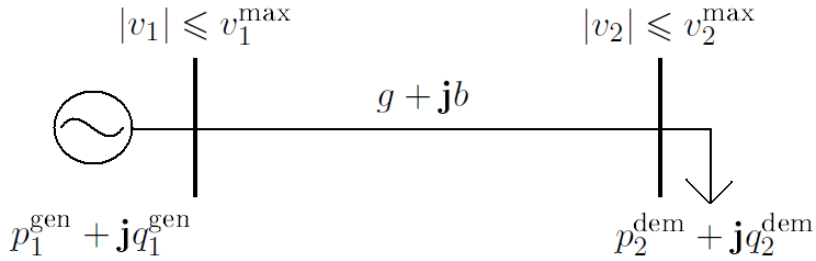


Figure G.1: Two-Bus System

Minimizing power loss reads

$$\inf_{v_1, v_2 \in \mathbb{C}} g |v_1|^2 - g \bar{v}_1 v_2 - g \bar{v}_2 v_1 + g |v_2|^2, \quad (\text{G.1})$$

subject to

$$-\frac{g - \mathbf{j}b}{2} \bar{v}_1 v_2 - \frac{g + \mathbf{j}b}{2} \bar{v}_2 v_1 + g |v_2|^2 = -p_2^{\text{dem}}, \quad (\text{G.2})$$

$$\frac{b + \mathbf{j}g}{2} \bar{v}_1 v_2 + \frac{b - \mathbf{j}g}{2} \bar{v}_2 v_1 - b |v_2|^2 = -q_2^{\text{dem}}, \quad (\text{G.3})$$

$$|v_1|^2 \leq (v_1^{\max})^2, \quad (\text{G.4})$$

$$|v_2|^2 \leq (v_2^{\max})^2. \quad (\text{G.5})$$

The feasible set is included in the ball defined by $|v_1|^2 + |v_2|^2 \leq (v_1^{\max})^2 + (v_2^{\max})^2$. In

accordance with Section 7.3, let's add a slack variable $v_3 \in \mathbb{C}$ and a constraint

$$|v_1|^2 + |v_2|^2 + |v_3|^2 = (v_1^{\max})^2 + (v_2^{\max})^2. \quad (\text{G.6})$$

The first and second orders (i.e., MSOS₁- \mathbb{C} and MSOS₂- \mathbb{C}) are written below where the notation $y_\beta^\alpha := y_{\alpha,\beta}$ ($\alpha, \beta \in \mathbb{N}^3$) is used to save space.

Example of MSOS₁- \mathbb{C} :

$$\inf_y g y_{100}^{100} - g y_{010}^{100} - g y_{100}^{010} + g y_{010}^{010}, \quad (\text{G.7})$$

subject to

$$-\frac{g - \mathbf{i}b}{2} y_{010}^{100} - \frac{g + \mathbf{i}b}{2} y_{100}^{010} + g y_{010}^{010} = -p_2^{\text{dem}} y_{000}^{000}, \quad (\text{G.8})$$

$$\frac{b + \mathbf{i}g}{2} y_{010}^{100} + \frac{b - \mathbf{i}g}{2} y_{100}^{010} - b y_{010}^{010} = -q_2^{\text{dem}} y_{000}^{000}, \quad (\text{G.9})$$

$$y_{100}^{100} \leq (v_1^{\max})^2 y_{000}^{000}, \quad (\text{G.10})$$

$$y_{010}^{010} \leq (v_2^{\max})^2 y_{000}^{000}, \quad (\text{G.11})$$

$$y_{100}^{100} + y_{010}^{010} + y_{001}^{001} = ((v_1^{\max})^2 + (v_2^{\max})^2) y_{000}^{000}, \quad (\text{G.12})$$

$$\begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \succcurlyeq 0, \quad (\text{G.13})$$

$$y_{000}^{000} = 1. \quad (\text{G.14})$$

Example of MSOS₂- \mathbb{C} :

$$\inf_y g y_{100}^{100} - g y_{010}^{100} - g y_{100}^{010} + g y_{010}^{010}, \quad (\text{G.15})$$

subject to

$$p_2^{\text{dem}} \begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \cdots$$

$$\begin{aligned}
 & -\frac{g - \mathbf{i}b}{2} \begin{pmatrix} y_{010}^{100} & y_{110}^{100} & y_{020}^{100} & y_{011}^{100} \\ y_{010}^{200} & y_{110}^{200} & y_{020}^{200} & y_{011}^{200} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{101} & y_{110}^{101} & y_{020}^{101} & y_{011}^{101} \end{pmatrix} \cdots \\
 & -\frac{g + \mathbf{i}b}{2} \begin{pmatrix} y_{100}^{010} & y_{200}^{010} & y_{110}^{010} & y_{101}^{010} \\ y_{100}^{110} & y_{200}^{110} & y_{110}^{110} & y_{101}^{110} \\ y_{100}^{020} & y_{200}^{020} & y_{110}^{020} & y_{101}^{020} \\ y_{100}^{011} & y_{200}^{011} & y_{110}^{011} & y_{101}^{011} \end{pmatrix} \cdots \\
 & +g \begin{pmatrix} y_{010}^{010} & y_{110}^{010} & y_{020}^{010} & y_{011}^{010} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{020} & y_{110}^{020} & y_{020}^{020} & y_{011}^{020} \\ y_{010}^{011} & y_{110}^{011} & y_{020}^{011} & y_{011}^{011} \end{pmatrix} = 0, \tag{G.16}
 \end{aligned}$$

$$\begin{aligned}
 & q_2^{\text{dem}} \begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \cdots \\
 & +\frac{b + \mathbf{i}g}{2} \begin{pmatrix} y_{010}^{100} & y_{110}^{100} & y_{020}^{100} & y_{011}^{100} \\ y_{010}^{200} & y_{110}^{200} & y_{020}^{200} & y_{011}^{200} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{101} & y_{110}^{101} & y_{020}^{101} & y_{011}^{101} \end{pmatrix} \cdots \\
 & +\frac{b - \mathbf{i}g}{2} \begin{pmatrix} y_{100}^{010} & y_{200}^{010} & y_{110}^{010} & y_{101}^{010} \\ y_{100}^{110} & y_{200}^{110} & y_{110}^{110} & y_{101}^{110} \\ y_{100}^{020} & y_{200}^{020} & y_{110}^{020} & y_{101}^{020} \\ y_{100}^{011} & y_{200}^{011} & y_{110}^{011} & y_{101}^{011} \end{pmatrix} \cdots \\
 & -b \begin{pmatrix} y_{010}^{010} & y_{110}^{010} & y_{020}^{010} & y_{011}^{010} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{020} & y_{110}^{020} & y_{020}^{020} & y_{011}^{020} \\ y_{010}^{011} & y_{110}^{011} & y_{020}^{011} & y_{011}^{011} \end{pmatrix} = 0, \tag{G.17}
 \end{aligned}$$

$$\begin{aligned}
 & (v_1^{\max})^2 \begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \cdots \\
 & - \begin{pmatrix} y_{100}^{100} & y_{200}^{100} & y_{110}^{100} & y_{101}^{100} \\ y_{100}^{200} & y_{200}^{200} & y_{110}^{200} & y_{101}^{200} \\ y_{100}^{110} & y_{200}^{110} & y_{110}^{110} & y_{101}^{110} \\ y_{100}^{101} & y_{200}^{101} & y_{110}^{101} & y_{101}^{101} \end{pmatrix} \succcurlyeq 0, \tag{G.18}
 \end{aligned}$$

$$\begin{aligned}
 & (v_2^{\max})^2 \begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \cdots \\
 & - \begin{pmatrix} y_{010}^{010} & y_{110}^{010} & y_{020}^{010} & y_{011}^{010} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{020} & y_{110}^{020} & y_{020}^{020} & y_{011}^{020} \\ y_{010}^{011} & y_{110}^{011} & y_{020}^{011} & y_{011}^{011} \end{pmatrix} \succcurlyeq 0, \tag{G.19}
 \end{aligned}$$

$$\begin{aligned}
 & ((v_1^{\max})^2 + (v_2^{\max})^2) \begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} \end{pmatrix} \cdots \\
 & - \begin{pmatrix} y_{100}^{100} & y_{200}^{100} & y_{110}^{100} & y_{101}^{100} \\ y_{100}^{200} & y_{200}^{200} & y_{110}^{200} & y_{101}^{200} \\ y_{100}^{110} & y_{200}^{110} & y_{110}^{110} & y_{101}^{110} \\ y_{100}^{101} & y_{200}^{101} & y_{110}^{101} & y_{101}^{101} \end{pmatrix} \cdots
 \end{aligned}$$

$$\begin{aligned}
 & - \begin{pmatrix} y_{010}^{010} & y_{110}^{010} & y_{020}^{010} & y_{011}^{010} \\ y_{010}^{110} & y_{110}^{110} & y_{020}^{110} & y_{011}^{110} \\ y_{010}^{020} & y_{110}^{020} & y_{020}^{020} & y_{011}^{020} \\ y_{010}^{011} & y_{110}^{011} & y_{020}^{011} & y_{011}^{011} \end{pmatrix} \cdots \\
 & - \begin{pmatrix} y_{001}^{001} & y_{101}^{001} & y_{011}^{001} & y_{002}^{001} \\ y_{001}^{101} & y_{101}^{101} & y_{011}^{101} & y_{002}^{101} \\ y_{001}^{011} & y_{101}^{011} & y_{011}^{011} & y_{002}^{011} \\ y_{001}^{002} & y_{101}^{002} & y_{011}^{002} & y_{002}^{002} \end{pmatrix} = 0, \tag{G.20}
 \end{aligned}$$

$$\begin{pmatrix} y_{000}^{000} & y_{100}^{000} & y_{010}^{000} & y_{001}^{000} & y_{200}^{000} & y_{110}^{000} & y_{101}^{000} & y_{020}^{000} & y_{011}^{000} & y_{002}^{000} \\ y_{000}^{100} & y_{100}^{100} & y_{010}^{100} & y_{001}^{100} & y_{200}^{100} & y_{110}^{100} & y_{101}^{100} & y_{020}^{100} & y_{011}^{100} & y_{002}^{100} \\ y_{000}^{010} & y_{100}^{010} & y_{010}^{010} & y_{001}^{010} & y_{200}^{010} & y_{110}^{010} & y_{101}^{010} & y_{020}^{010} & y_{011}^{010} & y_{002}^{010} \\ y_{000}^{001} & y_{100}^{001} & y_{010}^{001} & y_{001}^{001} & y_{200}^{001} & y_{110}^{001} & y_{101}^{001} & y_{020}^{001} & y_{011}^{001} & y_{002}^{001} \\ y_{000}^{200} & y_{100}^{200} & y_{010}^{200} & y_{001}^{200} & y_{200}^{200} & y_{110}^{200} & y_{101}^{200} & y_{020}^{200} & y_{011}^{200} & y_{002}^{200} \\ y_{000}^{110} & y_{100}^{110} & y_{010}^{110} & y_{001}^{110} & y_{200}^{110} & y_{110}^{110} & y_{101}^{110} & y_{020}^{110} & y_{011}^{110} & y_{002}^{110} \\ y_{000}^{101} & y_{100}^{101} & y_{010}^{101} & y_{001}^{101} & y_{200}^{101} & y_{110}^{101} & y_{101}^{101} & y_{020}^{101} & y_{011}^{101} & y_{002}^{101} \\ y_{000}^{020} & y_{100}^{020} & y_{010}^{020} & y_{001}^{020} & y_{200}^{020} & y_{110}^{020} & y_{101}^{020} & y_{020}^{020} & y_{011}^{020} & y_{002}^{020} \\ y_{000}^{011} & y_{100}^{011} & y_{010}^{011} & y_{001}^{011} & y_{200}^{011} & y_{110}^{011} & y_{101}^{011} & y_{020}^{011} & y_{011}^{011} & y_{002}^{011} \\ y_{000}^{002} & y_{100}^{002} & y_{010}^{002} & y_{001}^{002} & y_{200}^{002} & y_{110}^{002} & y_{101}^{002} & y_{020}^{002} & y_{011}^{002} & y_{002}^{002} \end{pmatrix} \approx 0, \tag{G.21}$$

$$y_{000}^{000} = 1. \tag{G.22}$$

Bibliography

- [1] *Power Systems Test Case Archive*.
- [2] *The MOSEK Optimization Toolbox for MATLAB Manual version 7*.
- [3] T. AITTO MAKI AND V. KOIVUNEN, *Beampattern Optimization by Minimization of Quartic Polynomial*, IEEE/SP 15th W. Stat. Signal Process., 51 (2009), pp. 437—440.
- [4] N. AKHIEZER, *The Classical Moment Problem and Some Related Questions in Analysis*, Hafner Publ. Co., New York, 1965.
- [5] C. ALIPRANTIS AND K. BORDER, *Infinite Dimensional Analysis, A Hitchhiker's guide*, Second Edition, Springer-Verlag Berlin Heidelberg, 1999.
- [6] F. L. ALVARADO AND E. ELKONYAL, *Reduction in Power Systems*, IEEE PES Summer Meeting, (1977).
- [7] M. ANDERSEN, A. HANSSON, AND L. VANDENBERGHE, *Reduced-Complexity Semidefinite Relaxations of Optimal Power Flow Problems*, IEEE Trans. Power Syst., 29 (2014), pp. 1855–1863.
- [8] E. ANDERSON AND P. NASH, *Linear Programming in Infinite-Dimensional Spaces, Theory and Applications*, Wiley Int. Ser. Disc. Math. Optim., 1987.
- [9] M. F. ANJOS AND J. B. LASSERRE, *Introduction to Semidefinite, Conic and Polynomial Optimization*, in Handbook on Semidefinite, Conic and Polynomial Optimization, M. F. Anjos and J. B. Lasserre, eds., vol. 166 of Int. Ser. Oper. Res. & Manag. Sci., Springer, 2012.
- [10] E. ARTIN, *Über die Zerlegung Definiten Funktionen in Quadrate*, Abhandlungen aus dem Math. Sem. der Univ. Hamburg, 5 (1927), pp. 100–115.
- [11] A. ATZMON, *A Moment Problem for Positive Measures on the Unit Disc*, Pacific J. Math., 59 (1975), pp. 317–325.

- [12] A. AUBRY, A. DE MAIO, B. JIANG, AND S. ZHANG, *Ambiguity Function Shaping for Cognitive Radar via Complex Quartic Optimization*, IEEE Trans. Signal Process., 61 (2013), pp. 5603—5619.
- [13] X. BAI, H. WEI, K. FUJISAWA, AND Y. WANG, *Semidefinite Programming for Optimal Power Flow Problems*, Int. J. Elec. Power, 30 (2008), pp. 383–392.
- [14] A. BANDEIRA, N. BOUMAL, AND A. SINGER, *Tightness of the Maximum Likelihood Semidefinite Relaxation for Angular Synchronization*, Preprint, available at: <http://arxiv.org/abs/1411.3272>, (2014).
- [15] L. BLANCHET, *Diplôme D'Ingénieur, Conservatoire National des Arts et Métiers*, (1986).
- [16] G. BLEKHERMAN, P. PARRILO, AND R. THOMAS, *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Ser. Optim., SIAM and MPS, Philadelphia, 2013.
- [17] J. BONNANS, J. GILBERT, C. LEMARÉCHAL, AND C. SAGASTIZÁBAL, *Numerical Optimization – Theoretical and Practical Aspects* (second edition), Universitext, Springer Verlag, Berlin, 2006. [authors] [editor].
- [18] S. BOSE, D. GAYME, K. CHANDY, AND S. LOW, *Quadratically Constrained Quadratic Programs on Acyclic Graphs with Application to Power*, IEEE Trans. Contr. Network Syst., (2015).
- [19] S. BOSE, D. GAYME, S. LOW, AND K. CHANDY, *Optimal power flow over tree networks*, in 49th Annu. Allerton Conf. Commun., Control, Comput., 2011.
- [20] S. BOSE, S. LOW, T. TEERARATKUL, AND B. HASSIBI, *Equivalent Relaxations of Optimal Power Flow*, IEEE Trans. Automat. Control, (2014), p. 99.
- [21] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, 2009.
- [22] W. BUKHSH, A. GROTHEY, K. MCKINNON, AND P. TRODDEN, *Local Solutions of the Optimal Power Flow Problem*, IEEE Trans. Power Syst., 28 (2013), pp. 4780—4788.
- [23] W. BUKHSH, A. GROTHEY, K. MCKINNON, AND P. TRODDEN, *Test case archive of optimal power flow (OPF) problem with local optima*, 2013.
- [24] E. CANDÈS, Y. C. ELДАР, T. STROHMER, AND V. VORONINSKI, *Phase Retrieval via Matrix Completion*, SIAM J. Imaging Sci., 6 (2013), pp. 199—225.

- [25] M. CARPENTIER, *Contribution à l'Étude du Dispatching Économique*, Bull. de la Soc. Fran. des Élec., 8 (1962), pp. 431—447.
- [26] A. CASTILLO AND R. O'NEILL, *Computational Performance of Solution Techniques Applied to the ACOPF (OPF Paper 5)*, tech. rep., US FERC, Jan. 2013.
- [27] —, *Survey of Approaches to Solving the ACOPF (OPF Paper 4)*, tech. rep., US FERC, Mar. 2013.
- [28] D. CATLIN AND J. D'ANGELO, *A Stabilization Theorem for Hermitian Forms and Applications to Holomorphic Mappings*, Math. Res. Lett., 3 (1996), pp. 149—166.
- [29] C. CHEN AND P. VAIDYANATHAN, *MIMO Radar Waveform Optimization With Prior Information of the Extended Target and Clutter*, IEEE Trans. Signal Process., 57 (2009), pp. 3533–3544.
- [30] C. COFFRIN, H. HIJAZI, AND P. VAN HENTENRYCK, *The QC Relaxation: Theoretical and Computational Results on Optimal Power Flow*, Preprint: <http://arxiv.org/abs/1502.07847>, (2015).
- [31] R. R. COIFMAN, Y. SHKOLNISKY, F. J. SIGWORTH, AND A. SINGER, *Graph Laplacian Tomography From Unknown Random Projections*, IEEE Trans. Image Process., 17 (2008), pp. 1349–1364.
- [32] R. CURTO AND L. FIALKOW, *Solution of the Truncated Complex Moment Problem for Flat Data*, Memoirs Amer. Math. Soc., 568 (1996).
- [33] —, *The Quadratic Moment Problem for the Unit Circle and Unit Disk*, Integral Equations Operator Theory, 38 (2000), pp. 377–409.
- [34] —, *The Truncated Complex K-Moment Problem*, Trans. Amer. Math. Soc., 353 (2000), pp. 2825–2855.
- [35] J. D'ANGELO, *Inequalities from Complex Analysis*, Carus Math. Monogr., MAA, 2002.
- [36] —, *Hermitian Analogues of Hilbert's 17th Problem*, Adv. Math., 226 (2011), pp. 4607–4637.
- [37] J. D'ANGELO AND M. PUTINAR, *Polynomial Optimization on Odd-Dimensional Spheres*, in Emerging Applications of Algebraic Geometry, Springer New York, 2008.
- [38] —, *Hermitian Complexity of Real Polynomial Ideals*, Int. J. Math., 23 (2012).
- [39] E. DE KLERK AND M. LAURENT, *On the Lasserre Hierarchy of Semidefinite Programming Relaxations of Convex Polynomial Optimization Problems*, SIAM J. Optim., 21 (2011), pp. 824–832.

- [40] E. DE KLERK, T. TERLAKY, AND K. ROOS, *Self-Dual Embeddings*, in Handbook of Semidefinite Programming – Theory, Algorithms, and Applications, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Acad. Publ., Boston, 2000.
- [41] M. FARIVAR AND S. LOW, *A survey of the optimal power flow literature*, IEEE Trans. Power Syst., 28 (2013), pp. 2565–2572.
- [42] S. FLISCOUNAKIS, P. PANCIATICI, F. CAPITANESCU, AND L. WEHENKEL, *Contingency Ranking with Respect to Overloads in Very Large Power Systems Taking into Account Uncertainty, Preventive and Corrective Actions*, IEEE Trans. Power Syst., 28 (2013), pp. 4909–4917.
- [43] F. FOGEL, I. WALDSPURGER, AND A. D’ASPREMONT, *Phase Retrieval for Imaging Problems*, Preprint, available at: <http://arxiv.org/abs/1304.7735>, (2014).
- [44] B. GHADDAR, J. MARECEK, AND M. MEVISSSEN, *Optimal Power Flow as a Polynomial Optimization Problem*, IEEE Trans. Power Syst., (2015).
- [45] A. GOPALAKRISHNAN, A. RAGHUNATHAN, D. NIKOVSKI, AND L. BIEGLER, *Global optimization of optimal power flow using a branch & bound algorithm*, in 50th Annu. Allerton Conf. Commun., Control, Comput., 2012.
- [46] R. GRONE, C. JOHNSON, E. SÁ, AND H. WOLKOWICZ, *Positive Definite Completions of Partial Hermitian Matrices*, Linear Algebra Appl., 58 (1984), pp. 109–124.
- [47] D. HENRION AND J. B. LASSERRE, *Inner Approximations for Polynomial Matrix Inequalities and Robust Stability Regions*, IEEE Trans. Automat. Control, 57 (2012), pp. 1456–1467.
- [48] D. HENRION, J. B. LASSERRE, AND J. LÖFBERG, *GloptiPoly 3: Moments, optimization and semidefinite programming*, Optim. Method Softw., 24 (2009), pp. 761–779.
- [49] J. HILLING AND A. SUDBERY, *The Geometric Measure of Multipartite Entanglement and the Singular Values of a Hypermatrix*, J. Math. Phys., 51 (2010).
- [50] M. HUNEULT AND F. GALIANA, *A Survey of the Optimal Power Flow Literature*, IEEE Trans. Power Syst., 6 (1991), pp. 762–770.
- [51] R. JABR, *Radial Distribution Load Flow using Conic Programming*, IEEE Trans. Power Syst., 21 (2006), pp. 1458–1459.
- [52] R. JABR, *Exploiting Sparsity in SDP Relaxations of the OPF Problem*, IEEE Trans. Power Syst., 27 (2012), pp. 1138–1139.

- [53] B. JIANG, Z. LI, AND S. ZHANG, *Approximation Methods for Complex Polynomial Optimization*, Springer Comput. Optim. Appl., 59 (2014), pp. 219–248.
- [54] ———, *Characterizing Real-Valued Multivariate Complex Polynomials and Their Symmetric Tensor Representations*, Preprint, available at: <http://arxiv.org/abs/1501.01058>, (2015).
- [55] C. JOSZ, S. FLISCOUNAKIS, J. MAEGHT, AND P. PANCIATICI, *AC Power Flow Data in MATPOWER and QCQP format: iTesla, RTE Snapshots, and PEGASE*, <https://arxiv.org/abs/1603.01533>, (2016).
- [56] C. JOSZ AND D. HENRION, *Strong Duality in Lasserre’s Hierarchy for Polynomial Optimization*, Springer Optim. Lett., (2015).
- [57] C. JOSZ, J. MAEGHT, P. PANCIATICI, AND J. GILBERT, *Application of the Moment-SOS Approach to Global Optimization of the OPF Problem*, IEEE Trans. Power Syst., 30 (2015), pp. 463–470.
- [58] B. KOCUK, S. DEY, AND X. SUN, *Inexactness of SDP Relaxation and Valid Inequalities for Optimal Power Flow*, to appear in IEEE Trans. Power Syst., (2015).
- [59] J. KRIVINE, *Anneaux préordonnés*, J. d’Anal. Math., 12 (1964), pp. 307–326.
- [60] J. B. LASSERRE, *Convergent LMI Relaxations for Nonconvex Quadratic Programs*.
- [61] ———, *Optimisation Globale et Théorie des Moments*, C. R. Acad. Sci. Paris, Série I, 331 (2000), pp. 929–934.
- [62] ———, *Global Optimization with Polynomials and the Problem of Moments*, SIAM J. Optim., 11 (2001), pp. 796–817.
- [63] ———, *Convexity in semialgebraic geometry and polynomial optimization*, SIAM J. Optim., 19 (2008), pp. 1995–2014.
- [64] ———, *Moments, Positive Polynomials and Their Applications*, vol. 1, Imperial College Press, 2010.
- [65] M. LAURENT, *A Comparison of Sherali-Adams, Lovasz-Schrijver, and Lasserre Relaxations for 0-1 Programming*, Math. Oper. Res., 28 (2003), pp. 470–496.
- [66] J. LAVAEI AND S. LOW, *Zero Duality Gap in Optimal Power Flow Problem*, IEEE Trans. Power Syst., 27 (2012), pp. 92–107.
- [67] B. LESIEUTRE, D. MOLZAHN, A. BORDEN, AND C. DEMARCO, *Examining the Limits of the Application of Semidefinite Programming to Power Flow Problems*, in 49th Annu. Allerton Conf. Commun., Control, and Comput., Sept. 2011, pp. 1492–1499.

- [68] Z. LI, S. HE, AND S. ZHANG, *Approximation Methods for Polynomial Optimization: Models, Algorithms, and Applications*, Comput. Optim. Appl., Springer, New York, 2012.
- [69] J. LÖFBERG, *YALMIP: A Toolbox for Modeling and Optimization in MATLAB*, in IEEE Int. Symp. Comput. Aided Contr. Syst. Des., 2004, pp. 284–289.
- [70] S. LOW, *Convex Relaxation of Optimal Power Flow: Parts I & II*, IEEE Trans. Control Network Syst., 1 (2014), pp. 15–27.
- [71] Z. LUO, W.-K. MA, A.-C. SO, Y. YE, AND S. ZHANG, *Semidefinite Relaxation of Quadratic Optimization Problems*, IEEE Signal Process. Mag., 27 (2010), pp. 20–34.
- [72] R. MADANI, M. ASHRAPHIJUO, AND J. LAVAEI, *Promises of Conic Relaxation for Contingency-Constrained Optimal Power Flow Problem*, in 52nd Annu. Allerton Conf. Commun., Control, Comput., Sept. 2014, pp. 1064–1071.
- [73] R. MADANI, J. LAVAEI, AND R. BALDICK, *Convexification of Power Flow Problem over Arbitrary Networks*, to appear in IEEE 54th Ann. Conf. Decis. Contr. (CDC), (2015).
- [74] R. MADANI, S. SOJOUDI, AND J. LAVAEI, *Convex Relaxation for Optimal Power Flow Problem: Mesh Networks*, IEEE Trans. Power Syst., 30 (2015), pp. 199–211.
- [75] B. MARICIC, Z.-Q. LUO, AND T. DAVIDSON, *Blind Constant Modulus Equalization via Convex Optimization*, IEEE Trans. Signal Process., 51 (2003), pp. 805–818.
- [76] M. MARSHALL, *Positive Polynomials and Sums of Squares*, no. 146 in Mathematical Surveys and Monographs, American Mathematical Society, 2008.
- [77] S. MELACCI AND M. BELKIN, *Laplacian Support Vector Machines Trained in the Primal*, J. Mach. Learn. Res., 12 (2011), pp. 1149–1184.
- [78] D. MOLZAHN, S. BAGHSORKHI, AND I. HISKENS, *Semidefinite Relaxations of Equivalent Optimal Power Flow Problems: An Illustrative Example*, in IEEE Int. Symp. Circ. Syst. (ISCAS), May 24-27 2015.
- [79] D. MOLZAHN AND I. HISKENS, *Moment-Based Relaxation of the Optimal Power Flow Problem*, 18th Power Syst. Comput. Conf. (PSCC), (2014).
- [80] ———, *Mixed SDP/SOCP Moment Relaxations of the Optimal Power Flow Problem*, in IEEE Eindhoven PowerTech, 29 June–2 July 2015.
- [81] ———, *Sparsity-Exploiting Moment-Based Relaxations of the Optimal Power Flow Problem*, IEEE Trans. Power Syst., 30 (2015), pp. 3168–3180.

- [82] D. MOLZAHN, J. HOLZER, B. LESIEUTRE, AND C. DEMARCO, *Implementation of a Large-Scale Optimal Power Flow Solver Based on Semidefinite Programming*, IEEE Trans. Power Syst., 28 (2013), pp. 3987–3998.
- [83] D. MOLZAHN, C. JOSZ, I. HISKENS, AND P. PANCIATICI, *A Laplacian-Based Approach for Finding Near Globally Optimal Solutions to OPF Problems*, To appear in IEEE Trans. Power Syst.
- [84] —, *Solution of Optimal Power Flow Problems using Moment Relaxations Augmented with Objective Function Penalization*, IEEE 54th Ann. Conf. Decis. Contr. (CDC), (2015).
- [85] D. MOLZAHN, B. LESIEUTRE, AND C. DEMARCO, *Investigation of Non-Zero Duality Gap Solutions to a Semidefinite Relaxation of the Power Flow Equations*, in 47th Hawaii Int. Conf. Syst. Sci. (HICSS), 6-9 Jan. 2014.
- [86] A. MONTICELLI, S. DECHMANN, A. GARCIA, AND B. STOTT, *Real-Time External Equivalents for Static Security Analysis*, IEEE Trans. on Power Appar. and Syst., (1979).
- [87] J. NIE, *Optimality Conditions and Finite Convergence of Lasserre’s Hierarchy*, Math. Program., 146 (2014), pp. 97–121.
- [88] K. PANDYA AND S. JOSHI, *A Survey of Optimal Power Flow Methods*, J. Theor. Appl. Inf. Tech., 4 (2008), pp. 450—458.
- [89] P. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, Cal. Inst. of Tech., May 2000.
- [90] —, *Semidefinite Programming Relaxations for Semialgebraic Problems*, Math. Program., 96 (2003), pp. 293–320.
- [91] A. PRESTEL AND C. DELZELL, *Positive Polynomials*, Springer Monogr. Math., 2001.
- [92] M. PUTINAR, *Positive Polynomials on Compact Semi-Algebraic Sets*, Indiana Univ. Math. J., 42 (1993), pp. 969–984.
- [93] —, *A Two-Dimensional Moment Problem*, J. Funct. Anal., 80 (1998), pp. 1–8.
- [94] —, *On Hermitian Polynomial Optimization*, Arch. Math., 87 (2006), pp. 41–51.
- [95] M. PUTINAR AND C. SCHEIDERER, *Quillen Property of Real Algebraic Varieties*, To appear in Muenster J. Math.
- [96] —, *Hermitian Algebra on the Ellipse*, Illinois J. Math., 56 (2012), pp. 213–220.

- [97] M. PUTINAR AND K. SCHMÜDGEN, *Multivariate Determinateness*, Indiana U. Math. J., 57 (2008), pp. 2931–2968.
- [98] D. QUILLEN, *On the Representation of Hermitian Forms as Sums of Squares*, Invent. Math., 5 (1968), pp. 237–242.
- [99] R. ROCKAFELLAR, *Conjugate Duality and Optimization*, no. 16 in Reg. Conf. Ser. Appl. Math., SIAM, Philadelphia, PA, USA, 1974.
- [100] W. RUDIN, *Real and Complex Analysis*, Math. Ser., Third Edition, McGraw Hill Int. Ed., 1987.
- [101] C. SCHEIDERER, *Positivity and Sums of Squares: A Guide to Recent Results*, vol. 149 of IMA Vol. Math. Appl., Springer, New York, 2009.
- [102] K. SCHMÜDGEN, *The K -Moment Problem for Semi-Algebraic Sets*, Math. Ann., 289 (1991), pp. 203–206.
- [103] ———, *Around Hilbert’s 17th Problem*, Doc. Math., Extra Vol. ISMP, (2012), pp. 433–438.
- [104] M. SCHWEIGHOFER, *Optimization of Polynomials on Compact Semialgebraic Sets*, SIAM J. Optim., 15 (2005), pp. 805–825.
- [105] A. SHAPIRO AND K. SCHEINBERG, *Duality, Optimality Conditions and Perturbation Analysis*, in Handbook of Semidefinite Programming – Theory, Algorithms, and Applications, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Acad. Publ., Boston, 2000.
- [106] N. SHOR, *Quadratic Optimization Problems*, Sov. J. Comput. Syst. Sci., 25 (1987), pp. 1–11.
- [107] SIEMENS PTI, *Volume II: Program Application Guide*, Power System Simulation for Engineering (PSS/E), 31.0 (2007).
- [108] A. SINGER, *Angular Synchronization by Eigenvectors and Semidefinite Programming*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 20–36.
- [109] A. J. SMOLA AND R. KONDOR, *Kernels and Regularization on Graphs*, in Learning Theory and Kernel Machines, Springer, 2003, pp. 144–158.
- [110] S. SOJODI AND J. LAVAEI, *Network Topologies Guaranteeing Zero Duality Gap for Optimal Power Flow Problem*, Submitted to IEEE Trans. Power Syst., (2012).
- [111] ———, *Physics of power networks makes hard optimization problems easy to solve*, in IEEE Pow. & Ener. Soc. (PES) Gen. Meeting, 2012.

- [112] L. SORBER, M. BAREL, AND L. DE LATHAUWER, *Unconstrained Optimization of Real Functions in Complex Variables*, SIAM J. Optim., 22 (2012), pp. 879–898.
- [113] G. STENGLE, *A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry*, Math. Ann., 207 (1974), pp. 87–97.
- [114] J. STOCHEL, *Solving the Truncated Moment Problem Solves the Full Moment Problem*, Glasg. Math. J., 43 (2001), pp. 335–341.
- [115] J. STURM, *Using SeDuMi 1.02, A Matlab Toolbox for Optimization over Symmetric Cones*, Optim. Method Softw., 11 (1999), pp. 625–653.
- [116] J. STURM AND S. ZHANG, *On Cones of Nonnegative Quadratic Functions*, Math. Oper. Res., 28 (2003), pp. 246–267.
- [117] R. TARJAN AND M. YANNAKAKIS, *Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs*, SIAM J. Comput., 13 (1984), p. 566.
- [118] J. TAYLOR, *Convex Optimization of Power Systems*, Cambridge University Press, 2015.
- [119] J. TAYLOR AND F. HOVER, *Conic AC Transmission System Planning*, IEEE Trans. Power Syst., 28 (2013), pp. 952–959.
- [120] J. A. TAYLOR AND F. S. HOVER, *Laplacians for Flow Networks*, SIAM J. Discrete Math., 25 (2011), pp. 1349–1364.
- [121] O. TOKER AND H. OZBAY, *On the Complexity of Purely Complex Mu Computation and Related Problems in Multidimensional Systems*, IEEE Trans. Automat. Control, 43 (1998), pp. 409–414.
- [122] M. TRNOVSKÁ, *Strong Duality Conditions in Semidefinite Programming*, J. Electr. Eng., 56 (2005), pp. 1–5.
- [123] UNIVERSITY OF EDINBURGH POWER SYSTEMS TEST CASE ARCHIVE, *GB Network*.
- [124] F. VASCILESCU, *Spectral Measures and Moment Problems*, Spectral Theory and Its Applications, (2003), pp. 173–215.
- [125] H. WAKI, S. KIM, M. KOJIMA, AND M. MURAMATSU, *Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity*, SIAM J. Optimiz., 17 (2006), pp. 218–242.
- [126] H. WAKI, S. KIM, M. KOJIMA, AND M. MURAMATSU, *Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity*, SIAM J. Optim., 17 (2006), pp. 218–242.

- [127] —, *Algorithm 883: SparsePOP – A Sparse Semidefinite Programming Relaxation of Polynomial Optimization Problems*, 35 (2008). [doi].
- [128] K. Q. WEINBERGER, F. SHA, Q. ZHU, AND L. K. SAUL, *Graph Laplacian Regularization for Large-Scale Semidefinite Programming*, in *Adv. Neural Informat. Process. Syst. (NIPS)*, 2006, pp. 1489–1496.
- [129] W. WIRTINGER, *Zur Formalen Theorie der Funktionen von Mehr Komplexen Veränderlichen*, *Math. Ann.*, 97 (1927), pp. 357–375.
- [130] B. ZHANG AND D. TSE, *Geometry of Feasible Injection Region of Power Networks*, in *50th Annu. Allerton Conf. Commun., Control, Comput.*, 2011.
- [131] —, *Geometry of Feasible Injection Region of Power Networks*, *IEEE Trans. Power Syst.*, 28 (2013), pp. 788–797.
- [132] R. ZIMMERMAN, C. MURILLO-SÁNCHEZ, AND R. THOMAS, *MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education*, *IEEE Trans. Power Syst.*, (2011), pp. 1–8.