NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES

DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

PROGRAM OF POSTGRADUATE STUDIES

PhD THESIS

Limitations of Linear Programming as a model of approximate computation

Yannis D. Moysoglou

ATHENS

July 2015

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Όρια του Γραμμικού Προγραμματισμού ως Μοντέλου Προσεγγιστικών Υπολογισμών

Ιωάννης Δ. Μωυσόγλου

ΑΘΗΝΑ

Ιούλιος 2015

PhD THESIS

Limitations of Linear Programming as a model of approximate computation

Yannis D. Moysoglou

SUPERVISOR: Stavros Kolliopoulos, Associate Professor UoA

THREE-MEMBER ADVISORY COMMITTEE:
Stavros Kolliopoulos, Associate Professor UoA
Vassilis Zissimopoulos, Professor UoA
George Karakostas, Associate Professor McMaster University

SEVEN-MEMBER EXAMINATION COMMITTEE

Stavros Kolliopoulos
Associate Professor UoA

Vassilis Zissimopoulos
Professor UoA

George Karakostas
Assoc. Professor McMaster U.

Ioannis Emiris
Professor UoA

Dimitris Fotakis
Assistant Professor NTUA

Leonidas Pitsoulis
Associate Professor AUTH

Dimitrios Thilikos
Professor UoA

Examination Date: 27/7/2015

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Όρια του Γραμμικού Προγραμματισμού ως Μοντέλου Προσεγγιστικών Υπολογισμών

Ιωάννης Δ. Μωυσόγλου

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ : Σταύρος Κολλιόπουλος, Αναπληρωτής Καθηγητής Ε.Κ.Π.Α.

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ :

Σταύρος Κολλιόπουλος, Αναπληρωτής Καθηγητής Ε.Κ.Π.Α.
Βασίλειος Ζησιμόπουλος, Καθηγητής Ε.Κ.Π.Α.
Γεώργιος Καρακώστας, Αναπληρωτής Καθηγητής Πανεπιστήμιο McMaster

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Σταύρος Κολλιόπουλος
Αναπληρωτής Καθηγητής Ε.Κ.Π.Α.

Βασίλειος Ζησιμόπουλος
Καθηγητής Ε.Κ.Π.Α.

Γεώργιος Καρακώστας
Αναπλ. Καθηγητής Παν. McMaster

Ιωάννης Εμίρης
Καθηγητής Ε.Κ.Π.Α.

Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Λεωνίδας Πιτσούλης
Αναπληρωτής Καθηγητής Α.Π.Θ.

Δημήτριος Θηλυκός
Καθηγητής Ε.Κ.Π.Α.

Ημερομηνία εξέτασης : 27/7/2015

ABSTRACT

Linear programming has proved to be one of the most powerful and widely used tools in algorithm design and especially in the design of approximation algorithms. It has proved its expressive power by modeling diverse types of problems in planning, routing, scheduling, assignment, and design. The popularity of linear programming raised the question whether for every algorithm there is a linear formulation or relaxation that captures its merits.

In the last decade, the virtual "completeness" of linear programming-based algorithms was questioned. Although this line of thought goes at least as back as the late 80s with the famous work of Yannakakis [66], it wasn't until the early 00's that the community started to systematically study the limitations of linear programming methods. In particular, the first line of results that incarnated that questioning was the proof of the limitations of lift-and-project methods for solving various problems. Those methods, when applied to a relaxation, define hierarchies of increasingly strengthened relaxations. Their power is such that proving a hardness result for some problem for a lift and project hierarchy can be taken as evidence that linear programming in general cannot efficiently solve that problem.

Exploring the power of linear programming for combinatorial optimization problems has been recently receiving renewed attention through the model of Extended Formulations. In this model we are concerned whether there are compact formulations or approximate relaxations for combinatorial optimization problems, allowing the use of any set or type of variables. The model of Extended Formulations is more general than that of lift-and-project hierarchies and is arguably the most general model that truly captures pure linear programming methods.

In this thesis we take the direction of exploring the limitation of linear programming. Most of the thesis's results are concerned with linear programming approximability of the capacitated versions of the metric facility location problem such as the capacitated facility location (CFL) and lower bounded facility location (LBFL), problems which, while they can be approximated within a constant factor using local search, they are not known to admit efficient relaxation based approximations. The failure for many years of the combinatorial optimization community to devise efficient relaxations for such problems has created a certain sensation. We give impossibility results in the hierarchy and in the extended formulations models and we also consider an other, independent family of relaxations. We note that our ideas and methods are of general value and independent interest. This is especially true for the extended formulations where we devise a general method for lower bounding the size of linear programs.

Regarding our hierarchy related results, we show that the relaxations obtained from the natural LP at $\Omega(n)$ levels of the semidefinite Lovász-Schrijver hierarchy for mixed programs, and at $\Omega(n)$ levels of the Sherali-Adams hierarchy, have an integrality gap of $\Omega(n)$, where n is the number of facilities of the instance. Both hierarchies yield an exact for-

mulation for CFL in at most n levels for the families of instances we consider and thus our bounds are asymptotically tight. For the families of instances we consider, both hierarchies yield at the nth level an exact formulation for CFL. Thus our bounds are asymptotically tight. Building on our methodology for the Sherali-Adams result, we prove that the standard CFL relaxation enriched with the submodular inequalities of [1], a generalization of the flow-cover valid inequalities, has also an $\Omega(n)$ gap and thus not bounded by any constant. This disproves a long-standing conjecture of [46].

We propose a framework for proving lower bounds on the size of extended formulations. We do so by introducing specific types of extended relaxations that we call *product and distributional relaxations*. Then we show that for every approximate extended formulation of a polytope P, there is a product or distributional relaxation that has the same size and is at least as strong. We provide a methodology for proving lower bounds on the size of approximate product and distributional relaxations by lower bounding the chromatic number of an underlying hypergraph, whose vertices correspond to gap-inducing vectors. As an application of our method we show for CFL an exponential lower bound on the size of a restricted type of extended formulations which we call mixed product relaxations.

We finally introduce the family of proper relaxations which generalizes to its logical extreme the classic star relaxation and captures general configuration-style LPs. We characterize the behavior of proper relaxations for CFL and LBFL through a sharp threshold phenomenon.

SUBJECT AREA: Computational Complexity, Approximation Algorithms

KEYWORDS: linear programming hierarchies, extended formulations, facility location, integrality gap, configuration linear program

ΠΕΡΙΛΗΨΗ

Ο Γραμμικός Προγραμματισμός έχει αποδειχθεί ως ένα από τα ισχυρότερα και ευρέως χρησιμοποιούμενα εργαλεία όσον αφορά τη σχεδίαση αλγορίθμων και ιδιαίτερα των προσεγγιστικών αλγορίθμων. Η εκφραστική του δύναμη είναι εμφανής στη μοντελοποίηση προβλημάτων διαφόρων τύπων όπως τη δρομολόγηση, τον χρονοπρογραμματισμό εργασιών και την ανάθεση πόρων. Η δημοφιλία του Γραμμικού Προγραμματισμού έγειρε το ερώτημα αν για κάθε αλγόριθμο υπάρχει ένα ισοδύναμο γραμμικό πρόγραμμα.

Την περασμένη δεκαετία η φαινομενική ¨πληρότητα¨ του γραμμικού προγραμματισμού αμφισβητήθηκε. Αν και η αμφισβήτηση αυτή μπορεί να ανιχνευτεί τουλάχιστον από τα τέλη της δεκαετίας του 1980 με την περίφημη εργασία του Γιαννακάκη [66], μόνο στις αρχές της δεκαετίας του 2000 η κοινότητα άρχισε συστηματικά να μελετά τους περιορισμούς μεθόδων γραμμικού προγραμματισμού. Πιο συγκεκριμένα, η πρώτη γραμμή αποτελεσμάτων που πραγμάτωνε αυτή την αμφισβήτηση είχε να κάνει με την απόδειξη των περιορισμών των lift-and-project μεθόδων για την επίλυση διαφόρων προβλημάτων. Αυτές οι μέθοδοι, όταν εφαρμοστούν σε μία γραμμική χαλάρωση, ορίζουν μία ιεραρχία γραμμικών χαλαρώσεων αυξανομένης ισχύος. Η δύναμή τους είναι τέτοια ώστε ένα αρνητικό αποτέλεσμα για κάποιο πρόβλημα για μία lift-and-project ιεραρχία μπορεί να ερμηνευθεί ως ένδειξη της αδυναμίας του γραμμικού προγραμματισμού να λύσει το πρόβλημα.

Η εξερεύνηση της δύναμης του γραμμικού προγραμματισμού έλαβε το ανανεωμένο ενδιαφέρον της κοινότητας μέσω του μοντέλου των Extended Formulations. Ο στόχος του μοντέλου αυτού είναι να απαντηθεί αν υπάρχει ή όχι μικρού μεγέθους γραμμική χαλάρωση για ένα πρόβλημα επιτρέποντας τη χρήση οποιουδήποτε συνόλου μεταβλητών. Το μοντέλο των Extended Formulations είναι πιο γενικό από αυτό των lift-and-project ιεραρχιών και μπορεί κάλλιστα να υποστηριχθεί ότι εύστοχα περιγράφει την δύναμη του γραμμικού προγραμματισμού.

Σε αυτή τη διατριβή επιλέχθηκε ο δρόμος της μελέτης των περιορισμών του γραμμικού προγραμματισμού. Τα περισσότερα από τα αποτελέσματα που παρουσιάζονται έχουν να κάνουν με την προσεγγισιμότητα μέσω γραμμικού προγραμματισμού facility location προβλημάτων με χωρητικότητες όπως το capacitated facility location (Cfl) και το lower bounded facility location (Lbfl), προβλήματα για τα οποία, αν και μπορούν να προσεγγιστούν με σταθερό λόγο προσέγγισης, δεν είναι γνωστό αν υπάρχει αποτελεσματικός προσεγγιστικός αλγόριθμος βασισμένος σε γραμμική χαλάρωση. Η επί πολλά έτη αποτυχία της κοινότητας να βρει μία τέτοια χαλάρωση έχει προκαλέσει αίσθηση. Δίνουμε αρνητικά αποτελέσματα στα μοντέλα των ιεραρχιών και των extended formulations και επίσης μελετούμε μία ανεξάρτητη από τα προηγούμενα οικογένεια χαλαρώσεων. Επισημαίνουμε ότι οι ιδέες και οι μέθοδοι είναι γενικότερης αξίας και όχι αναγκαστικά συνυφασμένες με κάποιο συγκεκριμένο πρόβλημα. Αυτό έχει ιδιαίτερη ισχύ για τα extended formulations όπου σχεδιάζουμε μία καινούρια μέθοδο για τη φραγή από κάτω του μεγέθους των γραμμικών προγραμμάτων.

Σχετικά με τα αποτελέσματά μας σε ιεραρχίες, δείχνουμε ότι οι χαλαρώσεις που προκύπτουν από το κλασσικό γραμμικό πρόγραμμα για το Cfl σε $\Omega(n)$ επίπεδα της semidefinite Lovász-Schrijver ιεραρχίας για mixed προγράμματα και σε $\Omega(n)$ επίπεδα της Sherali-Adams

ιεραρχίας, έχουν integrality gap $\Omega(n)$. Και οι δύο ιεραρχίες δίνουν ένα ακριβές formulation το πολύ σε n επίπεδα και επομένως τα αποτελέσματά μας είναι ασυμπτωτικά σφικτά. Χρησιμοποιώντας τις ιδέες και τις μεθόδους που αναπτύχθηκαν για το αποτέλεσμα για την Sherali-Adams ιεραρχία, αποδεικνύουμε ότι το κλασσικό γραμμικό πρόγραμμα για το CFL έχει μη φραγμένο integrality gap ακόμα και μετά την εισαγωγή των submodular inequalities του [1], μία γενίκευση των flow-cover valid inequalities. Το τελευταίο αποτέλεσμα απαντά αρνητικά σε μία επί χρόνια ανοικτή εικασία [46].

Προτείνουμε μία μεθοδολογία για το φράξιμο του μεγέθους των extended formulations. Αυτό το πετυχαίνουμε κάνοντας χρήση συγκεκριμένων τύπων extended formulations που ορίζουμε, τα product και τα distributional relaxations. Έπειτα αποδεικνύουμε ότι για οποιοδήποτε προσεγγιστικό extended formulation ενός πολυτόπου P υπάρχει ένα product ή distributional formulation που έχει το ίδιο μέγεθος και είναι τουλάχιστον το ίδιο ισχυρό. Δίνουμε μία μέθοδο απόδειξης κάτω φραγμάτων για προσεγγιστικά product και distributional relaxations με το να φράσσεται ο χρωματικός αριθμός ενός σχετικού υπεργραφήματος, του οποίου οι κορυφές αντιστοιχούν σε διανύσματα που προκαλούν gap. Ως εφαρμογή της μεθοδολογίας μας δείχνουμε εκθετικό φράγμα στο μέγεθος για το CFL ενός ειδικού τύπου extended formulations τα οποία ονομάζουμε mixed product relaxations.

Τέλος εισάγουμε και μελετούμε την οικογένεια των proper relaxations τα οποία γενικεύουν το star relaxation και μοντελοποιεί γενικά configuration-style LPs. Χαρακτηρίζουμε τη συμπεριφορά των proper relaxations για το CFL και LBFL μέσω ενός φαινομένου κατωφλίου.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Πολυπλοκότητα, Προσεγγιστικοί Αλγόριθμοι

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: ιεραρχίες γραμμικών προγραμμάτων, επεκταμένες διατυπώσεις, χωροθέτηση εγκαταστάσεων, χάσμα ακεραιότητας, γραμμικό πρόγραμμα διαμορφώσεων

ACKNOWLEDGEMENTS

First of all, I would like to thank my PhD advisor Stavros Kolliopoulos for his support, for his advice, for the time he spent to read and comment on my work including this thesis and, more importantly, for the freedom he granted me to pursue the research directions of my interest in a trouble-free environment, which in turn allowed me to produce the results presented in this thesis.

I would like to thank the faculty of the department of Informatics and Telecommunications and especially the professors Dimitris Achlioptas, Ioannis Emiris, Vassilis Zissimopoulos, Elias Koutsoupias, Nikolaos Mysirlis and Panos Rondogiannis for their important lessons that helped me build a solid backround in theoretical computer science and helped me to shape my interests. I would also like to thank the professors Dimitris Thilikos and Leonidas Pitsoulis for their valuable support.

I would also like to acknowledge the support of a University of Athens ELKE graduate fellowship during the first 3 years of my PhD work.

Special thanks are given to my dear friend Giorgos Kasetas for standing beside me through difficult times.

I give my warmest thanks to Pavlos Maragoudakis for inspiring in me the love of problem solving and for introducing me to a whole new world.

And, of course, I would like to thank my family and especially my parents for their love, for the sacrifices they made to raise me, and for nourishing everything that is good in me.

ΣΥΝΟΠΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ

ΤΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Ο Γραμμικός Προγραμματισμός έχει αποδειχθεί ως ένα από τα ισχυρότερα και ευρέως χρησιμοποιούμενα εργαλεία όσον αφορά τη σχεδίαση αλγορίθμων και ιδιαίτερα των προσεγγιστικών αλγορίθμων. Η εκφραστική του δύναμη είναι εμφανής στη μοντελοποίηση προβλημάτων διαφόρων τύπων όπως τη δρομολόγηση στα δίκτυα, το χρονοπρογραμματισμό εργασιών και την ανάθεση πόρων.

Την περασμένη δεκαετία η φαινομενική ¨πληρότητα¨ του γραμμικού προγραμματισμού αμφισβητήθηκε. Πιο συγκεκριμένα, η πρώτη γραμμή αποτελεσμάτων που πραγμάτωνε αυτή την αμφισβήτηση είχε να κάνει με την απόδειξη των περιορισμών των lift-and-project μεθόδων για την επίλυση διαφόρων προβλημάτων. Αυτές οι μέθοδοι, όταν εφαρμοστούν σε μία γραμμική χαλάρωση, ορίζουν μία ιεραρχία γραμμικών χαλαρώσεων αυξανομένης ισχύος. Η δύναμή τους είναι τέτοια ώστε ένα αρνητικό αποτέλεσμα για κάποιο πρόβλημα για μία lift-and-project ιεραρχία μπορεί να ερμηνευθεί ως ένδειξη της αδυναμίας του γραμμικού προγραμματισμού να λύσει το πρόβλημα αποδοτικά.

Η εξερεύνηση της δύναμης του γραμμικού προγραμματισμού έλαβε το ανανεωμένο ενδιαφέρον της κοινότητας μέσω του μοντέλου των Extended Formulations. Ο στόχος του μοντέλου αυτού είναι να απαντηθεί αν υπάρχει ή όχι μικρού μεγέθους γραμμική χαλάρωση για ένα πρόβλημα επιτρέποντας τη χρήση οποιουδήποτε συνόλου μεταβλητών. Το μοντέλο των Extended Formulations είναι πιο γενικό από αυτό των lift-and-project ιεραρχιών και μπορεί κάλλιστα να υποστηριχθεί ότι περιγράφει εύστοχα τη δύναμη του γραμμικού προγραμματισμού.

Σε αυτή τη διατριβή γίνεται μελέτη των περιορισμών του γραμμικού προγραμματισμού στα παραπάνω μοντέλα. Τα περισσότερα από τα αποτελέσματα που παρουσιάζονται αφορούν την προσεγγισιμότητα μέσω γραμμικού προγραμματισμού facility location προβλημάτων με χωρητικότητες όπως το capacitated facility location και το lower bounded facility location. Δίνουμε αρνητικά αποτελέσματα στα μοντέλα των γραμμικών ιεραρχιών και των extended formulations και επίσης μελετούμε μία ανεξάρτητη από τα προηγούμενα οικογένεια χαλαρώσεων. Οι ιδέες και οι πρωτότυπες μέθοδοι είναι γενικότερης αξίας, ιδιαίτερα για το μοντέλο των extended formulations όπου παρουσιάζουμε μία καινούργια τεχνική για τη φραγή από κάτω του μεγέθους των γραμμικών προγραμμάτων.

Στο υπόλοιπο κείμενο της παρούσας σύνοψης θα γίνει μία συνοπτική παρουσίαση των αποτελεσμάτων τα οποία προέκυψαν από την έρευνα που εκπονήθηκε στα πλαίσια της διατριβής, καθώς και σύνδεση αυτών των αποτελεσμάτων με σημαντηκά προβλήματα της περιοχής.

Το Facility location πρόβλημα είναι ένα από τα πιο μελετημένα προβλήματα στην συνδυαστική βελτιστοποίηση. Στην ανευ χωρητικότητες εκδοχή αυτού (*uncapacitated facility location*) ή *(*UFL*)* μας δίνετε ένα συνολο F από facilities και ένα σύνολο C από clients.Μπορούμε να ανοίξουμε το facility i πληρώνοντας το κόστος ανοίγματός του (opening cost) f_i και μπορούμε να αναθέσουμε τον client j στο facility i πληρώνοντας το κόστος σύνδεσης (ςοννεςτιον ςοστ) c_{ij}.Το ζητούμενο είναι να ανπίξουμε ένα υποσύνολο $F' \subseteq F$ των facilities και να αναθέσουμε τον κάθε client σε κάποιο ανοιχτό facility. Ο στόχος είναι να ελαχιστοποιηθεί το συνολικό opening και connection cost. Η προσεγγισιμότητα του UFL απαντήθηκε με έναν $O(\log|C|)$-προσεγγιστικό αλγόριθμο από το [33], το οποίο μέσα από μία αναγωγή στο Set Cover είναι ασυμπτωτικά βέλτιστο, εκτός αν P = NP [55].

Στο *metric* UFL τα κόστη ανάθεσης ικανοποιούν την ακόλουθη παραλλαγή της τριγωνικής ανισότητας: $c_{ij} \leq c_{ij'} + c_{i'j'} + c_{i'j}$ για κάθε $i, i' \in F$ και $j, j' \in C$. Αυτή η φυσική ειδική περίπτωση του UFL είναι προσεγγίσιμη με σταθερό λόγο, και πολλές βελτιωμένες προσεγγίσεις έχουν δημοσιευθεί τα τελευταία χρόνια. Σε αυτές, LP-based μέθοδοι, όπως filtering, randomized rounding και primal-dual έχουν πρωταγωνιστίσει (βλέπε π.χ. [65]). Ο καλύτερος μέχρι στιγμής προσεγγιστικός αλγόριθμος για το metric UFL έχει προσέγγιση 1.488 [47], ενώ το καλύτερο γνωστό κάτω φράγμα είναι 1.463, εκτός αν P = NP ([64]).

Το CFL είναι μία γενίκευση του metric UFL όπου το κάθε facility i έχει μία χωρητικότητα u_i που καθορίζει το μέγιστο πλήθος από clients που μπορούν να ανατεθούν στο i. Στο *uniform* CFL όλα τα facilities έχουν την ίδια χωρητικότητα U. Η εύρεση ενός προσεγγιστικού αλγορίθμου για το CFL ήταν έως πρόσφατα ένα απίμονο ανοικτό πρόβλημα. Έως την πρόσφατη εργασία των [7], οι μόνοι $O(1)$-προσεγγιστικοί αλγόριθμοι ήταν βασισμένοι σε local search, με την καλύτερη προσέγγιση μέχρι στιγμής να έχει λόγο 5 [11] για την μη uniform περίπτωση και προσέγγιση με λόγο 3 [4] για την uniform περίπτωση. Οι Williamson και Shmoys [65], κατέταξαν την εύρεση προσεγγιστικού αλγορίθμου βασισμένου σε χαλάρωση για το CFL ως ένα από τα 10 κορυφαία προβλήματα των προσεγγιστικών αλγορίθμων. Πρόσφατα οι An et al. [7] έδωσαν έναν LP-based 288-προσεγγιστικό αλγόριθμο. Όμως το LP του [7] έχει εκθετικό μέγεθος και δεν είναι γνωστό αν μπορεί να επιλυθεί σε πολυωνυμικό χρόνο. Επομένως το ερώτημα για το εάν υπάρχει μία πολυωνιμική γραμμική χαλάρωση με φραγμένο integrality gap για το CFL παραμένει ανοικτό. Τα αποτελέσματά μας σχετικά με το LP-(in)approximability του CFL μπορούν να ερμηνευθούν ως πολύ ισχυρές ενδείξεις ότι η απάντηση στο προηγούμενο ερώτημα είναι αρνητική.

Έχουν γίνει αρκετές προσπάθειες να κατανοηθούν τα 0-1 πολύτοπα που προκύπτουν από την εφαρμαγή επαναληπτικών lift-and-project διεργασιών. Αυτές οι διεργασίες παράγουν ιεραρχίες από προοδευτικά ισχυρότερες γραμμικές χαλαρώσεις, όπου valid ινεχυαλιτιες προστίθενται σε κάθε επίπεδο. Μετά από το πολύ d επίπεδα, όπου d είναι η διάσταση της αρχικής χαλάρωσης, παίρνουμε μία περιγραφή του ακέραιου πολυτόπου. Παραδείγματα τέτοιων μεθόδων είναι η μέθοδος των Balas et al. [10], των Lovász και Schrijver [49] (για γραμμικά και semidefinite προγράμματα), των Sherali και Adams [59], του Lasserre [43] (για semidefinite προγράμματα). Βλέπε [44] για μία συγκριτική μελέτη.

Στην εργασία των Arora et al. [9], μελετήθηκε το integrality gap χαλαρώσεων για το Vertex Cover, συμπεριλαμβανομένων και χαλαρώσεων που προκύπτουν από την Lovász-Schrijver (LS) ιεραρχία. Αυτή η εργασία εισήγαγε την χρήση των ιεραρχιών ως μοντέλο για απόδειξη LP-βασισμένης δυσκολίας προσέγγισης. Το να αποδειχτεί ότι το integrality gap παραμένει υψηλό για πολλά επίπεδα μιας ιεραρχίας, αυτό αποκλείει την χρήση των χαλαρώσεων αυτής για τον σχεδιασμό βελτιωμένων προσεγγιστικών αλγορίθμων. Από την άλλη, το να δείξει κάποιος ότι το integrality gap παραμένει υψηλό (βλέπε επίσης [58]) μπορεί να παρθεί ως ένδειξη αδυναμίας του γραμμικού προγραματισμού. Η προηγούμενη πεποίθηση είναι πλέον θεώρημα για τα maximum constraint satisfaction problems: όσον αφορά την προσέγγιση, LPs μεγέθους d^k, ακριβώς τόσο ισχυρά όσο $O(k)$-επιπέδου Sherali-Adams (SA) χαλαρώσεις [21]. Από την αλγοριθμική σκοπιά, χαλαρώσεις των πρώτων $O(1)$ επιπέδων των ιεραρχιών μπορούν να βελτιστοποιηθούν σε πολυωνιμικό χρόνο εφόσον υπάρχει ασθενής διαχωρισμός για την αρχική χαλάρωση [49, 59].

Σημειώνουμε ότι υπάρχουν παραδείγματα πολυτόπων, όπως το matching πολύτοπο, του οποίου το convex hull προκύπτει σε υψηλά επίπεδα της LS και της SA ιεραρχίας [53], ενώ υπάρχουν εκθετικές περιγραφές του πολυτόπου οι οποίες να βελτιστοποιηθούν σε πολυωνιμικό χρόνο. Ο αλγόριθμος διαχωρισμού για το matching γραμμικό πρόγραμμα όμως περιλαμβάνει βήματα που δεν ήταν γνωστό αν αυτά μπορούν να μεταφραστούν σε ένα πολυωνιμικού μεγέθους γραμμικό πρόγραμμα. Μετά το σημαντικό αποτέλεσμα του Rothvoß [56] γνωρίζουμε ότι αυτός ο αλγόριθμος διαχωρισμού δεν μπορεί να εξομοιωθεί από ένα πολυωνιμικού μεγέθους γραμμικό πρόγραμμα. Έτσι, τουλάχιστον για το matching, η αδυναμία των ιεραρχιών για το πρόβλημα είναι εξαιτίας της εγγενής δυσκολίας του προβλήματος για τον γραμμικό προγραμματισμό, επιβεβαιώνοντας την πεποίθηση στην οιπόια αναφερθήκαμε πιο πάνω (βλέπε π.χ. [58]). Για την ακρίβεια η τελευταία είναι πλέον θεώρημα για τα constraint satisfaction problems χάρις τη εργασία [21].

Δίνουμε αρνητικά αποτελέσματα σε υποσχόμενα μοντέλα γραμμικού προγραμματισμού για το CFL και παράλληλα απαντάμε σε ανοιχτά προβλήματα από την βιβλιογραφία.

Τα πρώτα μας αποτελέσματα σχετικά με αυτό το κομμάτι είναι ότι υπάρχουν στιγμιότυπα με $\Theta(n)$ facilities και $\Theta(n^4)$ clients για τα οποία οι χαλαρώσεις οι οποίες προκύπτουν από $\Omega(n)$ επίπεδα, όταν χρησιμοποιούμε την LS διεργασία στην κλασική γραμμική χαλάρωση για το CFL έχει integrality gap της τάξης του $\Omega(n)$. Η κλασική χαλάρωση περιλαμβάνει μεταβλητές y_i, για κάθε $i \in F$, και μεταβλητές ανάθεσης x_{ij}, για κάθε $i \in F$, και για κάθε client $j \in C$.

Ο ορισμός της LS ιεραρχίας είναι επαγωγικός (βλ. κεφάλαιο 3.3.3). Για να πάρουμε τη χαλάρωση του επιπέδου $t+1$, εφαρμόζουμε την LS διεργασία στη χαλάρωση επιπέδου t. Είναι γνωστό ότι η LS διεργασία επεκτείνεται στα μικτά 0-1 προγράμματα [49, 10] όπως το CFL με γενικά client demands. Στα στιγμιότυπα που θεωρούμε οι clients έχουν μοναδιαίο demand: είναι γνωστό ότι σε αυτή την περίπτωση το μικτό και το ακέραιο πολύτοπο ταυτίζονται. Στην απόδειξη του θεωρήματος 4.1.1, μεταχειριζόμαστε τις μεταβλητές ως δυαδικές. Στην απόδειξή μας χρησιμοποιούμε μία επαναδιατύπωση των συνθηκών του [49] το οποίο είναι ξεχωριστού ενδιαφέροντος.

Η LS_+ διεργασία είναι μία ισχυρότερη εκδοχή της LS όπου απαιτείται επιπλέον συνθήκη ημιθετικότητας. Η *mixed* LS_+ διεργασία για την εκδοχή του LS_+ όπου κάποιος απαιτεί ο πίνακας να είναι ημιθετικός στις 0-1 μεταβλητές (βλ, [10], [26]). Το θεώρημα 4.1.4, το οποίο έπεται απευθείας από το αποτέλεσμα για την LS ιεραρχία, δείχνει ότι το $\Omega(n)$ gap ισχύει για $\Omega(n)$ γύρους της μικτής LS_+ επίσης.

Έπειτα, δείχνουμε ότι τα LPs που παίρνουμε από την κλασική γραμμική χαλάρωση για το CFL σε $\Omega(n)$ επίπεδα της ισχυρότερης SA ιεραρχία έχουν gap $\Omega(n)$ για την ίδια οικογένεια στιγμιοτύπων που χρησιμοποιήσαμε για το LS αποτέλεσμα, με $|F| = \Theta(n)$ και $|C| = \Theta(n^4)$, και έτσι δίνουμε τη δεύτερη συνεισφορά μας στις γραμμικές ιεραρχίες. Αυτό το αποτέλεσμα απαντάει στα ανοιχτά ερωτήματα [48] και [6] που αναφέρθηκαν προηγουμένως, όσον αφορά το κλασικό LP. Τα φράγματά μας είναι ασυμπτωτικά βέλτιστα, καθώς SA ιεραρχία είναι ισχυρότερη από την LS .

Χρησιμοποιούμε μία παραλλαγή της *local-to-global* μεθόδου η οποία χρησιμοποιήθηκε για πρώτη φορά στην εργασία [9] για χαλαρώσεις με τοπικούς περιορισμούς και επεκτάθηκε στην SA ιεραρχία στην εργασία [28]. Βλ. επίσης, [30] για μία ρητή περιγραφή, καθώς και την [23] για μία εφαρμογή για το Max Cut και για άλλα προβλήματα. Σε αυτή την προσέγγιση, η εφικτότητα μιας λύσης για το t-επίπεδο SA επιτυγχάνεται μέσα από το σχεδιασμό πιθανοτικών κατανομών για κάθε περιορισμό του υψηλών διαστάσεων πολυτόπου, έτσι ώστε αυτές οι κατανομές να συμφωνούν τοπικά μεταξύ τους.

Για να αποδειχτεί η εφικτότητα για την κακή λύση CFL δίνουμε μία και μόνο πιθανοτική κατανομή, η οποία είναι πάνω σε λύσεις που ικανοποιούν όλους τους περιορισμούς, εκτός από έναν. Έπειτα, δείχνουμε ότι η λύση που ορίζεται από αυτήν την κατανομή είναι εφικτή για το πλήθος των επιπέδων που θεωρούμε, καθώς οι λύσεις της κατανομής οι οποίες ικανοποιούν τον περιορισμό το κάνουν με μεγάλο περιθώριο. Όσο αυξάνουν τα επίπεδα, καθώς οι περιορισμοί του υψηλών διαστάσεων πολυτόπου γίνονται περισσότερο καθολικοί και λαμβάνουν υπόψη ένα μεγαλύτερο κομμάτι του στιγμιοτύπου και επομένως, το περιθώριο με το οποίο ικανοποιούνται ολοένα και μειώνεται, μέχρι που σε κάποιο επίπεδο σταματάει πια η λύση να είναι εφικτή.

Σημειώνουμε ότι, ενώ το κάτω φράγμα για την SA ιεραρχία υπονοεί το φράγμα για την ασθενέστερη LS ιεραρχία, είναι μη συγκρίσιμο με το φράγμα για τη μικτή LS_+ ιεραρχία. Από την άλλη, υπάρχουν προβλήματα για την οποία η LS_+ έχει καλύτερη επίδοση από όαι οι γραμμικές ιεραρχίες (βλ. Stable Set [49]). Από μία ποιοτική άποψη, είμαστε οι πρώτοι που δίνουμε, τουλάχιστον από όσο γνωρίζουμε, φράγματα για ιεραρχίες σε χαλαρώσεις που έχουν περισσότερες από ενός τύπου μεταβλητές, όσον αφορά τη σημασιολογία τους.

Η τρίτη μας συνεισφορά σε αυτό το κομμάτι (βλ. Θεώρημα 4.3.1) είναι ότι τα *submodular* inequalities που εισήχθησαν στην εργασία [1, 2] για το CFL αποτυγχάνουν να μειώσουν το gap σε σταθερά. Αυτοί οι περιορισμοί γενικεύουν τις flow-cover ανισότητες για το CFL. Επομένως, διαψεύδουμε μια επί πολύ καιρό ανοιχτή εικασία του [46], σύμφωνα με την οποία η προσθήκη των τελευταίων περιορισμών στο κλασικό LP αρκεί για σταθερό integrality gap. Αν και αυτό το αποτέλεσμα δεν αφορά ιεραρχίες προγραμματισμού,το περιλαμβάνουμε σε αυτό το τμήμα της διατριβής, καθώς η μεθοδολογία που χρησιμοποιούμε είναι μία βελτιωμένη εκδοχή της local-global μεθόδου και επομένως, η απόδειξή μας ξεφεύγει από συνηθισμένες integrality gap κατασκευές. Η απόδειξη βασίζεται σε κάποιες δομικές ιδιότητες των ανισοτήτων αυτών, χωρίς να λαμβάνονται υπόψη οι συντελεστές των μεταβλητών.

Στο τρίτο και πιο γενικό μοντέλο γραμμικού προγραμματισμού στο οποίο επεκτείνουμε τα αποτελέσματά μας είναι αυτό των *extended formulations (EFs)*. Τα τελευταία χρόνια έχει γίνει μεγάλη προσπάθεια να κατανοηθεί η δύναμη των πολυωνυμικού μεγέθους χαλαρώσεων συνδυαστικών προβλημάτων. Ο στόχος είναι να δοθεί ένα κάτω φράγμα στο μέγεθος των διατυπώσεων ενός συγκεκριμένου προβλήματος. Οι επεκτεταμένες διατυπώσεις με την εισαγωγή καινούργιων μεταβλητών προσπαθούν να μειώσουν το μέγεθος της περιγραφής ενός πολυτόπου. Το ελάχιστο μέγεθος μιας τέτοιας διατύπωσης για ένα πρόβλημα ονομάζεται *extension complexity* του αντίστοιχου πολυτόπου. Ένα υπερπολυωνυμικό κάτω φράγμα στην extension complexity έχει ξεχωριστό ενδιαφέρον, τόσο στη συνδυαστική βελτιστοποίηση, όσο και στην υπολογιστική πολυπλοκότητα, καθώς αυτό σημαίνει ότι δεν υπάρχει μικρού μεγέθους γραμμικό πρόγραμμα που να λύνει το πρόβλημα. Δεν αποκλείει, βέβαια, την ύπαρξη γενικότερων αλγορίθμων βασισμένων σε γραμμικές χαλαρώσεις, οι οποίοι είναι αποδοτικοί, όπως primal-dual, reprocessing κ.τ.λ.

Στην εργασία του Γιαννακάκη, η οποία μπορούμε να πούμε ότι γέννησε την όλη περιοχή [66], το πρόβλημα της κάτω φραγής του μεγέθους των extended formulations μελετήθηκε για πρώτη φορά: δόθηκαν εκθετικά κάτω φράγματα στο μέγεθος συμμετρικών extended formulations του matching και του TSP πολυτόπου. Ο Γιαννακάκης [66] όρισε μία σημαντική συνδυαστική παράμετρο, την *nonnegative rank* του slack matrix του πολυτόπου P, και έδειξε ότι χαρακτηρίζει την extension complexity του P. Έδωσε, επίσης, μία μέθοδο για την κάτω φραγή της τελευταίας παραμέτρου χρησιμοποιώντας ιδέες από communication complexity [66]. Αυτή η μεθοδολογία χρησιμοποιήθηκε από όλα τα μετέπειτα σχετικά αποτελέσματα.

Οι Fiorini et al. [29] ήραν τη συνθήκη συμμετρίας από το αποτέλεσμα του [66] σχετικά με το TSP πολύτοπο, δίνοντας έτσι το πρώτο παράδειγμα ενός πολυτόπου με εκθετική extension complexity και επομένως, απαντώντας στο επί πολλά έτη του [66]. Πρόσφατα, ο Rothvoß [57] ήρε τη συνθήκη συμμετρίας για το πολύτοπο του matching, απαντώντας στη δεύτερη ανοιχτή ερώτηση του [66].

Ένα πιο γενικό μοντέλο είναι αυτό των προσεγγιστικών extended formulations. Αυτά μελετήθηκαν για πρώτη φορά στο [15] όπου η μέθοδος των [29] επεκτάθηκε σε προσεγγιστικά formulations. Έπειτα, οι Braverman και Moitra [19] έδωσαν κάτω φράγματα χρησιμοποιώντας μεθόδους θεωρίας πληροφορίας. Οι Braun και Pokutta στο [16] περαιτέρω επέκτειναν τα προηγούμενα αποτελέσματα χρησιμοποιώντας τη θεωρία πληροφορίας και πρόσφατα στο [17] επέκτειναν το αποτέλεσμα του [57] στην προσεγγιστική εκδοχή του matching πολυτόπου. Πολύ πρόσφατα στο [70] το μοντέλο των προσεγγιστικών extended formulations απλοποιήθηκε και ταυτόχρονα γενικεύθηκε σε αυτό των *LP Formulations*.

Τα αποτελέσματά μας σε συτό το μοντέλο (βλέπε Κεφ. 4.3) είναι τα εξής: προτείνουμε μια καινούργια ισχυρή μέθοδο για την κάτω φραγή του μεγέθους των προσεγγιστικών επεκτεταμένων διατυπώσεων η οποία στηρίζεται στη σύλληψη ότι ισχυρά σύνολα μεταβλητών με συγκεκριμένες κωδικοποιήσεις λύσεων είναι ικανά να προσομοιώσουν οποιοδήποτε γραμμικό πρόγραμμα χωρίς να αυξάνουν το μέγεθός του. Η συνεισφορά μας, συγκεκριμένα, είναι η ακόλουθη.

Πρώτα εισάγουμε δύο ισχυρές οικογένειες από extended formulations (ή relaxations) ενός δεδομένου πολυτόπου-προβλήματος τις οποίες οναμάζουμε *product formulations* και *distributional formulations* αντίστοιχα. Τα product relaxations έχουν εμπνευστεί από τη μελέτη της Sherali-Adams ιεραρχίας. Τα distributional στοχεύουν να κωδικοποιήσουν το πρόγραμμα με τέτοιο τρόπο, ώστε η εφικτή περιοχή να είναι μία κατανομή πάνω σε (κυρτός συνδυασμός) εφικτών ακέραιων λύσεων.

Αποδεικνύουμε στο Θεώρημα 5.2.1 (Θεώρημα 5.2.3) ότι για οποιοδήποτε ρ-προσεγγιστικό extended formulation ενός 0-1 πολυτόπου, υπάρχει ένα product (distributional) formulation του ίδιου μεγέθους το οποίο είναι τουλάχιστον τόσο ισχυρό. Η απόδειξη είναι σύντομη και κομψή. Στο θεώρημα 5.2.1 (5.2.3) ανάγεται η κάτω φραγή του μεγέθους κάποιου αυθαίρετου γραμμικού προγράμματος, το οποίο χρησιμοποιεί ένα μη γνωστό χώρο μεταβλητών και κωδικοποίηση, ενός πολυτόπου P, στην κάτω φραγή του μεγέθους των product (distributional) formulations του P. Στο χώρο των product (distributional) έχουμε το πλεονέκτημα της γνώσης των μεταβλητών και των κωδικοποιήσεων. Επεκτείνουμε τον ορισμό των product relaxations και τη μεθοδολογία μας σε mixed integer σύνολα, όμως σε αυτήν την περίπτωση μπορέσαμε να δείξουμε ότι τα *mixed product relaxations* είναι τόσο ισχυρά όσο μια ειδική οικογένεια extended formulations (βλ. Θεώρημα 5.2.3).

Σημειώνουμε ότι η προσέγγισή μας δε βασίζεται στην έννοια του slack matrix που εισήχθη από τον Γιαννακάκη [66], στην οποία βασίστηκαν όλα τα έως τώρα σχετικά αποτελέσματα.

Έπειτα, δίνουμε μία μέθοδο για την κάτω φραγή του μεγέθους των χαλαρώσεων όταν είναι γνωστές οι μεταβλητές, και συγκεκριμένα για τα product (distributional) formulations. Παρόμοια επιχειρήματα έχουν χρησιμοποιηθεί ανεξάρτητα για την κάτω φραγή των εδρών πολυτόπων (βλ. για παράδειγμα [37]), αλλά πριν από τη συνεισφορά μας φαινόταν να μην εφαρμόζονται σε γενικά EFs. Η μέθοδος έχει ως εξής: πρώτα ορίζουμε ένα σύνολο διανυσμάτων στο χώρο της χαλάρωσης έτσι ώστε για καθένα από αυτά να υπάρχει αντικειμενική συνάρτηση η οποία να πιστοποιεί integrality gap ρ. Έπειτα, δείχνουμε ότι για οποιοδήποτε μερισμό του προηγούμενου συνόλου σε κ μέρη, θα υπάρχει κάποιο το οποίο περιέχει conflicting διανύσματα. Ένα σύνολο διανυσμάτων θα λέγαται *conflicting* εάν το κυρτό του περίβλημα έχει μη κενή τομή με το σύνολο $\{z^x \mid x \in P(x) \cap \{0,1\}^n\}$, το οποίο πάντοτε περιέχεται στην εφικτή περιοχή ενός product relaxation – εδώ z^x είναι η κωδικοποίηση της ακέραιας λύσης x στις μεταβλητές του product formulations. Επομένως, τουλάχιστον κ ανισότητες χρειάζονται για να διαχωρίσουν τα διανύσματα αυτά από την εφικτή περιοχή και, άρα, το κ είναι ένα κάτω φράγμα στο μέγεθος οποιουδήποτε ρ-προσεγγιστικού product formulation. Τα προηγούμενα μπορούν να διατυπωθούν στη γλώσσα των υπεργραφημάτων και δίνουμε έτσι έναν χαρακτηρισμό της extension complexity με γραφοθεωρητικούς όρους (βλ. Θεώρημα 5.4.2).

Δίνουμε ένα παράδειγμα της μεθοδολογίας μας με το να δείξουμε με το Θεώρημα 5.6.3 εκθετικό κάτω φράγμα στο μέγεθος οποιουδήποτε $O(N)$-προσεγγιστικού mixed product relaxation για το CFL πολύτοπο, όπου N είναι το πλήθος των facilities του στιγμιοτύπου. Αυτό το αποτέλεσμα δείχνει (βλ. Θεώρημα 5.6.4) ότι το $\Omega(N)$-επιπέδου SA relaxation για το CFL, το οποίο λαμβάνουμε από οποιοδήποτε αρχικό LP μεγέθους $2^{o(N)}$ ορισμένο στις κλασικές μεταβλητές, έχει μη φραγμένο gap $\Omega(N)$. Αυτό απαντάει σε ένα ανοιχτό ερώτημα των [6] σχετικά με το αν υπάρχουν LP relaxations στα οποία εάν εφαρμόσουμε lift-and-project μεθόδους πετυχαίνουμε τη δύναμη των αλγοριθμικών βημάτων προδιεργασίας του CFL. Αυτό είναι το πρώτο, από όσο γνωρίζουμε, αποτέλεσμα για την *universal* SA διεργασία, η οποία είναι ανεξάρτητη της αρχικής χαλάρωσης K.

Στο τελευταίο κομμάτι της συνεισφοράς μας εισάγουμε και μελετούμε την οικογένεια των proper relaxations τα οποία γενικεύουν configuration τύπου γραμμικά προγράμματα. Το *Configuration LP* χρησιμοποιήθηκε από τους Bansal και Sviridenko [12] για το Santa Claus και έχει χρησιμοπιηθεί με επιτυχία σε resource allocation και scheduling προβλήματα (π.χ, [61]). Το ανάλογο του Configuration LP για το facility location υπάρχει ήδη, ονομάζεται *star relaxation* (βλ. [34]). Στο star relaxation κάθε μεταβλητή αντιστοιχεί σε ένα *star*, δηλαδή, ένα facility f και ένα σύνολο από clients που ανατίθενται στο f. Το φυσικό star relaxation για τα CFL και LBFL είναι ισοδύναμο με το standard LP και επομένως έχει μη φραγμένο integrality gap. Γενικεύουμε την ιδέα του αστεριού με την εισαγωγή τον επονομαζομένων *classes*. Μία *class* αποτελείτε από ένα σύνολο αυθέρετου πλήθους από facilities και clients μαζί με μία ανάθεση του κάθε client σε ένα facility του συνόλου. Ο ορισμός μίας κλάσης ποικίλει από απλές 'τοπικές' σε πιο 'καθολικά' κομμάτια του στιγμιοτύπου. Ένα *proper relaxation* για ένα στιγμιότυπο περιλαμβάνει ένα σύνολο $\mathcal{C}$ από classes και μία μεταβλητή απόφασης για κάθε κλάση. Επιτρέουμε μεγάλη ελευθερία στον ορισμό του $\mathcal{C}$· η μόνη απαίτητση είναι η χαλάρωση να είναι συμμετρική και valid. Η *complexity* α ενός proper relaxation είναι το μέγιστο ποσοστό από facilities που περιέχονται στο $\mathcal{C}$. Στο Θεώρημα 3.1.1 χαρακτηρίζουμε την συμπεριφορά των proper relaxations για το CFL και το LBFL μέσω ενός φαινομένου κατωφλίου: για πολυπλοκότητα μικρότερης της μέγιστης δυνατής έχουμε μη γραγμένο integrality gap, ενώ υπάρχουν relaxations μέγιστης πολυπλοκότητας με gap ακριβώς 1.

CONTENTS

FIGURES

1. INTRODUCTION

In the field of *Combinatorial Optimization* we are generally concerned with the efficient finding of an object s from a finite set of objects S which maximizes the value of a given objective function $f : S \to \mathbb{R}$. We call an object $s \in S$ a *feasible solution* to the problem, and we call S the *set of feasible solutions*. A typical problem $\Pi(n)$ of combinatorial optimization, parameterized by the *dimension* n, consists of a set of solutions $\mathcal{S}(n)$ and a set of *admissible objective functions* $W(n)$. An *instance* I of dimension n consists of a pair (S, w) with $S \in \mathcal{S}(n)$ and $w \in W(n)$. An *optimal solution* of an instance I is an $s^{opt} \in S$ such that $w(s^{opt}) = \max_{s \in S} w(s)$. For the combinatorial problem to be interesting from a complexity prospective we require that S is given in some efficient encoding that has length polynomial in n while $|S|$ might be exponential in n. We also require that the objective function w is given using some efficient encoding instead of, let us say, pairs $(s, w(s))$ and moreover we require that given an $s \in S$, $w(s)$ can be computed in time polynomial in n. For example, the problem of finding the shortest path between two designated vertices u, v of a weighted graph $G(V, E)$ is the following Combinatorial Optimization problem: $\mathcal{S}(n)$ contains, for every graph $G(V, E)$ on n vertices and every pair u, v with $u, v \in V$, the set of all possible paths from u to v in G. $W(n)$ consists of all possible assignments of weights to the edges $e \in E$. An efficient encoding of an instance could consist of the labels of the vertices u and v together with a weighted $n \times n$ adjacency matrix A. The set of feasible (u, v)-paths and their corresponding values with respect to w can be uniquely inferred from the latter encoding, which has size polynomial in n. We wish to solve a combinatorial problem using a number of algorithmic steps that is polynomial in n. Assuming that P $\neq$ NP, many interesting combinatorial optimization problems cannot be solved efficiently – these problems are called NP-hard optimization problems.

In the more specific field of approximation algorithms we are interested in understanding the approximability of such NP-hard optimization problems. On the positive side we would like to provide an efficient (polynomial-time) algorithm that is guaranteed to find a solution within a small multiplicative factor ρ of the optimal. We refer to such an algorithm as a *ρ-approximation algorithm*. On the negative side, we would like to prove our limits on designing approximation algorithms by proving an *inapproximability* or *hardness of approximation* result. Typically in such a result it is stated that the problem in question is not efficiently approximable within some factor ρ, given some complexity assumption such as P $\neq$ NP or the Unique Games Conjecture (*UGC*) [69]. The ultimate goal is to precisely characterize the complexity of an approximation problem, i.e., to prove that a ρ approximation is achievable while better than ρ is not possible.

Without complexity-theoretic assumptions, lower bounds on the time required by an algorithm that solves a combinatorial problem exactly or approximatively for a given approximation guarantee, or lower bounds on the quality of the solution of an efficient (polynomial) approximation algorithm are very hard to establish. In fact, even when we do not aim at exponential lower bounds, as this goal would settle the greatest open problem of theoretical computer science P vs NP, even non-trivial polynomial lower bounds for the majority of the optimization problems are not known.

1.1 The motivation and the goal of this thesis

Given the aforementioned difficulty in the establishment of true lower bounds, in this thesis we are concerned with the establishment of lower bounds that are independent of assumptions, but for restricted types of algorithmic models, where the algorithm consists entirely of formulating an instance of the problem as a linear program and solving it. As linear programming is one of the strongest tools in algorithm design and arguably the most popular one in the design of approximation algorithms, the direction of exposing the weakness of linear programming for solving or even estimating the value of the optimal solution is a rather interesting one. More specifically, in the first part of the thesis, we are concerned with the systems of the *Lift-and-Project* linear program hierarchies: in particular with the one defined by Lovász and Schrijver [49] and with the one defined by Sherali and Adams [59]. In the second part, we are concerned with the more general model of *Extended Formulations*. In our relevant results regarding the Lift-and-Project hierarchies and Extended Formulations we focus on various versions of the *facility location* problem and in particular the *capacitated* version of the facility location problem, although the methodology and techniques we develop, especially for extended formulation lower bounds, are general. In the third part we are concerned with the strength of general configuration-type linear programs, which are not captured by the previous models as they are in general exponential in size.

The message from those results is that "strict" linear programming, i.e., using a relaxation which is independent of the objective function and without relying on arbitrary algorithmic steps such as preprocessing of the input, fails to provide satisfactory approximations to assignment problems with restrictions such as capacity constraints.

1.2 Overview of the thesis's contributions: results and connections to previous work

As mentioned earlier, this thesis is naturally divided in three parts regarding the subfield in which the corresponding results reside: One part regarding the quality of solution for the CFL problem obtained by capitalizing on LP hierarchies, a second part, and perhaps the most important contribution of the thesis, regarding the development of a methodology for lower bounding the size of Extended Formulations and applications to the CFL problem, and one smaller part regarding the characterization of the strength of generalized configuration linear programs for CFL. In this overview we first define CFL and give a detailed backround, and then we give an introduction to each topic corresponding to each part: we mention relevant work and we briefly present the exact contribution and results obtained by the research conducted in the context of this thesis.

1.2.1 Approximating Facility Location

Facility location is one of the most well-studied families of models in combinatorial optimization. In the *uncapacitated facility location* problem (UFL) we are given a set F of facilities and a set C of clients. We may open facility i by paying its opening cost f_i and we may assign client j to facility i by paying the connection cost c_{ij}. We are asked to open a subset $F' \subseteq F$ of the facilities and assign each client to an open facility. The goal is to minimize the total opening and connection cost. The approximability of UFL was settled by an $O(\log |C|)$-approximation [33], which via a reduction from Set Cover is asymptotically best possible, unless P = NP [55]. In *metric* UFL the service costs satisfy the following variant of the triangle inequality: $c_{ij} \le c_{ij'} + c_{i'j'} + c_{i'j}$ for any $i, i' \in F$ and $j, j' \in C$. This natural special case of UFL is approximable within a constant-factor, and many improved results have been published over the years. In those, LP-based methods, such as filtering, randomized rounding and the primal-dual method have been particularly prominent (see, e.g., [65]). After a long series of papers the currently best approximation ratio for metric UFL is 1.488 [47], while the best known lower bound is 1.463, unless P = NP ([64]). The proximity of the known upper and lower bound imply that the approximation of the metric case of UFLis almost settled.

CFL is the generalization of metric UFL where every facility i has a capacity u_i that specifies the maximum number of clients that may be assigned to i. In *uniform* CFL all facilities have the same capacity U. Finding an approximation algorithm for CFL that uses a linear programming lower bound was until recently a notorious open problem. The natural LP relaxations have an unbounded integrality gap and up to the recent breakthrough of [7], the only known $O(1)$-approximation algorithms were based on local search, with the currently best ratios being 5 [11] for the non-uniform and 3 [4] for the uniform case respectively. In the special case where all facility costs are equal, CFL admits an LP-based 5-approximation [46]. Williamson and Shmoys [65], stated the design of a relaxation-based algorithm for CFL as one of the top 10 open problems in approximation algorithms. Apart from the apparent theoretical value of the previous question, the quest of finding LP-based algorithms have an important practical contribution: comparing the LP optimum against the solution output by an LP-based algorithm establishes a guarantee that is at least as strong as the one established a priori by worst-case analysis. In contrast, when a local search algorithm terminates, it is not at all clear what the lower bound is. Very recently, An et al. [7] gave a polynomial-time LP-based 288-approximation algorithm, thus answering the open question of [65]. The LP in [7] has exponential size and is not known to be separable in polynomial time. Therefore the question on the existence of an efficient, compact, linear relaxation for CFL remains open. The series of our results regarding the LP-(in)approximability of CFL can be taken as very strong evidence that such a relaxation does not exists.

The Lower Bound Facility Location (LBFL) is in a sense the opposite problem to CFL. In an LBFL instance every facility i comes with a lower bound b_i which is the minimum number of clients that must be assigned to i if we open it. In *uniform* LBFL all the lower bounds have the same value B. LBFL is even less well-understood than CFL. The first approximation algorithm for the uniform case had a performance guarantee of 448 [62], which has been improved to 82.6 [5]. Both use local search. Interestingly, the LBFL algorithms from [62, 5] both use a CFL algorithm on a suitable instance as a subroutine.

Other related work on the facility location literature include the following: Koropulu et al. [42] gave the first constant-factor approximation algorithm for uniform CFL. Chudak and Williamson [25] obtained a ratio of 6, subsequently improved to 5.83 [22]. Pál et al. [52] gave the first constant-factor approximation for non-uniform CFL. This was improved by Mahdian and Pál [50] and Zhang et al. [67] to a 5.83-approximation algorithm. As mentioned, the currently best guarantee is 5, due to Bansal et al. [11], while for the uniform case a 3-approximation exists [4]. All these results use local search. In the *soft-capacitated* facility location problem one is allowed to open multiple copies of the same facility. Work on this problem includes [60, 24, 25, 35]. As observed in [34] a ρ-approximation for UFL yields a 2ρ-approximation for the case with soft capacities. Mahdian, Ye and Zhang [51] noticed a sharper tradeoff and obtained a 2-approximation. A tradeoff between the blowup of capacities and the cost approximation for CFL was studied in [3]. Bicriteria approximations for LBFL appeared in [38, 32]. For hard capacities and general demands the feasibility of the *unsplittable* case, where the demand of each client has to be assigned to a single facility, is NP-complete, as PARTITION reduces to it. Bateni and Hajiaghayi [13] considered the unsplittable problem with an $(1+\epsilon)$ violation of the capacities and obtained an $O(\log n)$-approximation.

For related problems there are some recent interesting results. Improved approximations were given for k-median [48] and capacitated k-center [27, 6], problems closely related to facility location. For both, the improvements are obtained by LP-based techniques that include preprocessing of the instance in order to defeat the known integrality gap. For k-median, the authors of [48] state that their $(1+\sqrt{3}+\epsilon)$-approximation algorithm can be converted to a rounding algorithm on an $O(\frac{1}{\epsilon^2})$-level LP in the Sherali-Adams (SA) lift-and-project hierarchy. They propose exploring the direction of using SA for approximating CFL. An et al. [6] raise as an important question to understand the power of lift-and-project methods for capacitated location problems, including whether those methods capture "automatically" relevant preprocessing steps.

1.2.2 Lift-and-Project methods and the resulting Hierarchies as model of LP computation

A lot of effort has been devoted to understanding the quality of relaxations of 0-1 polytopes obtained by an iterative lift-and-project procedure. Such procedures define hierarchies of successively stronger relaxations, where valid inequalities are added at each level. At level at most d, where d is the number of variables, all valid inequalities have been added and thus the integer polytope is expressed. Relevant methods include those developed by Balas et al. [10], Lovász and Schrijver [49] (for linear and semidefinite programs), Sherali and Adams [59], Lasserre [43] (for semidefinite programs). See [44] for a comparative discussion.

The seminal work of Arora et al. [9], studied integrality gaps of families of relaxations for Vertex Cover, including relaxations in the Lovász-Schrijver (LS) hierarchy. This paper introduced the use of hierarchies as a restricted model of computation for obtaining LP-based hardness of approximation results. Proving that the integrality gap for a problem remains large for many levels of a hierarchy is an unconditional guarantee against the class of relaxation-based algorithms obtainable through the specific method. At the same time, if an LP relaxation maintains a gap of g for a linear number of levels, one can take this as evidence that compact, polynomial-size, relaxations are unlikely to yield approximations better than g (see also [58]). In fact, the former belief is now a theorem for maximum constraint satisfaction problems: in terms of approximation, LPs of size d^k, are exactly as powerful as $O(k)$-level Sherali-Adams (SA) relaxations [21]. From the algorithmic side, relaxations at the first $O(1)$ levels of LP hierarchies can be optimized in polynomial time as long as a weak separation oracle exists [49, 59].

We note that there are known examples of polytopes, such as the matching polytope, whose convex hull is obtained at high levels of the LS and the SA hierarchies [53], while at the same time there are known linear descriptions of the integer hull that can be optimized in polynomial time. The separation algorithm for the matching LP however has algorithmic steps that were not known to be translatable into a polynomial-size linear program. After the breakthrough result of Rothvoß [56] we positively know this separation algorithm cannot be captured by a compact linear program and thus a polynomial-size formulation cannot be obtained. So, at least for matching, the weakness of hierarchies is due to an intrinsic difficulty of any compact LP formulation for the problem supporting the general belief we stated earlier (se e.g. [58]). In fact the latter belief is now a theorem for maximum constraint satisfaction problems: in [21] it was proved that in terms of approximating maximum constraint satisfaction problems, LPs of size $O(n^k)$ are exactly as powerful as $O(k)$-level relaxations in the Sherali-Adams hierarchy.

1.2.2.1 Our results on hierarchy-based gaps for CFL

We give impossibility results on arguably the most promising directions for obtaining efficient linear strengthened relaxations for CFL and in doing so we answer open problems from the literature.

Our first result of this part of our contribution (cf. Theorem 4.1.1) is that there is an instance with $\Theta(n)$ facilities and $\Theta(n^4)$ clients on which the relaxations produced at $\Omega(n)$ levels when the LS procedure is applied on the natural CFL LP have an integrality gap of $\Omega(n)$. The natural LP, defined in Section 2.2, has a facility opening variable y_i, for every $i \in F$, and an assignment variable x_{ij}, for every $i \in F$, and client $j \in C$.

The definition of the LS hierarchy is inductive (see chapter 3.3.3). In order to obtain the relaxation at level $t + 1$, one applies the LS procedure to the relaxation at level t. Therefore, instead of levels, one may equivalently refer to the number of *rounds* for which the procedure has been applied. It is well-known that the LS procedure extends to mixed 0-1 programs [49, 10] such as CFL with general client demands. By lifting only the binary variables, the convex hull of the mixed integer feasible set is known to be obtained the latest at the pth level of the LS hierarchy, where p is the number of those binary variables ([49], [10, Theorem 2.6]). For CFL, p equals the number $|F|$ of facilities. In the instances we consider clients have unit demands and the capacities are integer; it is well-known that in this case the integer polytope and the mixed integer one (where fractional client assignments are allowed) are the same. In the proof of Theorem 4.1.1, we treat all variables as binary. In every round we obtain a polytope which is at least as tight as the one obtained when only the facility-opening variables are binary. Therefore our lower bound of $\Omega(n)$ applies also to the mixed integer LS procedure and is linear in the parameter p.

In our proof we use a simple reformulation of the feasibility conditions of [49] for the strengthened relaxations, which is of independent interest in the context of the LS literature. In a nutshell, the LS procedure reduces the survival at level t of the original bad solution s to the existence of a set of vectors with appropriate structure. These vectors act as “witnesses” of the survival of s and after suitable scaling they can be collected in the so-called *protection matrix* [9] at that level.

The LS_+ procedure is the stronger version of LS where one additionally requires that every protection matrix is positive semidefinite. The *mixed* LS_+ *procedure* for a mixed integer program is the version of LS_+ where one lifts only the 0-1 variables and requires that the resulting protection matrix is positive semidefinite (see, [10], [26]). Theorem 4.1.4, which follows almost immediately from the proof of our LS result, shows that the $\Omega(n)$ gap applies for $\Omega(n)$ rounds of mixed LS_+ as well. In fact, Theorem 4.1.4 shows this result for the stronger procedure where one lifts also the client assignment variables, but one requires that only the minor of the protection matrix that corresponds to the facility opening variables is positive semidefinite.

We then show that the LPs obtained from the natural relaxation for CFL at $\Omega(n)$ levels of the stronger SA hierarchy have a gap of $\Omega(n)$ on the same family of instances used for the LS result, with $|F| = \Theta(n)$ and $|C| = \Theta(n^4)$, giving the second contribution of this part.

This result answers the questions of [48] and [6] stated above as far as the natural LP is concerned. Our bound is asymptotically tight since the relaxation obtained at every level of the SA hierarchy is at least as strong as the one obtained at the same level of LS.

We use a variation of the *local-to-global* method which was implicit in [9] for local-constraint relaxations and was then extended to the SA hierarchy in [28]. See also [30] for an explicit description and [23] for applications to Max Cut and other problems. In this approach, the feasibility of a solution for the t-level SA relaxation is established through the design of a set of appropriate distributions over feasible integer solutions for each constraint such that these global distributions agree with each other locally on relevant events.

To prove the feasibility of a bad solution for CFL we first construct a single distribution whose support of assignments satisfies all but one type of constraints. Then we show the solution to the lifted polytope defined by this distribution satisfies that constraint too: this is because the assignments in the support of the distribution that satisfy the constraint do so with enough slackness so that the average vector of the lifted polytope corresponding to the distribution satisfies the relevant constraint of the lifted polytope too for an asymptotically tight number of levels. We note that, as the level of the hierarchy increases, the constraints of the lifted polytope become aware of a larger portion of the instance and, thus, it becomes increasingly more difficult to fool them and the slackness of the constraint reduces until the solution becomes infeasible at high levels of the hierarchy.

We note that, while the lower bound on the SA hierarchy implies the bound on the weaker LS hierarchy, it is in principle incomparable to the bound on the mixed LS_+ hierarchy. On the other hand, there are problems for which LS_+ is known to perform better than the LP hierarchies. For Stable Set one round of LS_+ is strictly stronger that one round of LS [49] and for Max Cut one round of LS_+ is strictly stronger than $\Omega(n^\delta)$ levels of SA [23]. From a qualitative aspect, we give the first, to our knowledge, hierarchy bounds for a relaxation where variables have more than one type of semantics, namely the facility opening and the client assignment type. Compare this, for example, with the Knapsack and Max Cut LPs that contain each one type of variable.

Our third contribution in this part (cf. Theorem 4.3.1) is that the *submodular* inequalities introduced in [1, 2] for CFL fail to reduce the gap of the classic relaxation to constant. These constraints generalize the flow-cover inequalities for CFL. Thus we disprove the long-standing conjecture of [46] that the addition of the latter to the classic LP suffices for a constant integrality gap. Although this is not a result that concerns linear programming hierarchies, we included its presentation in that part of the thesis because the methodology we use is inspired by the local-to-global method and thus our proof deviates from standard integrality gap constructions. In fact we take the idea of fooling local constraints a little further: the bad solution fools every inequality π because its part that is *visible* to π, i.e., the variables in the support of π, can be extended to a solution that is a convex combination of feasible integer solutions for that instance or it is a convex combination of feasible solutions to another instance for which the same inequality is valid. Our proof relies on simple structural properties of the inequalities, disregarding the exact coefficients of the variables. Our ideas can be extended to even more general families such as the *submodular inequalities* [1] (cf. Theorem 4.3.1)

1.2.3 Extended Formulations: the currently most general mode of LP computation

In the past few years there has been an increasing interest in exposing the limitations of compact LP formulations for combinatorial optimization problems. The goal is to show a lower bound on the size of *extended formulations (EFs)* for a particular problem. Extended formulations add extra variables to the natural problem space; the increase in dimension may yield a smaller number of facets. The minimum size over all extended formulations is the *extension complexity* of the corresponding polytope. A superpolynomial lower bound on the extension complexity is of intrinsic interest in both polyhedral combinatorics and combinatorial optimization and implies that there is no polynomial-time algorithm relying purely on the solution of a compact linear program. It does not however rule out efficient LP-based algorithms that combine algorithmic steps of arbitrary type, such as preprocessing, primal-dual, etc., with linear programming.

In the seminal paper of Yannakakis [66] the problem of lower bounding the size of extended formulations was considered for the first time: exponential lower bounds were proved for symmetric extended formulations of the matching and TSP polytopes. Yannakakis [66] identified also a crucial combinatorial parameter, the *nonnegative rank* of the slack matrix of the underlying polytope P, and he showed that it equals the extension complexity of P. A strong connection of the extension complexity of a polytope to communication complexity was made in [66], by showing that the nonnegative rank of the slack matrix is at least the size of its minimum rectangle cover. That connection has been exploited in several results on the extension complexity of polytopes.

Fiorini et al. [29] lifted the symmetry condition on the result of [66] regarding the TSP polytope, giving the first example of a polytope with exponential extension complexity and thus answering a long-standing open problem of [66]. The result was obtained by showing that the correlation polytope has exponential extension complexity which in turn was shown using the communication complexity framework established in [66]. Recently, Rothvoß [57] removed the symmetry condition for the matching polytope as well, answering the second long-standing open question of [66]. This was done by a breakthrough in bounding a refined version of the rectangle covering number.

A more general question is that of the size of approximate extended formulations. This problem was first considered in [15] where the methodology of [29] was extended to approximate formulations and an exponential bound for the linear encoding of the $n^{1/2-\varepsilon}$-approximate clique problem was given. Subsequently, Braverman and Moitra [19] extended the former bound to $n^{1-\varepsilon}$-approximate formulations of the clique, following a new, information theoretic, approach. Braun and Pokutta in [16] further strengthened the lower bounds by introducing the notion of common information. Recently, Braun and Pokutta [17] extended the result of [57] to approximate formulations of the matching polytope by combining ideas of the latter with the notion of common information. Very recently in [70] the model of approximate extended formulations was simplified and at the same time generalized to that of *LP Formulations*.

In [21] it was proved that in terms of approximating maximum constraint satisfaction problems (CSPs), LPs of size $O(n^k)$ are exactly as powerful as $O(k)$-level relaxations in the Sherali-Adams hierarchy. Their proof differs from previous work in showing that polynomials of low degree can approximate the functional version of the factorization theorem of [66]. This approach offers one more method of lower bounding the extension complexity and was generalized to SDP extension complexity for constraint satisfaction problems in [68].

1.2.3.1 Our contribution on Extended Formulations

In the relevant part of the thesis (see chapter 4.3) we propose a new intuitive, geometric approach for proving lower bounds on the size of approximate extended formulations that relies on an insight on the expressive strength of “strong" sets of variables and encodings. Our contribution is summarized by the following.

First we introduce two very strong families of extended formulations (or relaxations) of a given polytope which we call *product formulations* and *distributional formulations*. The product relaxations are inspired by the study of the Sherali-Adams hierarchy – the variables have the intuitive meaning of corresponding to products over sets of variables from the original space. The distributional are intended to encode the problem in such a way that the feasible region is a straightforward distribution of (convex combination of) feasible integer solutions. (See Section 5.1 for the necessary definitions).

We prove in Theorem 5.2.1 (Theorem 5.2.3) that for any ρ-approximate extended formulation of a 0-1 polytope there is a product (distributional) formulation of the same size that is at least as strong. The proof is short and accessible. Theorem 5.2.1 (5.2.3) reduces lower bounding the size of an extended formulation, which uses some unknown space and encoding, of a polytope P, to lower bounding the size of product (distributional) formulations of P. In the product (distributional) space we have the concrete advantage of knowing the section of the target relaxation. We extend the definition of product relaxations and our methodology to mixed integer sets. However in this case we are able to show that *mixed product relaxations* are at least as powerful as a special family of extended formulations (cf. Theorem 5.2.3).

We note that our approach does not rely on the notion of the slack matrix introduced by Yannakakis [66]. It differs from that of [21] in which the slack functions of the factorization theorem [66] were shown to be approximable, for max CSPs, by low-degree polynomials and thus SA gaps are transferred to general linear programs.

Then we propose a methodology for proving lower bounds for relaxations for which the encoding of solutions is known, and in particular for product (distributional) formulations. Similar arguments have been used in the context of bounding the number of facets of specific polyhedra (see for example [37]) , but prior to our work, they seemed inapplicable for lower bounding the size of arbitrary EFs which lift the polytope in arbitrary variable spaces. The method is the following: first we define a set of vectors in the space of the relaxation such that for each one of those vectors there is an admissible objective function witnessing an integrality gap of ρ. We call that set of vectors the *core*. Then we show that, for any partition of the core into fewer than κ parts, there must be some part containing a set of conflicting vectors. A set of infeasible vectors is *conflicting* if its convex hull has nonempty intersection with the convex hull of $\{z^x \mid x \in P(x) \cap \{0,1\}^n\}$, which is always included in the feasible region of a product relaxation – here z^x is the encoding of feasible solution x to the variables of product formulations. Thus, we get that at least κ inequalities are needed to separate the members of the core from the feasible region and so κ is a lower bound on the size of any ρ-approximate product formulation. By considering the hypergraph whose set of vertices corresponds to the aforementioned set of vectors and whose set of hyperedges corresponds to the sets of conflicting vectors, the chromatic number of the hypergraph is a lower bound on the size of every ρ-approximate extended formulation (cf. Theorem 5.4.2). Moreover, there is always a core such that the chromatic number of the resulting, possibly infinite, hypergraph equals the extension complexity of the polytope at hand. Thus we give a characterization of extension complexity in

Theorem 5.4.2 which can be seen as an alternative to the nonnegative rank of the slack matrix.

When arguing about the polyhedral complexity of a specific polytope, i.e., the minimum size of its formulation in the original variable space, the above method can always be simplified to finding a set of gap–inducing vectors with the property that (almost) any pair of them are conflicting. The underlying hypergraph reduces then to a simple graph that is very dense, almost a clique, and thus has high chromatic number. We used this idea in [40] to derive exponential bounds on the polyhedral complexity of approximate metric capacitated facility location, where only the classic variables are used (cf. Theorem 5.5.1) – the exposition can be found in chapter 4.3 of this thesis and serves as a warm up towards the main contributions mentioned in the previous paragraphs.

We exhibit a concrete application of our methodology by proving in Theorem 5.6.3 an exponential lower bound on the size of any $O(N)$-approximate mixed product relaxation for the CFL polytope, where N is the number of facilities in the instance. This result can be shown to imply (cf. Theorem 5.6.4) that the $\Omega(N)$-level SA relaxation for CFL, which is obtained from any starting LP of size $2^{o(N)}$ defined on the classic set of variables, has unbounded gap $\Omega(N)$. Note, that it is well-known that by lifting only the facility variables, at N levels the integer polytope is obtained for CFL [10]. This settles the open question of [6] whether there are LP relaxations upon which the application of lift-and-project methods captures the strength of preprocessing steps for CFL. This result establishes for the first time such a trade-off for a *universal* SA procedure that is independent of the starting relaxation K. The proof follows the methodology outlined above and is different from the standard arguments that apply only to the SA lifting of a specific LP. As we mentioned earlier, the SA construction in [41] applied a variation of the local-global method [28] that constructs an appropriate distribution of solutions for each explicit constraint of the starting LP.

1.2.4 The strength of generalized configuration linear programs for capacitated versions of the facility location problem

In the relevant part of this thesis we introduce and study the family of proper relaxations which are configuration-like linear programs. The so-called *Configuration LP* was used by Bansal and Sviridenko [12] for the Santa Claus problem and has yielded valuable insights, mostly for resource allocation and scheduling problems (e.g., [61]). The analogue of the Configuration LP for facility location already exists, it is the *star relaxation* (see, e.g., [34]). In a star relaxation every variable corresponds to a *star*, i.e., a facility f and a set of clients assigned to f. The natural star relaxation for CFL and LBFL is equivalent to the standard LPs so it has an unbounded integrality gap. We generalize the idea of a star by introducing what we call *classes*. A *class* consists of a set with an arbitrary number of facilities and clients together with an assignment of each client to a facility in the set. The definition of a class can thus vary from simple, "local" assignments of clients to a single facility, to "global" snapshots of the instance that express the assignment of clients to a large set of facilities. A *proper relaxation* for an instance is defined by a collection $\mathcal{C}$ of classes and a decision variable for every class. We allow great freedom in defining $\mathcal{C}$; the only requirement is that the resulting formulation is symmetric and valid. The *complexity* α of a proper relaxation is the maximum fraction of the available facilities that are contained in a class of $\mathcal{C}$. Proper LPs are stronger than the standard relaxation. One can easily construct infinite families of instances where, by increasing the complexity in a proper relaxation, one cuts off more and more fractional solutions. In Theorem 3.1.1 we characterize the behavior of proper relaxations for CFL and LBFL through a sharp threshold result: anything less than maximum complexity results in a gap that is not bounded by any constant, while there are proper relaxations of maximum complexity with a gap of 1. The proof of Theorem 3.1.1 is combinatorial in nature and relies on Lemma 3.2.2. The latter lemma establishes that any proper relaxation which is valid for a family of instances we define must include a variable for a certain class with specific properties. Using that class and its symmetric ones, one can then construct a bad fractional solution that is feasible.

1.3 Outline of the thesis

The outline of this thesis is as follows. In Chapter 1.3 we give some elementary definitions which are expected to be known by the reader in the subsequent chapters.

In Chapter 2.2 we of focus on a more specific family of relaxations that, because of their disregard on size, are not captured by the previous models, while at the same time generalize the popular configuration linear programs. We introduce the proper relaxations and in Sections 3.2 and in Section 3.3 and 3.3 we prove a theorem that characterizes their strength for approximating CFL and LBFL through a sharp threshold phenomenon.

Chapter 3.3.3 includes the hierarchy-related material: the definition of the LS hierarchy in Section 4.1.1, the reformulation of the latter that we use in our proofs in Section 4.1.2, and the proof of our results for the LS and LS_+ hierarchies for CFL in Sections 4.1.3 and 4.1.4 respectively. In Section 4.2 we present our result on the SA hierarchy for CFL. In particular, Subsection 4.2.1 contains necessary background, Subsection 4.2.2 the proof of the result, in Section 4.2.4 we give similar results for LBFL and in Section 4.2.3 we comment on the robustness of our proof. In Section 4.3 we give the definitions of the flow-cover, effective capacity and submodular inequalities and then the proof of our theorem regarding their strength as far as the quality of approximation is concerned.

In Chapter 4.3, we introduce the basic concepts of the field of Extended Formulations in Section 5.1 along with the product and distributional formulations which play a central role through the whole chapter. In Section 5.2 we show the completeness of the two type of formulations as far as compact expressibility of polytopes is concerned, simple or approximate. In Section 5.3 we are concerned with the dependence of the model of Extended Formulations from the enconding and, in particular, we prove the independence of the latter with the help of distributional formulations. In Section 5.4 we present a new methodology for proving lower bounds on the size of extended formulations. In Section 5.5 give a show as a warmup a first application of our method by deriving exponential lower bounds on the size of approximate relaxations for CFL that use the classic variables. In Section 5.6 we use our method once again to give tight bounds on the size of approximate mixed product formulations for CFL.

We conclude in Chapter 5.6.1 with a brief discussion.

The publications that resulted from the work presented in this thesis include the following: [41] contains the results on the SA hierarchy, the flow-cover inequalities and the proper relaxations. The results regarding the LS hierarchy combined with the results of [74] and [41] were published in [75]. The results regarding the Extended Formulations are contained in [39].

2. PRELIMINARIES

In this thesis it is assumed that the reader has basic knowledge of linear algebra, linear programming and discrete mathematics – in particular the reader is assumed to be familiar with basic concepts of discrete probability, combinatorics and graph theory. Familiarity with polytope theory is welcomed - although not essential since every concept used is defined in this chapter or in the upcoming chapters. In what follows, the more frequently used concepts are defined and briefly commented upon.

2.1 Basic Backround

We start with basic concepts of convexity and linear algebra.

Definition 2.1.1. *Let $x_1, ..., x_k \in \mathbb{R}^n$. A point $z \in \mathbb{R}^n$ is a convex combination of $x_1, ..., x_k$ if and only if there exist $\lambda_1, ..., \lambda_k \in \mathbb{R}$ such that $z = \sum_i \lambda_i x_i$, $\sum_i \lambda_i = 1$, and $\lambda_i \geq 0,\ \forall i$.*

So a convex combination of a set of vectors can be understood as a weighted average of those vectors.

Definition 2.1.2. *A set $C \subset \mathbb{R}^n$ is called convex if and only if the following holds: for every $x_1, ..., x_k \in C$, if z is a convex combination of $x_1, ..., x_k$ then $z \in C$.*

The area of convex optimization is concerned with maximizing an objective function over a convex set. Convex sets receive such attention because most of the efficient mathematical programming methods solve systems defining special convex sets such as polyhedra and spectahedra.

Definition 2.1.3. *Let $X \subset \mathbb{R}^n$. The convex hull of X is the set of all points that can be obtained as convex combinations of points in x:*

$$\mathrm{conv}(X) := \{\sum_i \lambda_i x_i \mid x_i \in X, \sum_i \lambda_i = 1, \lambda_i \geq 0,\ \forall i\}$$

The convex hull of X is the minimal convex set containing X. If X is finite then $\mathrm{conv}(X)$ is bounded and closed. When dealing with a pair of points, we abuse notation and denote by $\mathrm{conv}(x_1, x_2)$ the set $\mathrm{conv}(\{x_1, x_2\})$.

Definition 2.1.4. *A set $P \subseteq \mathbb{R}^n$ is a polyhedron if there is an $A \in \mathbb{R}^{m\times n}$ and a $b \in \mathbb{R}^m$ such than $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$.*

According to the above definition a polyhedron is the set of solutions of a system of linear inequalities. If the system is infeasible then we say that it defines an *empty polyhedron*.

Definition 2.1.5. *A set $P \subseteq \mathbb{R}^n$ is a polytope if there is a finite set $X \subseteq \mathbb{R}^n$ such that $P = \mathrm{conv}(X)$.*

The feasible region of a linear program is a polyhedron and moreover, in most cases, a polytope. Since we study the strength of linear programming it is natural that we will focus on polytopes. The equivalent definitions of a polytope, as the convex hull of a finite set of points and as an intersection of halfspaces that is bounded, are very helpful to have in mind (and very basic at the same time see, e.g., [72]). The *extreme points* or *vertices* of a polyhedron (polytope) $P \subseteq \mathbb{R}^n$ are those points $x \in P$ such that there are no $x_1, x_2 \in P$ with $x \in \text{conv}(x_1, x_2)$.

A *Linear Program* (LP) is an optimization problem where we are given a linear objective function $f : x \mapsto c^T x$ and a set of linear inequalities $Ax \geq b$, $x \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$. The goal is to find a vector $x' \in \mathbb{R}^n$ satisfying $Ax' \geq b$ that optimizes the value of the objective function (either maximizes or minimizes depending on the problem). All vectors satisfying the system of inequalities $Ax \leq b$ are called *feasible solutions* to the linear program. The *feasible region* of a linear program is the set of all its feasible solutions and, as mentioned previously, it defines a polyhedron. A popular special case of linear programs upon which we will focus are the 0-1 LPs where the feasible region is contained in the unit hypercube.

A popular approach in combinatorial optimization is to express a problem as 0-1 LP so that the vertices of the corresponding polytope encode the solutions of the instances of the combinatorial problem Π: each solution s to an instance I is associated with a 0-1 vector x^s. Then, given that the objective function can be encoded as a linear function, one can solve the LP and output not only the value of the optimal solution but also the optimum solution to the problem. Then we say that we have a linear *formulation* of the problem. We give to the corresponding polytope the name of the optimization problem (e.g. the TSP polytope, the Perfect Matching polytope) and, as the inequality system of the LP serves as a description of that polytope via inequalities, we often say that we have a description of the polytope instead of saying that we have a formulation of the problem.

Some combinatorial problems admit a *mixed integer formulation*: the feasible region of a mixed integer formulation is a set $X \subseteq \mathbb{R}^d$, for which there is $p \in \{1, \ldots, d-1\}$ such that $d = n + p$ and $X \subseteq \{0,1\}^n \times [0,1]^p$. The integer variables $x_i,\ i \in \{1, ..., n\}$ usually represent some boolean predicates of the solutions of the problem while the real variables $x_i,\ i \in \{n+1, ..., d\}$ usually represent the fraction of the assignment of some quantity which can be defined after the integer variables are fixed.

We say that $s \in S$ is a *ρ-approximate solution* to instance $I(S, f)$ if $f(s) \leq \rho\, \text{opt}(I)$ for a minimization problem Π or $f(s) \geq \frac{1}{\rho}\, \text{opt}(I)$ for a maximization problem. An algorithm A is said to be a *ρ-approximation algorithm* for Π if, on each instance, A produces a ρ-approximate solution, and the running time of A is bounded by a polynomial in the instance size (dimension). We call the factor ρ the *approximation ratio* of the algorithm. From now on we will focus on minimization problems as all the problems studied in the context of this thesis are such. Note that the performance guarantee is at least 1 for minimization problems; an algorithm with a performance guarantee of 1 means that the algorithms solves the problem optimally (or exactly). Some NP-hard optimization problems can be approximated to any desired degree. We say that Π admits a *polynomial time approximation scheme* (PTAS) if there is an algorithm A such that, on each instance I and for each fixed $\epsilon > 0$, A produces an $(1+\epsilon)$-approximate solution and the running time of A, for a fixed ϵ, is bounded by a polynomial in the instance size. Moreover, if the running time of A is bounded by a polynomial in both the instance size and $\frac{1}{\epsilon}$, then we say that A is a *fully polynomial time approximation scheme* (FPTAS) for problem Π.

Since there are polynomial algorithms that solve linear programs (e.g. the ellipsoid method), we cannot hope to formulate NP-hard optimization problems as compact linear programs and in fact even finding a succinct description of an exponential formulation is highly unlikely (see [71]). Following the concept of approximation algorithms we define (approximate) *relaxations* for NP-hard problems: the feasible region of a *valid relaxation* is a superset of the feasible region of the corresponding formulation. Very often we obtain such relaxations by relaxing an integer formulation of the problem of interest, that is by allowing the variables to take real values. Achieving polynomiality in the size of the description of the relaxation comes with a loss in the accuracy as one would expect. As linear relaxations are perhaps the most popular and widely used tool in the design of approximation algorithms (see e.g. [65]) the notion of *integrality gap* is of great importance – we give the definition for minimization problems.

Definition 2.1.6. *Let Π be a combinatorial optimization problem and let $R = R(I)$ be a relaxation of a formulation of Π, parameterized by the instance I of Π. Let $v(R(I))$ be the value of an optimal solution of $R(I)$. Then the integrality gap of R is defined as $\sup_I \frac{\mathrm{opt}(I)}{v(R(I))}$.*

So the integrality gap of a relaxation is the worst-case ratio over all instances of the optimum value of a solution of the instance, to the optimum value of a solution to the relaxation. The integrality gap measures the quality of estimation of the intended relaxation with respect to the admissible objective functions of the problem. Typically we require that the feasible region of a relaxation contains the encoding of solutions $x^s, s \in S$ and in fact in this thesis we will be concerned only with that type of relaxations which are called *valid relaxations*. We note that one could possibly drop the requirement of validity and obtain improved approximate relaxations or define an even more general mode of linear programming computation if one's goal is to show impossibility results.

We conclude this part of the introductory section by stating a way to think of the variables of a linear program that will be of great help to the reader that is interested in Chapter 4.3. A variable intuitively represents a predicate on the feasible solution and, especially, a 0-1 variable represents a predicate that is true or false for a feasible solution; for example the standard formulation of the matching problem for a graph $G(V, E)$ uses variables $x_{uv}, (u, v) \in E$ to denote the inclusion of edge (u, v) in a matching. Of course a variable may represent a more complex predicate like “the length of the shortest (u, v) path in a Hamiltonian tour". A 0-1 variable x can be seen as a function mapping feasible integer solutions to $\{0, 1\}$. In general, given that we encode the feasible solutions as vertices of the hypercube, a variable x can be also considered as a boolean function $x : \{0, 1\}^n \to \mathbb{R}$.

2.2 Facility Location preliminaries

As stated in the introduction we use the facility location problem, and more specifically capacitated versions of it, to apply our ideas and methods. While our hierarchy results are more problem-specific, our extended formulation methodology is of general value and CFL serves as an application for the developed methodology.

Let $I(F, C)$ be an instance of CFL with general client demand and capacities: each client $j \in C$ has a demand of $d_j \in \mathbb{R}^+$ that must be served and each facility $i \in F$ has a capacity U_i that upper bounds the total client demand it can serve. In the classic relaxation for UFL we have variables $y_i, i \in F$ and $x_{ij}, i \in F, j \in C$ with y_i having the meaning of opening facility $i \in F$ and x_{ij} stands for the fraction of the assignment of the demand of j to facility i. The relaxation is the following:

$$\min \sum_{i\in F} f_i y_i + \sum_{i\in F}\sum_{j\in C} x_{ij}c_{ij} \tag{2.1}$$
$$x_{ij} \le y_i \qquad \forall i\in F, \forall j\in C \tag{2.2}$$
$$\sum_{i\in F} x_{ij} = 1 \qquad \forall j\in C \tag{2.3}$$
$$\sum_{j\in C} d_j x_{ij} \le U_i y_i \qquad \forall i\in F \tag{2.4}$$
$$0 \le x_{ij} \le 1 \qquad \forall i\in F, \forall j\in C \tag{2.5}$$
$$0 \le y_i \le 1 \qquad \forall i\in F \tag{2.6}$$

In most of our results we will be concerned with the unpslittable version of CFL with unit client demands and uniform integer capacities U. Each client can be thought of as representing one unit of demand. It is well-known that in such a setting the splittable and unsplittable versions of the problem are equivalent. The following 0-1 IP is the standard valid formulation of capacitated facility location with unsplittable unit demands.

$$\min \sum_{i\in F} f_i y_i + \sum_{i\in F}\sum_{j\in C} x_{ij}c_{ij} \tag{2.7}$$
$$x_{ij} \le y_i \qquad \forall i\in F, \forall j\in C \tag{2.8}$$
$$\sum_{i\in F} x_{ij} = 1 \qquad \forall j\in C \tag{2.9}$$
$$\sum_{j\in C} x_{ij} \le U y_i \qquad \forall i\in F \tag{2.10}$$
$$0 \le x_{ij} \le 1 \qquad \forall i\in F, \forall j\in C \tag{2.11}$$
$$0 \le y_i \le 1 \qquad \forall i\in F \tag{2.12}$$

In the case of unpslittable LBFL with unit demands and integer uniform lower bounds, the standard relaxation is, naturally, the following:

$$\min \sum_{i\in F} f_i y_i + \sum_{i\in F}\sum_{j\in C} x_{ij}c_{ij} \tag{2.13}$$
$$x_{ij} \le y_i \qquad \forall i\in F, \forall j\in C \tag{2.14}$$
$$\sum_{i\in F} x_{ij} = 1 \qquad \forall j\in C \tag{2.15}$$
$$\sum_{j\in C} x_{ij} \ge B y_i \qquad \forall i\in F \tag{2.16}$$
$$0 \le x_{ij} \le 1 \qquad \forall i\in F, \forall j\in C \tag{2.17}$$
$$0 \le y_i \le 1 \qquad \forall i\in F \tag{2.18}$$

In the rest of the thesis we refer to the above LPs by the term *(LP-classic)* – it would be clear from the context to which version we are referring. It is easy to see that (LP-classic) has a super-constant integrality gap: consider the following family of instances $\mathcal{H}$. For every $n \geq 1$, $\mathcal{H}$ contains the following instance. There are 2 facilities with $f_1 = 0$ and $f_2 = 1$. The number of clients is $n + 1$, all facilities and clients are co-located at the same point and both capacities are equal to n. By opening the first facility integrally and the second at a fraction of $\frac{1}{n}$ and assigning each client with a fraction of $\frac{n}{n+1}$ to the first and with a fraction of $\frac{1}{n+1}$ to the second, we get a solution that is feasible for (LP-classic) with a cost of $1/n$ (see Figure 2.1). On the other hand, each integral solution has a cost of 1 as both facilities must be open in order to satisfy the client demand.

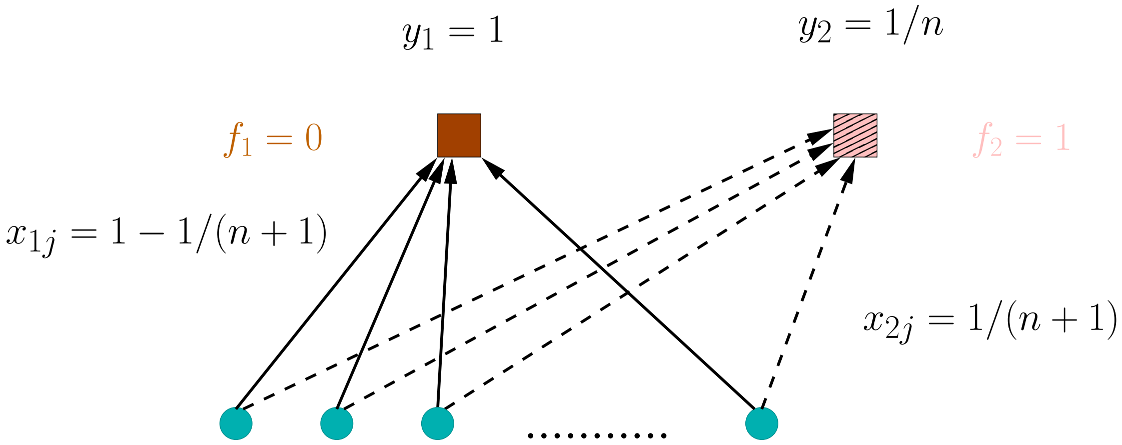

Figure 2.1: Representative instance from family $\mathcal{H}$ and the associated fractional solution to (LP-classic).

The instances and solutions that we will use in our proofs in sections 4.1, 4.2 and 4.3 generalize the former example. The parameterized family $\mathcal{I}$ of bad instances contains the following instance for every $n \geq 1$. Consider a set of n facilities which have 0 opening cost. We call that set *Cheap*. Moreover, consider a set of n facilities that have an opening cost of 1 each. Call that set *Costly*. The set of facilities F is $Cheap \cup Costly$. Let all the facilities have the same capacity $U = n^3$, and let there be a total of $nU + 1$ clients in the set C. All clients and facilities are at a distance of 0 from each other. Clearly all integer solutions to the instance have a cost of at least 1. The bad fractional solution s to (LP-classic) is the following: for each facility $i \in Cheap$, $y_i = 1$, and for each client j, set $x_{ij} = \frac{(1-a)}{n}$, $a = n^{-2}$. For each facility $i \in Costly$, $y_i = b$ where $b = 10/n^2$, and for each client j, set $x_{ij} = a/n$ (see Figure 2.2). The constructed solution incurs a cost of $\Theta(n^{-1})$.

For LBFL a simple metric space instance and bad solution witnessing unboundedness in the integrality gap of the classic relaxation is the following. We have two facilities 1 and 2 at a distance of 1. With each facility a set of $n - 1$ clients is co-located. The bound is $B = n$. An optimal solution to the classic relaxation opens both facilities with a fraction of $\frac{n-1}{n}$ and its client is assigned with a fraction of $\frac{n-1}{n}$ to each nearest facility and with a fraction $\frac{1}{n}$ to the other . Obviously this solution induces an integrality gap of $\Omega(n)$ as in an optimal solution only one facility can be opened and thus the faraway clients incure a connection cost of $\Theta(n)$.

A second well-known LP for facility location problems is the star relaxation. A *star* is a set consisting of some clients and one facility. Let $\mathcal{S}$ be a set of stars. For a star $s \in \mathcal{S}$, let x_s be an indicator variable denoting whether s is picked. The cost c_s of star s is equal to the opening cost of the corresponding facility plus the cost of connecting the star's clients to

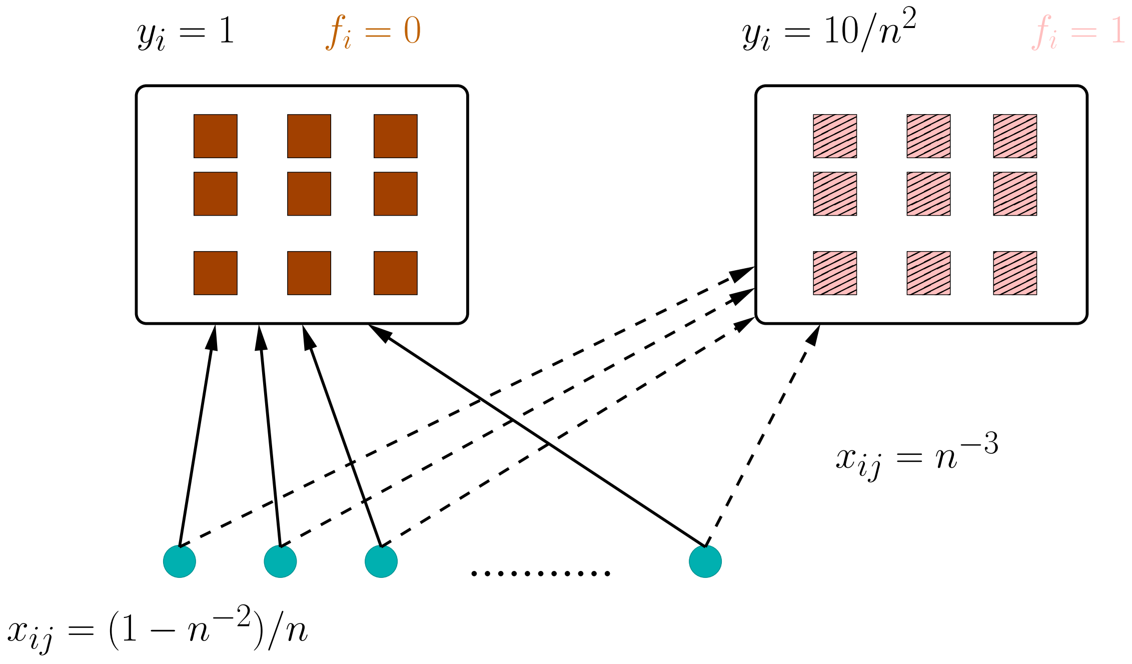

Figure 2.2: Representative instance from family $\mathcal{I}$ and the associated fractional solution s that will be used in the upcoming sections.

it.

$$\min \sum_s x_s c_s \qquad \text{(LP-star)}$$

$$\sum_{s \ni j} x_s = 1 \qquad \forall j \in C \qquad (2.19)$$

$$\sum_{s \ni i} x_s \leq 1 \qquad \forall i \in F \qquad (2.20)$$

$$x_s \geq 0 \qquad \text{for all stars } s \in \mathcal{S} \qquad (2.21)$$

Defining $\mathcal{S}$ as the set of all stars s where the total number of the clients in s is at most the capacity U (at least the bound B), we get corresponding relaxations for CFL (LBFL). As usual, when invoking the name *(LP-star)*, the appropriate version of facility location problem will be clear from the context.

It is well known , and quite easy to prove, that for both CFL and LBFL, (LP-classic) and (LP-star) are equivalent, therefore (LP-star) can be solved in polynomial time. For the sake of completion we include a simple proof.

Lemma 2.2.1. **(Folklore)** *Let $I(F, C)$ be an instance of* CFL *(*LBFL*). Then there is a map from each feasible solution (y, x) of LP-classic to a feasible solution x^s of LP-star that induces the same fractional assignment and openings and, thus, has the same cost and vice versa.*

Proof. It is easy to see that for any feasible solution to (LP-star) that satisfies $\sum_{s \ni i} x_s \leq 1$, for all $i \in F$, we get a solution to (LP-classic) of the same assignments, openings and cost: just set $y_i = \sum_{s \ni i} x_s$, and $x_{ij} = \sum_{s | \{i,j\} \in s} x_s$.

For the converse, given a feasible solution (y, x) to (LP-classic) and we produce a solution x_s to (LP-star) inducing the same fractional assignments and openings and having the same cost as follows: for a facility $i \in F$, consider a rectangle R_i of height y_i divided in $w_i = \lceil \sum_j x_{ij}/y_i \rceil$ columns. By picking a client after another we do the following: we start from the first column and foirst client j and we "fill" with the assignment of the first client from the top of the column to height x_{ij} from the top where we leave a mark. Then we pick the next client j', we go to last column filled and to the mark we left and we fill the column until an additional height of $x_{ij'}$ is filled from top to bottom and we leave one mark at the height we stop. If the column has not enough height left unfilled, then we fill all the column and we go to the next one and we fill with the remaining height from top to bottom and we leave a mark, and so on. At the end of this procedure we draw horizontal lines to all the marks we left. Now consider the all the rectangles defined by two consecutive horizontal lines, including the boundaries of R_i, which are divided in columns. For each such rectangle increase $x_{s'}$ by its height, where s is the set of clients that filled the columns of that rectangle – this is a valid set as we cannot have the same client at the same height of different columns by constraint $x_{ij} \leq y_i$ and more over s defines a proper star as it has at most (at least) U (B) members, the latter is ensured by the capacity (lower bound) constraint. The obtained solution for the LP-star is feasible and by construction induces the same assignments and openings and has the same cost.

□

Given two vectors $y = (y_1, \ldots, y_n)^T$ and $x = (x_1, \ldots, x_m)^T$ by (y, x) we denote the vector $(y_1, \ldots, y_n, x_1, \ldots, x_m)^T$. We slightly abuse notation and use $(1, x)$ to denote the vector $(1, x_1, \ldots, x_m)^T$.

3. A GENERALIZED CONFIGURATION LP APPROACH FOR FACILITY LOCATION

In this chapter we present the family of proper relaxations and characterize their strength. As stated in the introduction, the motivation behind proper relaxations is to capture configuration-like linear programs and study their behavior. More specifically we are interested in the behavior of the integrality gap in relation to a measure of complexity which indicates how global the generalized configurations are. We note that the results of this Chapter are not implied by the results regarding the Extended Formulations model since the size of proper relaxations is exponential.

3.1 Definition of proper relaxations

Given an instance of CFL (LBFL), a *class* is defined as a 0-1 vector $\eta = (y^\eta, x^\eta)$ which satisfies constraint (2.14) of (LP-classic). Note that the definition of a class η is quite general as the support of the vector (y^η, x^η) is not necessarily a subset of the support of some feasible integer solution for CFL (LBFL).

We associate with class η the *cost of the class*

$$c_\eta = \sum_{i \mid y_i^\eta = 1} f_i + \sum_{i,j \mid x_{ij}^\eta = 1} c_{ij}.$$

Let the *pairings of class* η be defined as

$$\text{Pairs}(\eta) = \{(i,j) \in F \times C \mid x_{ij}^\eta = 1\}.$$

We say that class η *opens* facility i, if $y_i^\eta = 1$. The set of facilities opened in η is denoted by $F(\eta)$.

Definition 3.1.1. (Constellation LPs) *Let $\mathcal{C}$ be a set of classes defined for an instance $I(F, C)$ of* CFL (LBFL)*. Let z_η be a variable associated with class $\eta \in \mathcal{C}$. The constellation LP with class set $\mathcal{C}$ is defined as*

$$\min \sum_{\eta \in \mathcal{C}} c_\eta z_\eta \qquad \text{(LP}(\mathcal{C})\text{)}$$

$$\textstyle\sum_{\eta \mid \exists i:(i,j) \in \text{Pairs}(\eta)} z_\eta = 1 \qquad \forall j \in C$$

$$\textstyle\sum_{\eta \mid i \in F(\eta)} z_\eta \le 1 \qquad \forall i \in F$$

$$z_\eta \ge 0 \qquad \forall \eta \in \mathcal{C}$$

We refer simply to a *constellation LP* when $\mathcal{C}$ is implied from the context. Given a solution z of LP($\mathcal{C}$), the induced solution to (LP-classic) is $(y, x) = \sum_{\eta \in \mathcal{C}} z_\eta (y^\eta, x^\eta)$.

We restrict our attention to constellation LPs that satisfy a symmetry property that is very natural for uniform capacities and unit demands.

Definition 3.1.2. **(P_1: Symmetry)** *We say that property P_1 holds for the constellation linear program LP($\mathcal{C}$) if for every class $\eta \in \mathcal{C}$, all classes resulting from a permutation that relabels the facilities and/or the clients of η are also in $\mathcal{C}$.*

Definition 3.1.3. **(Proper Relaxations)** *We call proper relaxation for* CFL *a constellation LP that is valid and satisfies property P_1.*

A simple example of a constellation LP is the well-known *(LP-star)* (see, e.g., [34]) where $\mathcal{C}$ corresponds to the set of all *stars*: a facility and a set of at most U clients assigned to it. Obviously (LP-star) is a proper relaxation, while (LP-classic) is equivalent to (LP-star). Therefore proper relaxations generalize the known natural relaxations for CFL. In order to characterize the strength of a proper LP we need the notion of complexity.

Definition 3.1.4. **(Complexity of proper relaxations)** *Given an instance $I(F, C)$ of* CFL *(*LBFL*), the complexity α of a proper relaxation $LP(\mathcal{C})$ for I is defined as the* $\max_{\eta\in\mathcal{C}}(|F(\eta)|/|F|)$.

The complexity of a proper LP represents the maximum fraction of the total number of feasibly openable facilities that is allowed in a single class. A complexity of nearly 1 means that there are classes that each take into consideration almost the whole instance at once. Low complexity means that all classes consider the assignments of a small fraction of the instance at a time. It is easy to create families of instances for which by increasing the complexity one obtains strictly better lower bounds. The following theorem characterizes the strength of proper relaxations with regard to their complexity.

Theorem 3.1.1. *The quality of solutions of proper relaxations is characterized by the following:*

(i) There is a family of instances $\mathcal{J}$ of uniform CFL *(*LBFL*), such that for every sufficiently large n, there is an instance in the family with $\Theta(n)$ facilities for which every proper relaxation with complexity $\alpha < 1$ has an integrality gap of $\Omega(n^2)$.*

(ii) There is a proper relaxation of complexity 1 for which the set of induced (y, x) vectors of its solutions is the integral CFL *(*LBFL*) polytope.*

3.1.1 Proof of (ii) of Theorem 3.1.1

We first prove the easy part, that there are proper relaxations for CFL (LBFL) with complexity 1 that express the integral polytope. For any instance, let $\mathcal{C}$ consist of the vectors of feasible 0-1 integer solutions. The resulting $LP(\mathcal{C})$ is clearly proper. Let z be any feasible solution of $LP(\mathcal{C})$ and let S be the support of the solution. For every $\eta \in S$, and for every client $j \in C$, there is an $i \in F$, such that $(i, j) \in \text{Pairs}(\eta)$. From this observation and the feasibility of z, for any $j \in C$ we have:

$$\sum\nolimits_{\eta|\exists i:(i,j)\in\text{Pairs}(\eta)} z_\eta = 1 \Rightarrow \sum_{\eta\in S} z_\eta = 1.$$

This implies that z is a convex combination of integral solutions. By the boundedness of the feasible region of $LP(\mathcal{C})$, the corresponding polytope is integral. Clearly not every LP with complexity 1 has an integrality gap of 1 since it might contain "strong" classes which correspond to feasible integer solutions together with "weak" ones, such as stars, whose conic combinations yield fractional solutions not in the integer CFL (LBFL) polytope.

3.2 Proof of (i) of Theorem 3.1.1 for CFL

In the next two subsections, we prove the first part of Theorem 3.1.1. In Subsection 3.2.1, we show that, for any proper relaxation $LP(\mathcal{C})$ of complexity strictly less than 1, there must be a specific set of classes contained in $\mathcal{C}$. In Subsection 3.2.2, we use these classes to construct a desired feasible fractional solution. By defining appropriately the cost structure of the instance, we establish an unbounded integrality gap.

3.2.1 Existence of a specific type of classes

Consider an instance I with n facilities, where n is sufficiently large to ensure that $\alpha n \leq n-1$. Let the capacity be $U = n^2$, and let the number of clients be $(n-1)U+1$. Notice that in every integer solution of the instance we must open all n facilities. The facility costs and the assignment costs will be defined later.

We assume, like before, that the facilities are numbered $1, 2, \ldots, n$. Consider an integral solution $s = (y^s, x^s)$ for I where all the facilities are opened, and furthermore facilities $1, \ldots, n-1$ are assigned U clients each and facility n is assigned one client. Since our proper relaxation is valid, there must be a solution s' in the space of feasible solutions of the proper relaxation whose induced (y, x) vector equals s. By Definition 3.1.1, it is easy to see that s' can only be obtained as a positive combination of classes η such that for every facility i and client j we have $x_{ij}^{\eta} \leq x_{ij}^{s}$. Recall that since the complexity of our relaxation is α, the classes in the support of any solution have at most $n-1$ facilities.

Now consider the support of s'. We will distinguish the classes η for which variable z_η is in the support of s' into 2 sets. The first set consists of the classes that assign exactly one client to facility n; call them *type A* classes. The second set consists of the classes that do not assign any client to facility n; call those *type B* classes. By the discussion above those sets form a partition of the classes in the support of s', and moreover they are both non-empty: this is by the fact that at most $n-1$ facilities are in any class, and by the fact that in s all n facilities are opened integrally. Notice also that no class of type B can open facility n even though the definition of a class does not exclude the possibility that a class opens a facility to which no clients are assigned.

We call *density* of a class η the ratio $d(\eta) = \frac{\sum_{i \neq n} \sum_{j \in C} x_{ij}^{\eta}}{\sum_{i \neq n} y_i^{\eta}}$. By the discussion above we have that $d(\eta) \leq U$ for all η in the support of s'. The following holds:

Lemma 3.2.1. *All classes in the support of s' have density U.*

Proof. The amount of demand that a class η contributes to the demand assigned to the set of the first $n-1$ facilities by s' is $d(\eta) z_\eta \sum_{i \neq n} y_i^{\eta}$. We have $\sum_\eta d(\eta) \sum_{i \neq n} y_i^{\eta} z_\eta = (n-1)U$. By the fact that in the induced (y, x) solution of s, for $i = 1, \ldots, n-1$, $y_i = \sum_\eta z_\eta y_i^{\eta} = 1$, we obtain that $\sum_\eta (z_\eta \sum_{i \neq n} y_i^{\eta}) = n-1$. Setting $m_\eta = \frac{z_\eta \sum_{i \neq n} y_i^{\eta}}{n-1}$ we have from the above $\sum_\eta m_\eta = 1$ and $\sum_\eta m_\eta d(\eta) = U$. The latter together with the fact that $d(\eta) \leq U$ we have that $d(\eta) = U$ for all classes η in the support of s'. $\square$

The following lemma is immediate from the above:

Lemma 3.2.2. *The support of s' contains a variable corresponding to a type B class, denoted η_0, that has density U.*

So far we have proved that in the class set of any proper relaxation for I, there is a class η_0 of type B with density $d(\eta_0) = U$. Let $|\sum_{i \in F} y_i^{\eta_0}| = t \le n-1$.

3.2.2 Construction of a bad solution

Consider the symmetric classes of η_0 for all permutations of the n facilities and for all permutations of the clients. Those classes are not necessarily in the support of s'. Take a quantity of measure ϵ and distribute it equally among all those classes. Since class η_0 has density U, all those symmetric classes assign on average U clients to each of their facilities. Due to symmetry, each facility is in a class $\epsilon\frac{t}{n}$ of the time and is assigned $\epsilon\frac{t}{n}U$ demand. Each client is assigned to each facility $\epsilon\frac{tU}{((n-1)U+1)n}$ of the time. We call that step of our construction *round* A.

Now consider the symmetric classes of η_0 for all permutations of the first $n-1$ facilities and for all permutations of the clients (those classes are well defined since $t \le n-1$). Again distribute a quantity of measure ϵ equally among all those classes. Similarly to the previous, each facility is in a class $\epsilon\frac{t}{n-1}$ of the time and is assigned $\epsilon\frac{t}{n-1}U$ demand. Each client is assigned to each facility $\epsilon\frac{tU}{((n-1)U+1)(n-1)}$ of the time. We call that step of our construction *round* B.

Spending $\phi = \frac{1}{nt}$ measure in round A and $\xi = \frac{(n-1)(1-1/n^2)}{t}$ measure in round B we construct a solution s_b whose corresponding (y, x) vector is the following (y^*, x^*): $y_i^* = 1$ for $i = 1, \dots, n-1$, $y_n^* = \frac{1}{n^2}$, and for every client j, $x_{nj}^* = \frac{U/n^2}{(n-1)U+1}$ and $x_{ij}^* = \frac{1-x_{nj}^*}{n-1}$ for $i = 1, \dots, n-1$. It is easy to see that s_b is a feasible solution for our proper relaxation.

Now simply set all distances to 0, and define the facility opening costs as $f_n = 1$ and $f_i = 0$ for $i \le n-1$. Simple calculations show that the integrality gap of the proper relaxation is $\Omega(n^2)$. The proof of Theorem 3.1.1 is now complete.

3.3 Proof of Theorem 3.1.1 for LBFL

Our proof includes the following steps. We define an instance I and consider any proper relaxation $LP(\mathcal{C})$ for I that has complexity $\alpha < 1$. Given α, we use the validity and symmetry properties to show the existence of a specific set of classes in $\mathcal{C}$. Then we use these classes to construct a desired feasible fractional solution, relying again on symmetry. In the last step we specify the distances between the clients and the facilities, so that the instance is metric and the constructed solution has an unbounded integrality gap.

3.3.1 Existence of a certain type of classes

Let us fix for the remainder of the section an instance I with $n+1$ facilities, where n is sufficiently large to ensure that $\alpha n \le n-2$. Let the bound $B = n^2$, and let the number of clients be n^3. Notice that there are enough clients to open n facilities, with exactly n^2 clients assigned to each one that is opened. The facility costs and the assignment costs will be defined later. Recall that the space of feasible solutions of a proper relaxation is independent of the costs.

We assume that the facilities are numbered $i = 1, 2, \ldots, n+1$. For a solution p we denote by $Clients_p(i)$ the set of clients that are assigned to facility i in solution p, and likewise for a class η we denote by $Clients_\eta(i)$ the set of clients that are assigned to facility i. Consider an integral solution s to the instance where facilities $1, \ldots, n$ are opened. Since our proper relaxation is valid, it must have a feasible solution s' to the proper relaxation whose projection to (y, x) gives the characteristic vector of s, let that vector be (y^s, x^s). We prove the existence of a class η_0, with some desirable properties, in the support of s'.

By Definition 3.1.1, s' can only be obtained as a positive combination of classes η such that for $(i,j) \in Pairs(\eta) \Rightarrow x^s_{i,j} = 1$. Otherwise, if the variables of a class η with $Pairs(\eta) \setminus \{(i,j) \mid x^s_{i,j} = 1\} \neq \varnothing$ have non-zero value, then in s' there will be some client assigned to some facility with a positive fraction, while the projection of s', namely s, does not include the particular assignment. Moreover, since exactly B clients are assigned to each facility in s, for every facility i that is contained in such a class η, $(i,j) \in Pairs(\eta) \Leftrightarrow x^s_{i,j} = 1$. To see why this is true, since in s we have $y_i = 1$, for all $i \leq n$, it follows that for every facility $i \leq n$, $\sum_{\eta \mid \exists (i,j) \in Pairs_\eta} z_\eta = 1$. But then we have that $|Clients_s(i)| = B = \sum_{\eta \mid \exists (i,j) \in Pairs(\eta)} z_\eta |(i,j) \mid Pairs(\eta)|$. We have already established that $z_\eta > 0 \implies |(i,j) \mid Pairs(\eta)| \leq B$. Then B is a convex combination of quantities less than or equal to B, so for all such classes η we have $(i,j) \mid Pairs(\eta)| = B$ for any facility i.

Therefore in the class set of any proper relaxation for I, there is a class η_0 that assigns exactly B clients to each of the facilities in $F(\eta_0)$. By the value of α, $|F(\eta_0)| = n - c \leq n - 2$ for some integer c.

3.3.2 Construction of a bad solution

In the present section we will use the class η_0 along with the symmetric classes to construct a solution to the proper LP with the following property: there are some q facilities that are almost integrally opened while the number of distinct clients assigned to them will be less than Bq.

Recall that by property P_1 every class that is isomorphic to η_0 is also a class of our proper relaxation. This means that every set of $n - c$ facilities and every set of $B(n - c)$ clients assigned to those facilities so that each facility is assigned exactly B clients, defines a class, called *admissible*, that belongs to the set of classes of a proper relaxation for the instance I.

Let us turn again to the solution s to provide some more definitions. For every facility i, $i = 1, \ldots, n-1$, we choose arbitrarily a client j' assigned to it by s. For each such facility i we denote by $Exclusive(i)$ the set of clients $Clients_s(i) - \{j'\}$, i.e., the set of clients assigned to i by s after we discard j' (we will also call them the *exclusive clients of* i). For facilities n, $n+1$ the sets $Exclusive(n)$, $Exclusive(n+1)$ are identical and defined to be equal to the union of $Clients_s(n)$ with all the discarded clients from the other facilities. In the fractional solution that we will construct below, the clients in $Exclusive(i)$ will be almost integrally assigned to i for $i = 1, \ldots, n-1$.

We are ready to describe the construction of the fractional solution. We will use a subset S of admissible classes that do not contain both n and $n+1$. S contains all such classes η that assign to each facility $i \leq n-1$ in the class the set of clients $Exclusive(i)$ plus one more client selected from the sets $Exclusive(i')$ for those facilities $i' \leq n-1$ that do not belong to η (there are at least $c-1$ of them). As for facility n (resp. $n+1$), if it is contained in η, then it is assigned some set of B clients out of the total $B+n-1$ in $Exclusive(n)$ (resp. $Exclusive(n+1)$). All classes not in S will get a value of zero in our solution. We will distinguish the classes in S into two types: the classes of *type* A that contain facility n or $n+1$ but not both, and classes of *type* B that contain neither n nor $n+1$.

We consider first classes of type A. We give to each such class a very small quantity of measure ϵ. Let ϕ be the total amount of measure used. We call this step $Round_A$. The following lemma shows that after $Round_A$, the partial fractional solution induced by the classes has a convenient and symmetric structure:

Lemma 3.3.1. *After $Round_A$, each client $j \in Exclusive(i)$, $i \leq n-1$, is assigned to i with a fraction of $\frac{n-c-1}{n-1}\phi$ and is assigned to each other facility i', $i' \neq i$, $i' \leq n-1$, with a fraction of $\frac{n-c-1}{(n-1)(n-2)(n^2-1)}\phi$. Each client $j \in Exclusive(n)$ ($= Exclusive(n+1)$) is assigned to n and to $n+1$ with a fraction of $\frac{n^2}{2(n^2+n-1)}\phi$.*

Proof. Consider a facility $i, i \leq n-1$. Since exactly one of facilities $n, n+1$ is present in all the classes of type A and each class contains $n-c$ facilities, i is present in the classes of $Round_A$ $\frac{n-c-1}{n-1}$ of the time due to symmetry of the classes. Each time i is present in a class η that class η assigns all $j \in Exclusive(i)$ to i. So client j is assigned to i with a fraction of $\frac{n-c-1}{n-1}\phi$. When i is not present in class η, which happens $\frac{c}{n-1}$ of the time, then its exclusive clients along with the exclusive clients of all the other $c-1$ facilities other than $n, n+1$ that are also not present in η are used to help the $n-c-1$ facilities $i \leq n-1$, reach the bound B of clients (recall that the number of exclusive clients of each such facility is equal to $B-1$). Each time this happens, the $n-c-1$ facilities $i \leq n-1$ in η need $n-c-1$ additional clients, while the exclusive clients of the c facilities other than $n, n+1$ that are not present in η are $c(n^2-1)$ in total. Due to symmetry once again, a specific client $j \in Exclusive(i)$ is assigned to one of those $n-c-1$ facilities $\frac{n-c-1}{c(n^2-1)}$ of the time of those cases. So in total this happens $\frac{c}{n-1} \times \frac{n-c-1}{c(n^2-1)} = \frac{n-c-1}{(n-1)(n^2-1)}$ of the time, so it follows that client j is assigned to a specific facility i', $i' \neq i$, $i' \leq n-1$, $\frac{n-c-1}{(n-1)(n-2)(n^2-1)}$ of the time. The fraction with which j is assigned to i' after $Round_A$ is $\frac{n-c-1}{(n-1)(n-2)(n^2-1)}\phi$.

For the proof of the second part of the lemma, consider facilities $n, n+1$. Each one of those is present in the classes of type A an equal fraction $1/2$ of the time. The only clients that are assigned to them are their exclusive clients. Each class η assigns exactly $B = n^2$ out of those n^2+n-1 clients. So, due to symmetry, each client $j \in Exclusive(n)$ is present in η $\frac{n^2}{n^2+n-1}$ of the time, so j is assigned to n and $n+1$ with a fraction of $\frac{n^2}{2(n^2+n-1)}\phi$ to each. □

Note that after $Round_A$ each facility $i, i \leq n-1$, has a total amount $\frac{(n-c-1)B}{(n-1)}\phi$ of clients (since it is present in a class $\frac{(n-c-1)}{(n-1)}$ of the time and when this happens it is given B clients). Similarly, facilities $n, n+1$ after $Round_A$ have a total amount $B\phi/2$ each.

Now we can explain the underlying intuition for distinguishing between the two types of classes. The feasible fractional solution (y^*, x^*) we intend to construct is the following: for each facility $i \leq n-1$, its exclusive clients are assigned to it with a fraction of $\frac{n^2-1}{n^2}$ each, while they are assigned with a fraction of $\frac{1}{(n^2)(n-2)}$ to each other facility $i' \leq n-1$. As for facilities $n, n+1$, all of their exclusive clients are assigned with a fraction of $1/2$ to each. If we project the solution to (y, x), the y variables will be forced to take the values $y_i^* = \frac{n^2-1}{n^2}$, for $i \leq n-1$, and $y_n^* = y_{n+1}^* = \frac{n^2+n-1}{2n^2}$. Observe as we give some amount of measure to $Round_A$, the variables concerning the assignments to facilities $n, n+1$ tend to their intended values in the solution we want to construct "faster" than the variables concerning the assignments to the other facilities. This is because, by Lemma 3.3.1 after $Round_A$ each exclusive client of $n, n+1$ is assigned to each of them with a fraction of $\frac{n^2}{2(n^2+n-1)}\phi$ which is $\frac{n^2}{n^2+n-1}\phi$ of the intended value. At the same time, every exclusive client of each other facility is assigned to it with a fraction of $\frac{n-c-1}{n-1}\phi$ which is $\frac{\frac{n-c-1}{n-1}\phi}{\frac{n^2-1}{n^2}}$ of the intended value. For sufficiently large instance I, as n tends to infinity, the assignments to n and $n+1$ will reach their intended values while there will be some fraction of every other client left to be assigned. Subsequently we have to use classes of type B, to achieve the opposite effect: the variables concerning the assignments of the first $n-1$ facilities should tend to their intended values "faster" than those of n and $n+1$ (since n and $n+1$ are not present in any of the classes of type B, the corresponding speed will actually be zero).

We proceed with giving the details of the usage of type B classes. As before, we give to each such class a very small quantity of measure ϵ. Let ξ be the total amount of measure used. We call this step $Round_B$.

Lemma 3.3.2. *After $Round_B$, each client $j \in Exclusive(i)$, $i \leq n-1$, is assigned to i with a fraction of $\frac{n-c}{n-1}\xi$ and is assigned to each other facility i', $i' \neq i$, $i' \leq n-1$, with a fraction of $\frac{n-c}{(n-1)(n-2)(n^2-1)}\xi$.*

Proof. The proof follows closely that of Lemma 3.3.1. A facility $i, i \leq n-1$, is present in a class of type B $\frac{n-c}{n-1}$ of the time (since $c \geq 2$ this fraction is less than 1). Each such time, every client $j \in Exclusive(i)$ is assigned to it (again this is due to the definition of classes of type B). So after $Round_B$, j is assigned to i with a fraction of $\frac{n-c}{n-1}\xi$. Also, when i is present in a class, it is assigned exactly one client which is exclusive to a facility not in the class. Since in total there are $(n-2)(B-1)$ such candidate clients, and by symmetry, after round B each one of them is picked an equal fraction of the time to be assigned to i, we have that each client j is assigned to a facility for which j is not exclusive with a fraction $\frac{n-c}{(n-1)(n-2)(n^2-1)}\xi$. □

To construct the aforementioned fractional solution (y^*, x^*), set $\phi = \frac{n^2+n-1}{n^2}$ and $\xi = (\frac{n^2-1}{n^2} - \frac{n-c-1}{n-1}\phi)\frac{n-1}{n-c}$, and add the fractional assignments of the two rounds.

It is easy to check that the facility and assignment variables of facilities $n, n+1$ take the value they have in (y^*, x^*). Same is true for the facility variables for $i \leq n-1$ and the assignment variables of the clients to the facilities they are exclusive. To see that the same goes for the non-exclusive assignments, observe that since every class assign exactly B clients to its facilities we have that $\sum_j x_{ij} = By_i$. So each $i \leq n-1$ takes exactly $1 - 1/n^2$ demand from non-exclusive clients which are $(n-2)(B-1)$ in total. Thus, by symmetry of the construction, each one them is assigned to i with a fraction of $\frac{B-1}{n^2(n-2)(B-1)} = \frac{1}{n^2(n-2)}$.

3.3.3 Proof of unbounded integrality gap of the constructed solution

In the present subsection, we manipulate the costs of instance I, which we left undefined, so as to create a large integrality gap while ensuring that the distances form a metric.

Set each facility opening cost to zero. As for the connection costs (distances) consider the $(n-2)$-dimensional Euclidean space $\mathbb{R}^{n-2}$. Put every facility i, $i \leq n-1$, together with its exclusive clients on a distinct vertex of an $(n-2)$-dimensional regular simplex with edge length D. Put facilities $n, n+1$ together with their exclusive clients to a point far away from the simplex, so that the minimum distance from a vertex is $D' >> D$. Setting $D' = \Omega(nD)$ is enough.

Since the distance between a facility and one of its exclusive clients is 0, the cost of the fractional solution we constructed is $O(nD)$. This cost is due to the assignments of exclusive clients of facility i, $i \leq n-1$, to facilities i' with $i' \neq i$, $i' \leq n-1$. As for the cost of an arbitrary integral solution, observe that since the n^2+n-1 exclusive clients of $n, n+1$ are very far from the rest of the facilities, using n of them to satisfy some demand of those facilities and help to open all of them, incurs a cost of $\Omega(nD') = \Omega(n^2 D)$. On the other hand, if we do not open all of the $n-1$ facilities on the vertices of the simplex (since they have in total $(n-1)(B-1)$ exclusive clients which is not enough to open all of them), there must be at least one such facility not opened in the solution, thus its $B-1 = \Theta(n^2)$ exclusive clients must be assigned elsewhere, incurring a cost of $\Omega(n^2 D)$.

This concludes the proof of Theorem 3.1.1.

4. HIERARCHY-BASED HARDNESS OF APPROXIMATING CAPACITATED VERSIONS OF FACILITY LOCATION

In this chapter we expose the weakness of linear programming hierarchies to provide satisfactory approximations to Facility Location problems with capacity restrictions. Section 4.1 contains the results regarding the Lovász-Schrijver hierarchy which were extended in Section 4.2 to the Sherali-Adams hierarchy. Then in Section 4.3, by using the techniques we devised in the previous sections, we show unboundedness in the integrality gap of the classic relaxation reinforced with the submodular inequalities.

4.1 LS gaps for CFL

We start with Section 4.1.1 which contains contains the definition of the Lovász-Schrijver hierarchy, then we give in Section 4.1.2 a reformulation of the Lovász-Schrijver procedure that we use in our proofs. We proceed in Section 4.1.3 with the proof of unboundedness of the gap of the LS relaxation for an asymptotically tight number of rounds.

4.1.1 The LS Hierarchy

The Lovász-Schrijver hierarchy of relaxations was defined in [49]. For a comprehensive presentation and various reformulations see [63]. In this section we give the definitions necessary for our proof.

In [49] an operator N was defined through the use of which we can refine a convex set $P \subseteq [0,1]^d$. Starting with the polytope of interest $P \subseteq [0,1]^d$ we define $\mathrm{cone}(P) = \{y = (\lambda, \lambda z_1, ..., \lambda z_d)^T \mid \lambda \geq 0, (z_1, ..., z_d)^T \in P\}$. For such a conified polytope in $\mathbb{R}^{d+1}_{\geq 0}$ we follow the convention that the indices of the coordinates range from 0 to d. Observe that the projection of $\mathrm{cone}(P)$ onto $x_0 = 1$ equals P. One can define the N operator through lift-and-project operations over the linear system defining P (see, e.g., [49, 10]). For our purposes it is sufficient to use the following formalism. For a cone $K \subseteq [0,1]^{d+1}$, and $t \geq 1$, we define the notation $N^0(K) = K$ and $N^t(K) = N(N^{t-1}(K))$.

Definition 4.1.1 ([49]). *Let K be a cone in $\mathbb{R}^{d+1}$. For $m \geq 1$, $N^m(K)$ is the set of vectors $z \in \mathbb{R}^{d+1}$ for which there is a $(d+1) \times (d+1)$ symmetric matrix Y satisfying*

1. $Ye_0 = diag(Y) = z$.
2. *For $1 \leq i \leq d$, both Ye_i and $Y(e_0 - e_i)$ are in $N^{m-1}(K)$.*

Y is called the protection matrix of z.

Lovász and Schrijver [49] showed that for any polytope $P \subseteq [0,1]^d$, the projection of $N^d(\mathrm{cone}(P))$ onto $x_0 = 1$ equals the integer hull of P.

The operator N_+ is defined as above with the additional restriction that Y has to be positive semidefinite. The LS *procedure* consists of repeated applications of the operator

N starting from $K = \text{cone}(P)$, where P is the polytope of the original relaxation. Every such application is called a *round*. If there is a protection matrix Y for z, we say that z *survives* one round of LS . Given a cost vector c, the LS procedure *on* P *produces after* m *rounds*, $m \geq 1$, the linear relaxation

$$\min\{c^T x \mid (1, x) \in N^m(K)\}.$$

Similarly, the LS_+ procedure on P *produces after* m *rounds*, $m \geq 1$, the linear relaxation

$$\min\{c^T x \mid (1, x) \in N_+^m(K)\}.$$

The discussion so far has treated P as the relaxation of an integer 0-1 program with d variables. For a mixed integer program with $d + p$ variables where one optimizes over $\text{conv}(P \cap (\{0,1\}^d \times \mathbb{R}^p))$, to derive an exact formulation it is sufficient to lift in each round only the d integer variables [10, 26]. To reflect this we define the operators M and M_+ as follows.

Definition 4.1.2 ([49, 10])**.** *Let K be a cone in $\mathbb{R}^{d+p+1}$ and let $M^0(K) = K$. For $m \geq 1$, $M^m(K)$ is the set of vectors $z = (y, x) \in \mathbb{R}^{d+p+1}$, $y \in \mathbb{R}^{d+1}$, $x \in \mathbb{R}^p$, for which there is a $(d+1) \times (d+1)$ symmetric matrix Y satisfying*

1. $Ye_0 = diag(Y) = y$.
2. *For $1 \leq i \leq d$, there exists $x^{(i)} \in \mathbb{R}^p$ such that both $(Ye_i, x^{(i)})$ and $(Y(e_0 - e_i), x - x^{(i)})$ are in $M^{m-1}(K)$.*

When in the definition above Y is constrained to be positive semidefinite, we obtain the operator M_+. Clearly, for any cone K, $N(K) \subseteq M(K)$ and $N_+(K) \subseteq M_+(K)$. The *mixed* LS_+ *procedure* applies repeatedly the M_+ operator.

4.1.2 Reformulation of the LS procedure

In this section we highlight some simple properties of protection matrices and explain how they will be used in our proofs.

Since we are interested in the projection of the cones resulting from the successive application of the operator N on the hyperplane $z_0 = 1$ which contains our original polytope P, we restate the conditions of survival of z as the following lemma which is immediate from Definition 4.1.1.

Lemma 4.1.1 ([8])**.** *Let $P \subseteq [0,1]^d$ be a polytope and denote by K the set $\text{cone}(P)$ in $\mathbb{R}^{d+1}_{\geq 0}$. A vector $z \in \mathbb{R}^{d+1}$ whith $z_0 = 1$ is in $N^m(K)$, $m \geq 1$, if and only if there is a $(d+1) \times (d+1)$ symmetric matrix Y satisfying*

1. $Ye_0 = diag(Y) = z$.
2. *For $1 \leq i \leq d$: If $z_i = 0$ then $Ye_i = \mathbf{0}$; If $z_i = 1$ then $Ye_i = z$; Otherwise, Ye_i/z_i, $Y(e_0 - e_i)/(1 - z_i)$ both lie in the projection of $N^{m-1}(K)$ onto the hyperplane $z_0 = 1$.*

Let Y_i denote the vector $Y^T e_i$, i.e., the ith row of Y. Lemma 4.1.1 makes it convenient to work with individual vector solutions that can be combined as rows to build the protection matrix. Given a protection matrix Y of z, we define a set of at most $2d$ *witnesses* of vector z. For each variable z_i, $1 \leq i \leq d$, there are at most 2 such witnesses: the one that equals

Y_i/z_i (if $z_i \neq 0$), which we call *type 1 witness of z corresponding to variable z_i*, and the vector $\frac{Y_0-Y_i}{1-z_i}$ (if $z_i \neq 1$), which we call *type 2 witness of z corresponding to variable z_i*.

Before proceeding, we give a brief preview of the proof strategy for our main LS result, Theorem 4.1.1. To prove the existence of a protection matrix Y for a vector z, we will use the following steps. We will define a set $S(z)$ of vectors, which consists of a “candidate” type 1 and a “candidate” type 2 witness for every non-integer variable and one of the appropriate type for each integer variable, with respect to some as of yet unknown protection matrix. We will ensure that the vectors in $S(z)$ meet a set of conditions (described in Lemma 4.1.2 below) which are sufficient so that $S(z)$ corresponds to a set of *actual* witnesses from a correct protection matrix Y of z. To prove the survival of a vector for many rounds we just embed the strategy above in an inductive argument. The following fact is immediate from Lemma 4.1.1: if, for all i the vectors Y_i/z_i and $\frac{Y_0-Y_i}{1-z_i}$ (that are defined) witnessing the survival of our initial vector z, survive themselves k rounds of LS, then z survives actually $k+1$ rounds of LS.

We are now ready to start presenting a reformulation of Lemma 4.1.1. For the validity of the following observation recall that if $z_i = 0$, and hence the type 1 witness corresponding to i is undefined, $Y_i = \mathbf{0}$.

Observation 4.1.1. *The fact that Y's main diagonal is equal to the vector Y_0 is equivalent to the following Condition (i): the variable z'_i of the type 1 witness z' corresponding to variable $z_i \neq 0$ is equal to 1, and equal to 0 if $z_i = 0$.*

The rows of Y that correspond to zero variables in z are filled with zeros. Moreover if $z_i = 1$, $Y_i = z$. To account for these requirements it is not enough that the integer values in Y_0 appear on the main diagonal. The following observation states that they are replicated across all witnesses.

Observation 4.1.2. *Let z' be a witness of z. [Condition (ii):] if for some i, $z_i \in \{0, 1\}$, then $z'_i = z_i$.*

To enforce symmetry for a row $Y_i = \mathbf{0}$ that corresponds to a variable $z_i = 0$, it must be the case that the ith column is set to zero as well. This is ensured by Observation 4.1.2 for all entries Y_{ki} of the column for which $z_k \neq 0$. (The remaining entries of the column belong to zero rows and are equal to zero anyway). For the remaining rows, it will be convenient to express the type 1 witness z' of z corresponding to some variable z_i, by defining the factors by which the variables of z' differ from the corresponding variables of z. Then the symmetry condition of Y is satisfied by ensuring that the condition of the following proposition on those factors holds.

Proposition 4.1.1. *Let indices q, t take values in $\{1, \ldots, d\}$. The symmetry condition of the protection matrix of z holds if and only if Condition (ii) from Observation 4.1.2 holds together with the following Condition (iii): for each type 1 witness z' of z corresponding to variable z_q, for which $z'_t = z_t f$, $z_t \neq 0$, then, for the type 1 witness z'' of z corresponding to variable z_t, we have $z''_q = z_q f$.*

The above proposition is obtained by just observing that $Y_{qt} = z_q z'_t = z_q z_t f = z_t z''_q = Y_{tq}$ for non-integer z_q, z_t and by the fact that $Y_i = z$ for integer z_i. Note that when we construct a type 1 witness z' corresponding to z_t, the type 2 witness z'' corresponding to z_t is automatically defined. We say that z'' is the *twin* of z' .

Proposition 4.1.2. *Let indices q, t take values in $\{1, \dots, d\}$. If Y is a protection matrix of z, the following Condition (iv) must hold: if $z'_q = z_q(1+\epsilon)$ in the type 1 witness z' corresponding to z_t, $z_t \neq 1$, then $z''_q = z_q(1 - \frac{z_t \epsilon}{1-z_t})$ where z'' is the type 2 twin of z'.*

Based on Observation 4.1.1 and Propositions 4.1.1, 4.1.2, the following lemma is now a straightforward reformulation of Lemma 4.1.1.

Lemma 4.1.2. *Let $z \in \mathbb{R}^d$ and let $S(z)$ be a collection of vectors that lie in the projection of $N^{m-1}(K)$ onto the hyperplane $z_o = 1$. The vectors in $S(z)$ satisfy the Conditions (i)-(iv) stated in Observation 4.1.1 and Propositions 4.1.1 and 4.1.2 if and only if there is a protection matrix Y for z such that for each $z_i \neq 0$ there is exactly one $z' \in S(z)$ such that $z' = Y_i/z_i$ and for each $z_i \neq 1$ there is exactly one $z'' \in S(z)$ such that $z'' = \frac{Y_0 - Y_i}{1-z_i}$.*

Lemma 4.1.2 shows how to produce a protection matrix that ensures the survival of a vector z for a single round of LS . To show survival for several rounds, we implement the inductive argument mentioned above by defining inductively an appropriate tree structure. In particular, we define the *evolution tree T_z of z*. Every node in T_z is *associated* with a vector. The tree is defined recursively: vector z is associated with the root of the tree, and if v is a node of T_z, associated with vector $z(v)$, then a set of vectors witnessing the survival of $z(v)$ is associated in one-to-one manner with the children of v. If there is no such set of witnesses, the fractional solution $z(v)$ does not survive one round of LS. The length of the shortest path from the root of an evolution tree T_z to a childless node is a lower bound on the number of rounds that our initial vector z survives.

Given a root vector, we will show that as long as we have walked down the evolution tree at depth k, where k is the target number of rounds, then the protection matrices of the root and all its descendants are well-defined. The inductive step shows how to define all children of a node v and therefore increase the depth of the tree by one. From now on we refer interchangeably to a node and its associated solution vector. Accordingly, if v' is a child of node v, z' (z) is associated with v' (resp. v) and z' is a type 1 (2) witness of z corresponding to variable z_i, we will refer to node v' as a *type 1 (resp. 2) child of node-solution z corresponding to variable z_i.*

Finally, the following fact will be useful for the feasibility proof.

Lemma 4.1.3. *Given a solution z in the evolution tree that satisfies an equality constraint $\sum_i a_i z_i = b$, and given a child of z that is a type 1 solution z' corresponding to some z_t that satisfies $\sum_i a_i z'_i = b$, then the twin type 2 solution z'' of z' also satisfies $\sum_i a_i z''_i = b$.*

Proof. Let $z'_i = z_i(1 + \epsilon_i)$. From $\sum_i a_i z_i = b$ and $\sum_i a_i z'_i = b$ we get $\sum_i a_i z_i \epsilon_i = 0$. Then by Proposition 4.1.2,

$$\sum_i a_i z''_i = \sum_i a_i z_i (1 - \frac{z_t \epsilon_i}{1 - z_t}) = \sum_i a_i z_i - \frac{z_t}{1 - z_t} \sum_i a_i z_i \epsilon_i = b.$$

□

4.1.3 LS gaps for approximate CFL

In this section we show that the integrality gap on a suitable instance of CFL remains poor even after a large number of iterations of the LS procedure. More precisely we prove the following result.

Theorem 4.1.1. *There is a family of instances $\mathcal{I}$ of uniform* CFL*, such that for every sufficiently large n, there is an instance in the family with $2n$ facilities and $n^4 + 1$ clients for which the relaxation produced by the* LS *procedure in $\Omega(n)$ rounds has integrality gap $\Omega(n)$.*

For the families of instances we consider (unit demands and integer capacities) our bound is the best possible asymptotically. The following is well-known and is included for the sake of completeness.

Theorem 4.1.2. *[10] For every instance of* CFL *with unit demands and integer capacities, the* LS *and mixed* LS_+ *procedures obtain each the integer hull after a number of rounds that is equal to the number of the facilities.*

Proof. In the case of unit demands and integer capacities the fully integer and the mixed CFL polytopes are the same. The LS procedure applies to mixed integer programs as well where only the integer variables are lifted and the mixed integer polytope is obtained after a number of rounds that is equal to the number of the integer variables [10]. For CFL, the lifting of just the facility variables produces a weaker relaxation than the lifting of both the facility and the assignment variables that our LS procedure implements. Thus our LS procedure obtains the integral polytope after at most a number of rounds that is equal to the number of the facilities. □

It is well-known that the LS procedure may take a large number of rounds to produce some simple valid inequalities. In the case of CFL, Theorem 4.1.1 implies that $\Theta(n)$ rounds are required to obtain, e.g., the inequality

$$\sum_{i \in F} y_i \geq \lceil |C|/U \rceil \tag{4.22}$$

which is facet-inducing [45, p. 283] for our instance. However, this inequality is not critical for our proof. It is easy to modify the input by adding one facility f' with zero opening cost and one client c' which are co-located at a large distance from the rest of the instance. There is a bad fractional solution in which f' is fully opened and client c' is integrally assigned to f'; this solution satisfies the inequality above. Theorem 4.1.1 continues to hold for the augmented instance: variables with integer values are handled trivially in our construction, cf. Observation 4.1.2 and the proof of Theorem 4.1.3 below.

To prove Theorem 4.1.1, we will consider an instance with $2n$ facilities from the family $\mathcal{I}$ and the associated bad solution $s = (y, x)$ defined in Section 2.2. Before proving a lower bound on the number of rounds s survives, we establish an easy upper bound.

Observation 4.1.3. *Solution s survives less than n rounds of the LS procedure.*

For the proof of Observation 4.1.3 the following more general lemma suffices:

Lemma 4.1.4. *Let $K \subseteq [0, 1]^d$ be a polytope. Let S be a set of indices of variables of vector $g \in [0, 1]^d$, s.t. $\sum_{i \in S} g_i < 1$. If $(1, g)$ belongs to $N^{|S|}(\mathrm{cone}(K))$, then there is a vector g^* such that $(1, g^*)$ belongs to $\mathrm{cone}(K)$ and for all $i \in S$ $g^*_i = 0$.*

Proof. Every vector in the feasible set obtained after r rounds of LS has the property that for every tuple of r coordinates it can be expressed as a convex combination of vectors that are feasible for the starting relaxation K and have integer values on those coordinates [49]. Thus, at least one such vector g^* in the corresponding convex combination that yields g must contain 0 values at all positions indexed by $i \in S$. □

To prove Observation 4.1.3, assume that s survives n rounds. Apply Lemma 4.1.4 with K being the feasible set of (LP-classic). There must be a vector in K in which all the y_i variables, for $i \in Costly$, are set to 0. This cannot be a feasible solution to (LP-classic) since at least one costly facility has to be opened by a nonzero fraction, a contradiction.

We are ready to state the main theorem of this section which implies that the solution s survives $n/10$ rounds of LS. We do not make any attempt to optimize the constant. The intuition is that at every level of the induction the new witness solutions cannot differ drastically from their parent node. We identify a set of invariants that express this controlled evolution of the values. Recall that $b = 10/n^2$ is the value of the costly facility variables in s.

Theorem 4.1.3. *Let n be sufficiently large. We can construct an evolution tree T_s with root s such that any node u of T_s at depth $k \leq \frac{n}{10}$ is associated with a feasible solution (y, x) that satisfies the following invariants:*

1 *For variable* $y_i \notin \{0, 1\}$, $i \in Costly$, $b - 2k\frac{a}{n} \leq y_i \leq b + 2k\frac{a}{n}$.

2 (a) *For variable* $x_{ij} \notin \{0, 1\}$, $i \in Cheap$, $\frac{1-a}{n} - 2k\frac{a}{n^2}b^{-1} \leq x_{ij} \leq \frac{1-a}{n} + 2k\frac{1-a}{n^2}$.

  (b) *For variable* $x_{ij} \neq 0$, $i \in Costly$, *and* $y_i \notin \{0, 1\}$, $\frac{a}{n} \leq x_{ij} \leq \frac{a}{n} + 2k\frac{a(1-a)}{n^2}$.

  (c) *For variable* $x_{ij} \notin \{0, 1\}$, $i \in Costly$, *and* $y_i = 1$, $\frac{a}{n} \leq x_{ij} \leq (\frac{a}{n} + 2k\frac{a(1-a)}{n^2})b^{-1}(1+\frac{1}{10})$.

3 *For* $i \in Cheap$, $\sum_j x_{ij} \leq (nU + 1)\frac{1-a}{n} + 2k(nU + 1)\frac{a}{n^2}$.

4 *For* $i \in Costly$,

  (a) *if* $y_i \neq 1$, $\sum_j x_{ij} \leq (nU + 1)\frac{a}{n} + k$.

  (b) *if* $y_i = 1$, $\sum_j x_{ij} \leq ((nU + 1)\frac{a}{n} + k)(1 + \frac{1}{10})b^{-1}$.

Theorem 4.1.3 implies that solution s survives $\Omega(n)$ rounds. It remains to prove Theorem 4.1.3 and then the proof of Theorem 4.1.1 will be complete.

4.1.3.1 Proof of Theorem 4.1.3

The proof is by induction on the depth of node u. More specifically, by assuming that the invariants hold for an arbitrary node v at depth less than $n/10$, we show how to construct all the children nodes of v so that they are well-defined and the invariants are met.

In the proof, whenever we give the construction of a type 1 or type 2 child of v corresponding to some variable z_i, we refer to z_i as the *touched variable* – we also say that z_i is *touched* as type 1 or type 2 in the current step. We will consider cases according to which variable is touched and whether it is touched as type 1 or as type 2. When we touch a variable $z_i \notin \{0, 1\}$ as type 1, the z'_i variable of the corresponding type 1 witness z' always takes the value 1 so by Observation 4.1.1 we satisfy the condition that the diagonal of the underlying protection matrix is equal to the 0th row. Note that we will not give the construction for the case in which y_i, $i \in Cheap$, is touched, since y_i is always 1 and the construction is trivial in those cases. The same applies to the cases of all variables that have integral values in the node-solution v of the inductive hypothesis, as we simply enforce Observation 4.1.2.

Another feature of our construction, which is actually a necessary property towards LS feasibility, is the following: when a fractional variable $x_{i'j}$ is touched as type 1, it is set to 1, and for all $i \neq i'$, x_{ij} becomes 0. If $x_{i'j}$ is touched as type 2, it is set to 0 and its previous value must be distributed among the other assignments of client j. Thus for every j, either there is some i' such that $x_{i'j} = 1$ and for all other $i \neq i'$, $x_{ij} = 0$ (e.g., when cases $1b$, $1c$ below have happened for an ancestor of v), or there are at most k facilities to which the assignment of j is 0, where k is the depth of the tree (if there are type 2 nodes, through cases $2a$, $2b$, $2c$, along the path of the tree that leads to v). In fact, as far as assignments to cheap facilities are concerned, the upper bound of k holds cumulatively across all clients, since no more than k assignment variables can be touched as Type 2 along a path of length k. Specifically, let C' be the set of clients j for which, for all $i \in F$, $x_{ij} < 1$. We will use the fact that $|\{x_{ij}, i \in Cheap, j \in C' \mid x_{ij} = 0\}| < k$.

Note that the invariants of Theorem 4.1.3 imply the satisfaction of constraints (2.14),(2.17),(2.18) and (2.10) for the number of rounds we consider.

Lemma 4.1.5. *Let (y, x) be a node-solution defined at depth $k \leq \frac{n}{10}$ of the evolution tree T_s. If (y, x) satisfies Invariants 1-4, then (y, x) meets constraints* (2.14)*,* (2.17)*,* (2.18) *and* (2.10).

Proof. Constraints (2.17) and (2.18) are obviously satisfied by Invariants 1 and 2. For $i \in Costly$, if $y_i \notin \{0, 1\}$ we have, by Invariant 1, that y_i is always "close" to b while, by Invariant $2.b$, for every j a non-integer x_{ij} is "close" to $\frac{a}{n}$ and the total demand assigned to i is, by Invariant 4, $\sum_j x_{ij} \leq (nU + 1)\frac{a}{n} + k \leq U(b - 2k\frac{a}{n}) \leq Uy_i$. So (2.14) and (2.10) are satisfied.

Also note that if $y_i = 0$, by our construction $\sum_j x_{ij} = 0$, thus again (2.14) and (2.10) are satisfied.

For $i \in Costly$, if $y_i = 1$ we have by Invariant $2.c$, that for every j a non-integer x_{ij} is at most $(\frac{a}{n} + 2k\frac{a(1-a)}{n^2})b^{-1}(1 + \frac{1}{10})$ and the total demand assigned to i is by Invariant 4 at most $\sum_j x_{ij} \leq ((nU + 1)\frac{a}{n} + k)(1 + \frac{1}{10})b^{-1} \leq U$. So (2.14) and (2.10) are satisfied.

Similarly for (2.14) and (2.10) involving some facility in $Cheap$. □

Thus, each time we prove the feasibility of the constructed solution, we only have to ensure that (2.15) holds.

We now explain the structure of the inductive step that produces the children of node v, where v is at depth $k < n/10$. We distinguish cases according to the variable that is touched. For every case, the proof has three parts. First, we give the construction of the child node (y', x') based on the current node-solution (y, x). Second, we prove that (y', x') is feasible for (LP-classic). Third, we show that (y', x') satisfies the four invariants in the statement of the theorem.

In order to unerstand intuitively the upcoming calculations for the invariants we suggest that the reader should focus on the dominant term in the rhs (or lhs) of each invariant. The loss in accuracy in the upper (lower) bounds obtained this way will be offset in the inductive step by increasing the coefficients of the minor terms. For example, in Invariant 2a the dominant term in the lhs is $(1 - a)/n$ and the minor term is $2kab^{-1}/n^2$. In the inductive step, $2k$ will become $2k + 2$.

Case 1: type 1 children

<u>subcase $1a$</u>: touched variable is y_{i_k}, $i_k \in Costly$

CONSTRUCTION

Consider the type 1 child (y', x') of v corresponding to variable y_{i_k}. Variables y_{i_k}, x_{i_kj} for all j are multiplied by a factor of $1/y_{i_k}$ and so $y'_{i_k} = 1$. Note that since we only consider cases where y_{i_k} is fractional, by the inductive hypothesis we have that for all variables x_{i_kj}, Invariant 2.b holds. For the variables involving facilities $i' \in Costly - \{i_k\}$ and for all j, we set $y'_{i'} = y_{i'}$, $x'_{i'j} = x_{i'j}$. For all j and for all $i' \in Cheap$ such that $x_{i'j} \neq 0$ we set $x'_{i'j} = x_{i'j}(1 - \frac{(1/y_{i_k}-1)x_{i_kj}}{x_{i'j}t})$, where t is the number of facilities in $Cheap$ for which j is assigned with a non-zero fraction (so $t \geq n - k$).

FEASIBILITY

Constraint (2.15) is satisfied by construction:

$$\sum_i x'_{ij} = \sum_i x_{ij} + (1/y_{i_k} - 1)x_{i_kj} - \sum_{i \in Cheap | x_{ij}>0} \frac{(1/y_{i_k}-1)x_{i_kj}}{t} = \sum_i x_{ij} = 1$$

INVARIANTS

Invariant 1. For $i \in Costly - \{i_k\}$, y_i remains unchanged so Invariant 1 holds by the inductive hypothesis (from now on abbreviated as *i.h.*).

Invariant 2. For $i \in Cheap$ we have $2.a$:

$$x'_{ij} = x_{ij} - \frac{(1/y_{i_k}-1)x_{i_kj}}{t} \geq \quad \text{(by Invariants 1, 2 of i.h. and being generous)}$$
$$\frac{1-a}{n} - 2k\frac{a}{n^2}b^{-1} - 2b^{-1}\frac{a}{n^2} \geq$$
$$\frac{1-a}{n} - 2(k+1)\frac{a}{n^2}b^{-1}$$

For $i \in Costly - \{i_k\}$, $2.b$ holds since variables x_{ij} were not changed. For x'_{i_kj} we need to show $2.c$:

$$x'_{i_kj} = x_{i_kj}\frac{1}{y_{i_k}} \leq \quad \text{(by Invariants } 2.b\text{, 1)}$$
$$(\frac{a}{n} + 2k\frac{a(1-a)}{n^2})(b + 2k\frac{a}{n})^{-1} \leq$$
$$(\frac{a}{n} + 2k\frac{a(1-a)}{n^2})b^{-1}(1 + \frac{1}{10})$$

Invariant 3. Observe than the total demand assigned to each facility in $Cheap$ was decreased so Invariant 3 holds by the inductive hypothesis.

Invariant 4. For $i \in Costly - \{i_k\}$ Invariant 4 holds by inductive hypothesis. For i_k we have $4.b$:

$$\sum_j x'_{i_kj} = 1/y_{i_k} \sum_j x_{i_kj} \leq \quad \text{(by the invariants of i.h.)}$$
$$b^{-1}(1 + \frac{1}{10})((nU + 1)\frac{a}{n} + k)$$

<u>subcase $1b$:</u> touched variable is x_{i_kj*}, $i_k \in Costly$

CONSTRUCTION

Consider the type 1 children (y', x') of v corresponding to variable x_{i_kj*}. We obtain y'_{i_k} by multiplying y_{i_k} by a factor of $1/y_{i_k}$ and so $y'_{i_k} = 1$ (and of course $x'_{i_kj*} = 1$, and $x'_{ij*} = 0$ for $i \neq i_k$). Every other variable of (y', x') is the same as in (y, x).

FEASIBILITY

The feasibility of this case is trivial.

INVARIANTS

The Invariants 1, 2, 3 in this case are satisfied trivially. For 4 we have for facility i_k:

$\sum_j x'_{i_k j} \leq$ (variable $x_{i_k j*}$ becomes 1)
$\sum_j x_{i_k j} + 1 \leq$ (by 4 of i.h.)
$(nU+1)\frac{a}{n} + k + 1$ if $y_{i_k} \neq 1$ or
$((nU+1)\frac{a}{n} + k + 1)b^{-1}$ if $y_{i_k} = 1$

In either of the two cases Invariant $4.b$ holds for the new value y'_{i_k}.

<u>subcase $1c$</u>: touched variable is $x_{i_k j*}$, $i_k \in Cheap$

CONSTRUCTION

Consider the type 1 children (y', x') of v corresponding to variable $x_{i_k j*}$. We obtain $y'_i, i \in Costly$ with $y_i \notin \{0,1\}$ by multiplying y_i by a factor of $(1 - \frac{(1/y_i - 1)x_{ij*}}{x_{i_k j*} t})$, where t is again the number of facilities in $Cheap$ for which $j*$ is assigned with a non zero fraction (so $t \geq n-k$). Of course $x'_{i_k j*} = 1$, and $x'_{ij*} = 0$ for $i \neq i_k$ as usual. Every other variable of (y', x') is the same as in (y, x).

FEASIBILITY

Obviously (2.15) is satisfied. All other constraints are satisfied by Lemma 4.1.5.

INVARIANTS

Invariant 1. For each $i \in Costly$ such that $y_i \notin \{0,1\}$ we have:

$y'_i = y_i(1 - \frac{(1/y_i - 1)x_{ij*}}{x_{i_k j*} t}) \geq$ (by Invariant 1 of i.h.)
$b - 2k\frac{a}{n} - (\frac{(1-y_i)x_{ij*}}{x_{i_k j*} t}) \geq$ (by Invariants 1, $2.b$ of i.h.)
$b - 2k\frac{a}{n} - (\frac{(1-b+2k\frac{a}{n})(\frac{a}{n}+2k\frac{a(1-a)}{n^2})}{(\frac{1-a}{n} - 2k\frac{a}{n^2}b^{-1})t})$
Observe that the numerator of the last fraction is almost $\frac{a}{n}$ and the denominator is almost $9/10$ since $t \geq \frac{9n}{10}$. The additive constant 2 in the last term of the following equation is enough to absorb the "noise" in our arguments. Therefore the last quantity is at least
$b - 2k\frac{a}{n} - 2\frac{a}{n} = b - (2k+2)\frac{a}{n}$

Invariant 2. Variables x'_{ij} remain unchanged for $j \neq j*$. For $j*$, $x'_{i_k j*} = 1$ while for $i \neq i_k$ we have $x'_{ij*} = 0$, so 2 is trivially satisfied.

Invariant 3. For $i \in Cheap - \{i_k\}$ the total demand is decreased (because of $j*$). For i_k:

$\sum_j x'_{i_k j} \leq \sum_j x_{i_k j} + 1 \leq$ (by 3 of i.h.)
$(nU+1)\frac{1-a}{n} + 2k(nU+1)\frac{a}{n^2} + 1 \leq (nU+1)\frac{1-a}{n} + 2(k+1)(nU+1)\frac{a}{n^2}$

Invariant 4. The demand assigned to facilities in $Costly - \{i_k\}$ is decreased (because of $j*$) so $4.a$, $4.b$ trivially hold.

Case 2: type 2 children

<u>subcase $2a$</u>: touched variable is y_{i_k}, $i_k \in Costly$

CONSTRUCTION

Consider the type 2 children (y', x') of v corresponding to variable $y_{i_k} \notin \{0,1\}$. Let $f =$

$\frac{y_{i_k}}{1-y_{i_k}}$. Solution (y', x') is dictated by its twin type 1 solution (case 1a): variables y_{i_k}, x_{i_kj} for all j, are multiplied by a factor of $(1 - f(1/y_{i_k} - 1))$ and so $y'_{i_k} = 0$ and $x'_{i_kj} = 0$, that is facility i_k closes. The variables involving facilities $i' \in Costly - \{i_k\}$, namely $y'_{i'}, x'_{i'j}$ for all j, have the same value as in (y, x). For all j and for $i' \in Cheap$ such that $x_{i'j} \neq 0$ we have $x'_{i'j} = x_{i'j}(1 + \frac{f(1/y_{i_k}-1)x_{i_kj}}{x_{i'j}t})$, where t is again the number of facilities in $Cheap$ for which j is assigned with a non zero fraction (so $t \geq n - k$).

FEASIBILITY

Constraint (2.15) is satisfied by Lemma 4.1.3.

INVARIANTS

Invariant 1. For $i \in Costly - \{i_k\}$, y_i remain unchanged so Invariant 1 holds by inductive hypothesis.

Invariant 2. For $i \in Cheap$ we have $2.a$:

$$x'_{ij} = x_{i'j} + \frac{f(1/y_{i_k}-1)x_{i_kj}}{t} \leq \qquad \text{(by Invariants 1, 2 of i.h.)}$$
$$\frac{1-a}{n} + 2k\frac{1-a}{n^2} + 2\frac{a}{n^2} \leq \qquad \text{(being very generous)}$$
$$\frac{1-a}{n} + (2k+2)\frac{1-a}{n^2}$$

For $i \in Costly - \{i_k\}$, $2.b$ holds since variables x_{ij} were not changed.

Invariant 3. For $i \in Cheap$ we have:

$$\sum_j x'_{ij} = \sum_j x_{ij} + \frac{1}{n}\sum_j x_{i_kj} + o(1) \leq \qquad \text{(by Invariants 3, 4 of i.h.)}$$
$$(nU+1)\frac{1-a}{n} + 2k(nU+1)\frac{a}{n^2} + \frac{(nU+1)\frac{a}{n}+k}{n} + o(1) \leq$$
$$(nU+1)\frac{1-a}{n} + (2k+2)(nU+1)\frac{a}{n^2}$$

The $o(1)$ above is due to the fact that at most k assignment variables for some cheap facilities may have been touched as type 2 and are 0. For those same clients the assignment to i_k is fractional, so the demand corresponding to them that was assigned to i_k, must be distributed among the, at least $n - k$, available cheap facilities. That additional demand is at most $\frac{k(\frac{a}{n}+2k\frac{a(1-a)}{n^2})}{n-k} = o(1)$.

Invariant 4. For $i \in Costly - i_k$ Invariant 4 holds by inductive hypothesis. For i_k we have $\sum_j x_{i_kj} = 0$.

<u>subcase $2b$:</u> touched variable is x_{i_kj*}, $i_k \in Costly$

CONSTRUCTION

Consider the type 2 children (y', x') of v corresponding to variable x_{i_kj*}. Let $f = \frac{x_{i_kj*}}{1-x_{i_kj*}}$. Solution (y', x') is dictated by its twin type 1 node-solution (case $1b$): variable y'_{i_k} is obtained by multiplying y_{i_k} by a factor of $(1 - f(1/y_{i_k} - 1))$ and for $i \neq i_k$, $x'_{ij*} = x_{ij*}(1 + f)$ and $x'_{i_kj*} = 0$. Every other variable of (y', x') is the same as in (y, x).

FEASIBILITY

The feasibility of this case is trivial by Lemma 4.1.3.

INVARIANTS

Invariant 1. For facilities $i \in Costly - \{i_k\}$ the proof is trivial (no change). Same if $y_{i_k} = 1$. If $y_{i_k} \notin \{0, 1\}$ we have:

$$\begin{aligned} &y'_{i_k} = y_{i_k}(1 - f(1/y_{i_k} - 1)) = \\ &y_{i_k} - (1 - y_{i_k})\tfrac{x_{i_k j*}}{1 - x_{i_k j*}} \geq && \text{(by Invariants 1, 2 of i.h.)} \\ &b - 2k\tfrac{a}{n} - 2\tfrac{a}{n} \geq \\ &b - (2k+2)\tfrac{a}{n} \end{aligned}$$

Invariant 2. For client $j*$ and facility $i \in Cheap$ we have $2.a$:

$$\begin{aligned} &x'_{ij*} = x_{ij*}(1 + f) \leq && \text{(by Invariant 2 of i.h.)} \\ &\tfrac{1-a}{n} + 2k\tfrac{1-a}{n^2} + 2\tfrac{(1-a)ab^{-1}}{n^2} \leq \\ &\tfrac{1-a}{n} + (2k+2)\tfrac{(1-a)}{n^2} && \text{(recall } a, b = \Theta(n^{-2})\text{)} \end{aligned}$$

For client $j*$ and facility $i \in Costly$ we have $2.b$:

$$\begin{aligned} &x'_{ij*} = x_{ij*}(1 + f) \leq \tfrac{a}{n} + 2k\tfrac{a(1-a)}{n^2} + 2\tfrac{a^2 b^{-1}}{n^2} \leq \\ &\tfrac{a}{n} + 2(k+1)\tfrac{a(1-a)}{n^2} \end{aligned}$$

Similarly for Invariant $2.c$.

Invariant 3. For $i \in Cheap$:

$$\begin{aligned} &\textstyle\sum_j x'_{ij} \leq \sum_j x_{ij} + 1 \leq && \text{(by 3 of i.h.)} \\ &(nU+1)\tfrac{1-a}{n} + 2(k+1)(nU+1)\tfrac{a}{n^2} \end{aligned}$$

Invariant 4. For i_k the total demand is decreased while for $i \in Costly - \{i_k\}$:

$$\begin{aligned} &\textstyle\sum_j x'_{ij} \leq \sum_j x_{ij} + 1 \leq && \text{(by 3 of i.h.)} \\ &(nU+1)\tfrac{a}{n} + k + 1 && \text{if } y_i \neq 1 \text{ or} \\ &((nU+1)\tfrac{a}{n} + k + 1)b^{-1} && \text{if } y_i = 1 \end{aligned}$$

<u>subcase $2c$</u>: touched variable is $x_{i_k j*}$, $i_k \in Cheap$

CONSTRUCTION

Consider the type 2 children (y', x') of v corresponding to variable $x_{i_k j*}$. Let $f = \frac{x_{i_k j*}}{1 - x_{i_k j*}}$. Solution (y', x') is dictated by its twin type 1 node-solution (case $1c$): variables $y'_i \notin \{0, 1\}, i \in Costly$, are obtained by multiplying y_i by a factor of $(1 + f\frac{(1/y_i - 1)x_{ij*}}{x_{i_k j*} t})$, where t is again the number of facilities in $Cheap$ for which j is assigned with a non zero fraction (so $t \geq n - k$). For $i \neq i_k$, $x'_{ij*} = x_{ij*}(1 + f)$ while $x'_{i_k j*} = 0$. Every other variable of (y', x') is the same as in (y, x).

FEASIBILITY

The satisfaction of (2.15) is ensured by Lemma 4.1.3.

INVARIANTS

Invariant 1. For facility $i \in Costly$ such that $y_i \notin \{0, 1\}$ we have:

$y'_i = y_i(1 + f\frac{(1/y_i-1)x_{ij*}}{x_{i_kj*}n}) =$
$y_i + (1 - y_i)\frac{x_{ij*}}{(1-x_{i_kj*})t} \leq$ (by Invariants 1, 2 of i.h.)
$b + 2k\frac{a}{n} + 2\frac{a}{n^2} \leq$
$b + (2k + 2)\frac{a}{n}$

Invariant 2. For client $j*$ and facility $i \in Cheap$ we have $2.a$:

$x'_{ij*} = x_{ij*}(1 + f) \leq$ (by Invariant 2 of i.h.)
$\frac{1-a}{n} + 2k\frac{1-a}{n^2} + 2\frac{(1-a)^2}{n^2} \leq$
$\frac{1-a}{n} + (2k + 2)\frac{1-a}{n^2}$

For client $j*$ and facility $i \in Costly$ with $y_i \notin \{0, 1\}$ we have $2.b$:

$x'_{ij*} = x_{ij*}(1 + f) \leq$ (by Invariant 2 of i.h.)
$\frac{a}{n} + 2k\frac{a(1-a)}{n^2} + 2\frac{(1-a)a}{n^2} \leq$
$\frac{a}{n} + (2k + 2)\frac{a(1-a)}{n^2}$

For client $j*$ and facility $i \in Costly$ with $y_i = 1$ we have $2.c$:

$x'_{ij*} = x_{ij*}(1 + f) \leq$ (by Invariant 2 of i.h.)
$(\frac{a}{n} + 2k\frac{a(1-a)}{n^2})b^{-1}(1 + \frac{1}{10}) + 2\frac{(1-a)ab^{-1}}{n^2} \leq$
$\frac{a}{n} + (2k + 2)\frac{a(1-a)}{n^2}$

Invariant 3. The demand assigned to i_k is decreased. For $i \in Cheap - \{i_k\}$:

$\sum_j x'_{ij} \leq \sum_j x_{ij} + 1 \leq$ (by 3 of i.h.)
$(nU + 1)\frac{1-a}{n} + 2(k + 1)(nU + 1)\frac{a}{n^2}$

Invariant 4. For $i \in Costly$:

$\sum_j x'_{ij} \leq \sum_j x_{ij} + 1 \leq$ (by 3 of i.h.)
$(nU + 1)\frac{a}{n} + k + 1$ if $y_i \neq 1$ or
$((nU + 1)\frac{a}{n} + k + 1)b^{-1}$ if $y_i = 1$

The case analysis is complete. It is easy to verify that the constructed vectors satisfy the conditions stated in Observation 4.1.1 and Proposition 4.1.2. It remains to show that the vectors we constructed for node v satisfy the symmetry requirements, i.e., the conditions in Proposition 4.1.1 and thus are indeed witnesses.

Lemma 4.1.6. *The symmetry Conditions (ii) and (iii) stated in Proposition 4.1.1, is satisfied for the children of node-solution v.*

Proof. By construction we never alter integer values of variables, as dictated by the LS procedure, therefore Condition (ii) of Observation 4.1.2 holds.

When a variable y_i, $i \in Costly$, is touched then for the symmetry between y_i and each other variable we have:

For all j, variables x_{ij} are multiplied by $1/y_i$ (case $1a$), and when some x_{ij} is touched, variable y_i is multiplied by $1/y_i$ (case $1b$).

For all j, variables $x_{i'j}$, $i' \in Cheap$, are multiplied by $(1 - (1/y_i - 1)\frac{x_{ij}}{x_{i'j}t})$ (case $1a$), and when some $x_{i'j}$ is touched, variable y_i is multiplied by $(1 - (1/y_i - 1)\frac{x_{ij}}{x_{i'j}t})$ (case $1c$).

For all j, variables $y_{i''}$, $x_{i''j}$, $i'' \in Costly - \{i\}$, are multiplied by 1 (case $1a$), and when $y_{i''}$ or some $x_{i''j}$ is touched, variable y_i is multiplied by 1 (cases $1a$, $1b$).

When a variable x_{ij}, $i \in Costly$, is touched then for the symmetry between x_{ij} and each other variable we have:

For all $j' \neq j$ and all i', variables $x_{i'j'}$ are multiplied by 1 (case $1b$), and when some $x_{i'j'}$ is touched, variable x_{ij} is multiplied by 1 (cases $1b$, $1c$).

For $i' \neq i$, variables $x_{i'j}$ are multiplied by 0 (case $1b$), and when some $x_{i'j}$ is touched, variable x_{ij} is multiplied by 0 (cases $1b$, $1c$).

Finally, when variable x_{ij}, $i \in Cheap$, is touched then for the symmetry between x_{ij} and each other variable, the remaining cases that have not been covered above are:

For all $j' \neq j$ and all $i' \in Cheap$, variables $x_{i'j'}$ are multiplied by 1 (case 1c), and when $x_{i'j'}$ is touched, variable x_{ij} is multiplied by 1 (case $1c$).

For all $i' \in Cheap$, variables $x_{i'j}$ are multiplied by 0 (case $1c$), and when $x_{i'j}$ is touched, variable x_{ij} is multiplied by 0 (case $1c$).

□

By Lemma 4.1.2, the proof of Theorem 4.1.3 is now complete.

The proof can yield a tradeoff between the number of rounds as a function of the dimension of the instance and the integrality gap, which can be obtained by toying with the quantities U, a, and b that are left as parameters. One can obtain a higher gap that survives for a smaller number of rounds.

4.1.4 The mixed LS_+ case

It is not hard to see that the proof of Theorem 4.1.1 also yields the same lower bound for the mixed LS_+ procedure: simply restrict the constructed protection matrices to the facility opening y variables. The resulting matrices are of the form $(1, y)(1, y)^T + \mathrm{Diag}(0, y - y^2)$ which are well-known to be positive semidefinite (see Theorem 4.1 in [31]). The following theorem is immediate.

Theorem 4.1.4. *There is a family of instances $\mathcal{I}$ of uniform* CFL*, such that for every sufficiently large n, there is an instance in the family with $2n$ facilities and $n^4 + 1$ clients for which the relaxation produced by the mixed LS_+ procedure in $\Omega(n)$ rounds has integrality gap $\Omega(n)$.*

4.2 Sherali-Adams gap for CFL

In Section 4.2.1 the definition of the Sherali-Adams hierarchy is given along with some basic facts. In Section 4.2.2 we give the proof of unboundedness of the gap of the SA CFL relaxation for an asymptotically tight number of rounds. In Section 4.2.3 we argue on the robustness of our proof and in Section 4.2.4 we extend our results to the LBFL problem. Then in Section 4.3 we give our results regarding the submodular inequalities.

4.2.1 The SA hierarchy

We proceed to define the Sherali-Adams hierarchy [59]. Consider a polytope $P \subseteq [0,1]^d$ defined by the linear constraints $Ax - b \leq 0$ We define the polytope $\mathrm{SA}^k(P) \subseteq \mathbb{R}^d$ as follows. For every constraint $\pi(x) \leq 0$ of P, for every set of variables $X \subseteq \{x_i \mid i = 1, \ldots, d\}$ such that $|X| \leq k$, and for every $W \subseteq X$, consider the nonlinear inequality: $\pi(x) \prod_{x_i \in X-W} x_i \prod_{x_i \in W}(1-x_i) \leq 0$. We call $\prod_{x_i \in X-W} x_i \prod_{x_i \in W}(1-x_i)$ a (X, W)-*multiplier*. When the sets X, W are clear from the context, we will refer simply to a *mutliplier*. After expanding the lhs of the inequalities, we linearize the resulting non-linear system as follows: (i) substitute x_i for x_i^2 for all i (ii) replace $\prod_{x_i \in I} x_i$ with x_I for each set $I \subseteq \{x_i \mid i = 1, \ldots, d\}$. Call the resulting lifted polyhedron Q. $\mathrm{SA}^k(P)$ is the projection of Q onto the original variables x_i, where variables $x_{\{x_i\}}$ are treated as being equal to x_i. We call $\mathrm{SA}^k(P)$ the polytope *obtained from* P *at level* k of the SA hierarchy. Given a cost vector $c \in \mathbb{R}^d$, the *relaxation obtained from* P *at level* k of SA is $\min\{c^T x \mid x \in \mathrm{SA}^k(P)\}$.

The definition of the SA procedure extends to the mixed integer case as well: given a valid relaxation $P \in [0,1]^d$ of a mixed integer set X, such that $\mathrm{conv}(X) \cap (\{0,1\}^n \times [0,1]^p) = P \cap (\{0,1\}^n \times [0,1]^p)$, the *level* k *Sherali-Adams (*SA*) procedure*, $k \geq 1$, is defined as follows [59]. Let P be defined by the linear constraints $Ax - b \leq 0$. For every constraint $\pi(x) \leq 0$ of P, for every set of variables $U \subseteq \{x_i \mid i = 1, \ldots, n\}$ such that $|U| \leq k$, and for every $W \subseteq U$, consider the *lifted* valid constraint: $\pi(x) \prod_{x_i \in U-W} x_i \prod_{x_i \in W}(1 - x_i) \leq 0$. Linearize the system obtained this way by replacing (i) x_i^2 with x_i for all i (ii) $\prod_{x_i \in I \subseteq [n]} x_i$ with x_I and (iii) $x_k \prod_{x_i \in I \subseteq [n]} x_i$, where $k \in \{n+1, \ldots, d\}$ with v_{Ik}. $\mathrm{SA}^k(P)$ is the projection of the resulting linear system onto the original variables $\{x_1, \ldots, x_d\}$. We call $\mathrm{SA}^k(P)$ the relaxation *obtained from* P *at level* k of the SA hierarchy. It is well-known that $\mathrm{SA}^n(P) = \mathrm{conv}(X)$ (see, e.g., [10]). If X is a 0-1 set, i.e., $X \subseteq \{0,1\}^d$, then it is easy to see that the definition is actually the one of the previous paragraph.

Recall the family of instances $\mathcal{I}$ and the solution s. We will show in Section 4.2.2 that the solution s to this same instance is feasible for a number of SA levels, which is linear in the number $2n$ of facilities, more specifically for $n/10$ levels. On the other hand, by Theorem 4.1.2, at level $2n$ the relaxation obtained expresses the integral polytope since SA refines LS. As an example of the robustness of our SA construction against the addition of new constraints, we show in Section 4.2.3 that a bad fractional solution remains feasible for $\Theta(n)$ levels of SA even if we add the valid inequality (4.22).

The following lemma, which is implicit in previous works [28, 23] and is explicitly stated in [30] (in a slightly different form) gives sufficient conditions for a solution to be feasible at level k of the SA hierarchy.

Lemma 4.2.1. *[28, 23, 30] Let $v(\pi, q)$ be the set of variables appearing in a nonlinear constraint obtained from π when multiplied by a (X, W)-multiplier $q = \prod_{i \in X-W} x_i \prod_{i \in W}(1 - x_i)$, for some X and W. A solution z is feasible for the lifted k-level* SA *relaxation if for every* *constraint π and each such multiplier z with at most k distinct variables:*

1. *There is a distribution $E_{\pi,q}$ over 0-1 assignments of the original variables that satisfy π and*
2. *(Consistency Condition) For any set $J \subseteq v(\pi, q)$ with $|J| \leq k$ we have $P_{E_{\pi,q}}[$all elements of J have value 1$] = z_J$.*

The above lemma is based on the fact that distributions over (convex combinations of) feasible solutions are by convexity feasible. The connection of the SA hierarchy to distributions turns out to be even deeper. In fact the feasible solution of the SA relaxations try to mimic distributions, and they become more effective to this end as the level increases. Moreover, one can exclude all redundant inequalities from the starting relaxation without loss in the tightness of the obtained SA relaxation. We give below a few properties of the feasible solutions of the SA relaxations that are known but, to our knowledge, not explicitly stated in the literature.

Lemma 4.2.2. *Consider some polytope $P \subseteq [0,1]^n$ defined by the system $Ax \leq b$. Let $\pi_i : a_i^T x \leq b_i$ be an inequality of the system such that $\pi_i = \sum_j \lambda_j \pi_j$ where π_j are also inequalities of the system. Let $P' \subseteq [0,1]^d$ be the polytope defined by the system $A'x \leq b'$ which is obtained from $Ax \leq b$ by omitting π_i. Then $\mathrm{SA}^k(P) = \mathrm{SA}^k(P')$, i.e. the SA system does not depend on the syntactical description of P.*

Proof. It suffices to show that the lifted polytopes are actually the same. Obviously $\mathrm{SA}^k(P) \supseteq \mathrm{SA}^k(P')$. Now consider a constraint $\pi_i z$ of the lifted P for some k-level (U, W)-multiplier z. This constraint is redundant in the system of the lifted P' as it can obtained as $\pi_i z = \sum_j \lambda_j \pi_j z$. Thus $\mathrm{SA}^k(P) = \mathrm{SA}^k(P')$. □

The follow lemma certifies that the values of the variables of SA solutions share some properties with the probabilities of the corresponding events, with respect to some distribution.

Lemma 4.2.3. *Let $\bar{x} \in \mathrm{SA}^k(P)$ and $\bar{x}_i = 1$ ($\bar{x}_i = 0$). Then for a feasible vector z for the lifted polytope that projects to $\bar{x}$ it is true that for $I : x_i \notin I$ $z_I = z_{I \cup \{x_i\}}$ ($z_{I \cup \{x_i\}} = 0$). (For notational simplicity assume $z_\emptyset = 1$.)*

Proof. All box constraints $0 \leq x_j \leq 1$ are valid by definition. First observe that by multiplying $x_i \leq 1$ by $\prod_{l \in I} x_l$ we get $z_{I \cup \{x_i\}} \leq z_I$.

The proof is by induction on $|I|$. For $|I| = 0$ the claim obviously holds. Now, if $|I| > 0$ is even, consider the constraint of the lifted polytope obtained by multiplying the constraint $x_j \geq 0$ for some $x_j \in I$ with the multiplier $(1 - x_i) \prod_{x_l \in I - \{x_j\}} (1 - x_l)$. The obtained constraint is $\sum_{I' \subseteq I \cup \{x_i\} | x_j \in I'} (-1)^{|I'|-1} z_{I'}$. By the inductive hypothesis the term $z_{I'}$ with $|I'| < |I|$ and $x_i \notin I'$ cancels the term $z_{I' \cup \{x_i\}}$. So what is left is $z_{I \cup \{x_i\}} - z_I \geq 0$ which, together with $z_I \geq z_{I \cup \{x_i\}}$, gives the claim. If $|I| > 0$ is odd the proof is similar, we just start from $(1 - x_j) \geq 0$ and we multiply by $(1 - x_i) \prod_{x_l \in I - \{x_j\}} (1 - x_l)$. For $x_i = 0$ the claim is proved again similarly.

□

4.2.2 SA gaps for approximate CFL

In this subsection, we state and prove our main SA result, namely Theorem 4.2.1. We give a solution z^b to the lifted k-level relaxation whose projection to the singleton variables is equal to the bad solution s. Our strategy is the following. We define z^b by defining a distribution E over 0-1 assignments to the original variables that are not necessarily feasible integer solutions and we let $z_I = P_E[\text{variables in } I \text{ have value } 1]$. More precisely, all of the assignments in the support of distribution E satisfy constraints (2.14), (2.17), (2.18) and (2.15) of the LP but some of them do not satisfy constraints (2.10). Thus, by Lemma 4.2.1, solution z^b is feasible for the lifted k-level relaxation if we ignore constraints (2.10): for every constraint π (other than (2.10)) and every (X, W)-multiplier q just let $E_{\pi,q} = E$. Then we employ a special treatment in order to prove the feasibility of z^b for constraints (2.10) and any k-level (X, W)-multiplier q: we bound the ratio of the values of the involved variables z_I by relating the probabilities of the corresponding events and then we prove that the inequality holds.

4.2.2.1 Defining Distribution E

We describe a probabilistic experiment which induces the distribution E over integer solutions satisfying constraints (2.14), (2.17), (2.18) and (2.15) of the LP and moreover the expected (y, x) of E is solution s. We begin by defining some parameters for the integer solutions that will make up the support of the distribution E. The solutions in the support of E have the property that at most one costly facility is opened while all of the cheap facilities are opened.

Let $w_{i_{co}} = \frac{\sum_{j \in C} x_{i_{co} j}}{y_{i_{co}}}$ be the number of clients assigned to facility i_{co} in the integer solutions in the support of E when facility $i_{co} \in Costly$ is opened. To simplify the presentation let us assume for now that $w_{i_{co}}$ and the values we subsequently define are integers (we discuss at the end of the proof how to handle fractional values). Let $w^1_{i_{ch}} = \frac{|C| - w_{i_{co}}}{|Cheap|}$ be the number of clients assigned to facility $i_{ch} \in Cheap$ when i_{co} is the only opened costly facility. Likewise let $w^2_{i_{ch}} = \frac{|C|}{|Cheap|}$ be the number of clients assigned to facility $i_{ch} \in Cheap$ in each integer solution in E where no costly facility is opened. Note that in the latter case the capacities are not respected. The following procedure produces distribution E.

Pick a costly facility i_{co} with probability $y_{i_{co}}$, with probability $1 - \sum_{i \in Costly} y_i$ no costly facility is picked. If some i_{co} is picked, we call this *Case 1*, then consider n bins corresponding to the n cheap facilities each one having w^1_{ch} slots and 1 bin corresponding to i_{co} having $w_{i_{co}}$ slots. Randomly distribute $|C|$ balls to the slots of the $n + 1$ bins, with exactly one ball in each slot. If no costly is picked - we call this *Case 2* - then consider n bins corresponding to the n cheap facilities each one having w^2_{ch} slots. Randomly distribute $|C|$ balls to the slots of the n bins, with exactly one ball in each slot - the outcome of the last case does not respect the capacities.

Note that the above experiment induces a distribution over integer solutions satisfying (2.14), (2.17), (2.18) and (2.15), since every client is assigned to exactly one opened facility in each outcome. It remains to show that the expected (y, x) vector with respect to E (and thus the projection of z^b) is the solution s.

The cheap facilities are always open, and the costly are open a fraction of the time that is equal to the value of their corresponding y variable. The expected demand assigned to each $i_{co} \in Costly$ is $y_{i_{co}} w_{i_{co}}$ which is the total demand assigned to i_{co} by s. Since the clients have the same probability of being tossed in the bin corresponding to i_{co}, the expected assignment of each client j to i_{co} is the same as in s.

Similarly, we can prove that the expected assignments to the cheap facilities are as required. Observe that in every outcome of the experiment the demand not assigned to costly facilities is exactly the demand assigned to cheap facilities. Since we have proved that the expected assignments to the costly facilities are those of the bad solution s, by linearity of expectation we get that the total assignments to all cheap facilities are $\sum_{i \in Cheap} \sum_j x_{ij}$ (the total assignment of each client adds up to 1). By the symmetric way the cheap facilities are handled in the experiment, we have that the total expected demand assigned to each $i \in Cheap$ is $\sum_{j \in C} x_{ij}$. Moreover, by the symmetric way the clients are assigned to i in the experiment, we get that the expected assignment of each $j \in C$ to i is $\frac{\sum_{j \in C} x_{ij}}{|C|} = x_{ij}$.

To handle the case where the $w_{i_{co}}, w^1_{i_{ch}}, w^2_{i_{ch}}$ are not integers (which is actually always the case), we do the following: each time costly facility i_{co} is picked, we set the number of slots of the corresponding bin to $\lfloor w_{i_{co}} \rfloor$ with probability $1 - (w_{i_{co}} - \lfloor w_{i_{co}} \rfloor)$, otherwise set the slots to $\lceil w_{i_{co}} \rceil$; this ensures that the expected number of slots is $w_{i_{co}}$. The same rationale applies to the remaining cases of the construction. If the number of slots of i_{co} is set to $\lfloor w_{i_{co}} \rfloor$, then we pick some $n(\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n} - \lfloor (\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n}) \rfloor)$ cheap facilities at random and set their corresponding number of slots to $\lceil \frac{|C|-\lfloor w_{i_{co}} \rfloor}{n} \rceil$ and the number of slots of the rest of the cheap facilities to $\lfloor \frac{|C|-\lfloor w_{i_{co}} \rfloor}{n} \rfloor$. Otherwise, pick some $n(\frac{|C|-\lceil w_{i_{co}} \rceil}{n} - \lfloor (\frac{|C|-\lceil w_{i_{co}} \rceil}{n}) \rfloor)$ cheap facilities at random and set their corresponding number of slots to $\lceil \frac{|C|-\lceil w_{i_{co}} \rceil}{n} \rceil$ and the number of slots of the rest to $\lfloor \frac{|C|-\lceil w_{i_{co}} \rceil}{n} \rfloor$. In every case the expected number of slots per facility is the same as in the initial description of the experiment where we assumed that $w_{i_{co}}, w^1_{i_{ch}}, w^2_{i_{ch}}$ are integers.

4.2.2.2 Satisfaction of constraints (2.10)

We will show that for any capacity constraint π of type (2.10) and any k-level (X, W)-multiplier q, $k \leq n/10$, the constraint of the lifted polytope obtained by π and q is satisfied by z^b.

We first give a general fact about solutions of the lifted SA relaxation which are defined via some distribution D that will prove useful when arguing on the feasibility of solution z^b. Let $\mathcal{E}_{X,W}$ be the event that all variables in $X - W$ have value 1 and all variables in W have value 0.

Lemma 4.2.4. *Let z^b be a solution for the lifted SA relaxation which is defined via a distribution D over 0-1 assignments, i.e. $z^b_I = P_D[$all variables in I have value $1]$. Let q be a (X, W)-multiplier and consider the expression $A_q(z)$ of variables z_I obtained after expanding q and linearizing. Then the value of the expression $A_q(z^b)$ for the lifted solution z^b is equal to $P_D[\mathcal{E}_{X,W}]$.*

Proof. The proof is by induction on $|W|$. For $|W| = 0$ the expression A_q is simply z_X which by definition has value equal to $P_D[$all variables in X have value $1]$ in z^b. Suppose the statement is true for $|W| \leq t$. Then for $|W| = t + 1$ we have
$q = \prod_{x_i \in X-W} x_i \prod_{x_i \in W} (1 - x_i) =$
$\prod_{x_i \in X-W} x_i \prod_{x_i \in W-\{x*\}} (1 - x_i) - \prod_{x_i \in X-(W-\{x^*\})} x_i \prod_{x_i \in W-\{x*\}} (1 - x_i) = q_1 - q_2$
where q_1 is a $(X - \{x*\}, W - \{x*\})$ multiplier and q_2 is a $(X, W - \{x*\})$ multiplier. By the inductive hypothesis we have that the expanded expressions $A_{q_1}(z), A_{q_2}(z)$ corresponding

to q_1, q_2 have values $p_1 = P_D[\mathcal{E}_{X-\{x*\},W-\{x*\}}]$ and $p_2 = P_D[\mathcal{E}_{X,W-\{x*\}}]$ for z^b. Thus $A_q(z^b) = A_{q_1}(z^b) - A_{q_2}(z^b) = p_1 - p_2 = P_D[\mathcal{E}_{X,W}]$. □

We will also denote $\mathcal{E}_{X,W}$ by $\mathcal{E}_p$ when a (X, W)-multiplier p is given. We refer to the events of the form $x_i = 1$ or $x_i = 0$ in the description of $\mathcal{E}_{X,W}$, where x_i is either a facility or an assignment variable, as *elementary events*.

So let us consider the constraint $\pi : \sum_{j \in C} x_{i'j} \leq U y_{i'}$ for some $i' \in Cheap$ and let q be some k-level (X, W)-multiplier for $k \leq n/10$ - the solutions in the support of E satisfy the constraint for costly facilities. By Lemma 4.2.4 it is enough to prove $\sum_{j \in C} P_E[(x_{i'j} = 1) \wedge \mathcal{E}_{X,W}] \leq U P_E[(y_{i'} = 1) \wedge \mathcal{E}_{X,W}]$. Without loss of generality we can assume that there is no facility opening variable of some costly facility or assignment variable to some costly facility in $X - W$ since in E whenever a costly facility is opened, all the inequalities are satisfied and thus z^b satisfies the inequality of the lifted relaxation corresponding to π and q. Likewise we can assume that there is no cheap facility opening variable in $X - W$, since such facilities are always opened and thus we can omit them from the relevant events without changing the values of the corresponding probabilities and, similarly, we can assume that there is no cheap facility variable in W, since in that case all the involved probabilities are 0. Finally, we further assume that the are no two assignment variables of the same client in $X - W$ as in this case the probabilities are again 0 since in each outcome of the experiment that defines E each client is assigned to exactly one facility.

We will distinguish between two cases of the experiment defining the distribution E. If event $\mathcal{E}_{X,W}$ is true in Case 1 then, on average, there are w^1_{ch} assignments to facility i' that are true. Thus, if we condition on Case 1, by linearity of expectation, the lhs of the inequality of the lifted relaxation has value equal to $w^1_{ch} P_E[\mathcal{E}_{X,W} \mid case\ 1]$ and the rhs has value $U P_E[\mathcal{E}_{X,W} \mid case\ 1]$. Similarly, if we condition on Case 2, the lhs of the inequality of the lifted relaxation has value equal to $w^2_{ch} P_E[\mathcal{E}_{X,W} \mid case\ 2]$ and the rhs has value $U P_E[\mathcal{E}_{X,W} \mid case\ 2]$. Thus it is enough to prove that $(w^2_{ch} - U) P_E[\mathcal{E}_{X,W} \mid case\ 2] P_E[case\ 2] \leq (U - w^1_{ch}) P_E[\mathcal{E}_{X,W} \mid case\ 1] P_E[case\ 1]$ for the number of levels we consider. Let $\mathcal{E}^*_{X,W}$ be the event that results from $\mathcal{E}_{X,W}$ if we ignore costly facility opening variables in W.

Lemma 4.2.5. *Let q be a k-level (X, W)-multiplier, $k \leq n/10$. Then* $P_E[\mathcal{E}^*_{X,W} \mid case\ 2] \leq e^2 P_E[\mathcal{E}^*_{X,W} \mid case\ 1]$.

Proof. Recall that by our previous assumptions X, W do not contain any facility opening variable, $X - W$ does not contain assignment variables to costly. Thus the variables that $\mathcal{E}^*_{X,W}$ requires to be equal to 1 are assignment to cheap facilities and the variables that it requires to be equal to 0 are assignment variables either to cheap or costly. Let $\mathcal{E}^*_{X,W} = ((x_{i_{a_1} j_{b_1}} = 1) \wedge \ldots \wedge (x_{i_{a_t} j_{b_t}} = 1) \wedge (x_{i_{a_{t+1}} j_{b_{t+1}}} = 0) \wedge \ldots \wedge (x_{i_{a_k} j_{b_k}} = 0))$. In order to bound the ratio $\frac{P_E[\mathcal{E}^*_{X,W} | case2]}{P_E[\mathcal{E}^*_{X,W} | case1]}$ we use that $P_E[\mathcal{E}^*_{X,W} \mid case\ i] = \prod_{l=1}^{l=t} P[x_{i_{a_l} j_{b_l}} = 1 \mid (\bigwedge_{z<l} (x_{i_{a_z} j_{b_z}} = 1) \wedge case\ i)] \prod_{l=t+1}^{l=k} P_E[(x_{i_{a_l} j_{b_l}} = 0) \mid \bigwedge_{z \leq t} (x_{i_{a_z} j_{b_z}} = 1) \bigwedge_{t+1 \leq z \leq l} (x_{i_{a_z} j_{b_z}} = 0) \wedge case\ i]$ and we bound the ratio of the corresponding products of the two cases.

The idea is that conditioning on at most k elementary events such as $x_{ij} = 1$ or $x_{ij} = 0$ doesn't affect much the probability of $x_{i^* j^*} = 1$ or $x_{i^* j^*} = 0$. Without loss of generality assume that in the experiment that defines E the balls are tossed in the order that they appear in $\mathcal{E}^*_{X,W}$. Let us first consider Case 1. In Case 1 the cheap facilities have $w^1_{ch} = n^3 - \Theta(n^2)$ slots so after we condition on the appropriate events - call the condition $Cond$ - we have that the empty slots of a cheap facility i^* are at least $w^1_{ch} - k$ and at most w^1_{ch}, and so the probability of $x_{i^* j^*}$, given the conditioning, is $\frac{w^1_{ch} - k}{|C| - k} \leq P_E[x_{i^* j^*} = 1 | \ Cond \wedge case\ 1] \leq$

$\frac{w^1_{ch}}{|C|-k}$ for a positive event $x_{i^*j^*}=1$. For a negative event $x_{i^*j^*}=0$ (here i^* is allowed to be costly) the probability is bounded as $\frac{|C|-w^1_{ch}-k}{|C|-k} \leq P_E[x_{i^*j^*}=0|\ Cond \wedge case\ 1] \leq \frac{|C|-w^1_{ch}+k}{|C|-k}$ if i^* is a cheap facility, or as $\frac{|C|-w_{co}-k}{|C|-k} \leq P_E[x_{i^*j^*}=0|\ Cond \wedge case\ 1] \leq \frac{|C|-w_{co}+k}{|C|-k}$ if i^* is a costly facility – the bounds follow from the fact that in the worst case for the outcome of throwing the k first balls, all balls end up in facilities other than i^*, and in the best case none of them ends up in facility i^*.

Now consider Case 2. In Case 2 the cheap facilities have $w^2_{ch}=n^3+o(1)$ slots so after we condition on the appropriate events – call again the condition $Cond$ - we have that the empty slots of a cheap facility i^* are at least $w^2_{ch}-k$ and at most w^2_{ch}, and so the probability of $x_{i^*j^*}=1$, given the conditioning, is $\frac{w^2_{ch}-k}{|C|} \leq P_E[x_{i^*j^*}=1|\ Cond \wedge case\ 2] \leq \frac{w^2_{ch}}{|C|-k}$ for a positive event $x_{i^*j^*}$. For a negative event $x_{i^*j^*}=0$ (here i^* is allowed to be costly) the probability is bounded as $\frac{|C|-w^2_{ch}-k}{|C|-k} \leq P_E[x_{i^*j^*}=0|\ Cond \wedge case\ 2] \leq \frac{|C|-w^2_{ch}+k}{|C|-k}$ if i^* is a cheap facility, or is equal to 1 if i^* is a costly facility, since no costly facility is opened in Case 2. The bounds follow from the same arguments as in Case 1.

So we have

$$\frac{P_E[\mathcal{E}^*_{X,W} \mid case\ 2]}{P_E[\mathcal{E}^*_{X,W} \mid case\ 1]} = \prod_{l=1}^{l=t} \frac{P_E[x_{i_{a_l}j_{b_l}}=1 \mid (\bigwedge_{z<l} x_{i_{a_z}j_{b_z}}=1 \wedge case\ 2)]}{P_E[x_{i_{a_l}j_{b_l}}=1 \mid (\bigwedge_{z<l} x_{i_{a_z}j_{b_z}}=1 \wedge case\ 1)]}\times$$

$$\prod_{l=t+1}^{l=k} \frac{P_E[x_{i_{a_l}j_{b_l}}=0 \mid \bigwedge_{z\leq t} x_{i_{a_z}j_{b_z}}=1 \bigwedge_{t+1\leq z\leq l} x_{i_{a_z}j_{b_z}}=0 \wedge case\ 2]}{P_E[x_{i_{a_l}j_{b_l}}=0 \mid \bigwedge_{z\leq t} x_{i_{a_z}j_{b_z}}=1 \bigwedge_{t+1\leq z\leq l} x_{i_{a_z}j_{b_z}}=0 \wedge case\ 1]} \leq$$

$$(\frac{\frac{w^2_{ch}}{|C|-k}}{\frac{w_{co}-k}{|C|-k}})^t(\frac{\frac{|C|-w^2_{ch}+k}{|C|-k}}{\frac{|C|-w_{co}-k}{|C|-k}})^{k-t} \leq$$

recalling that $|C|=n^4+1$, $w^2_{ch}=n^3+o(1)$, $w^1_{ch}=n^3-\Theta(n^2)$ and $w_{co}=n^3-o(1)$, the latter expression is bounded by

$$(1+2/n)^k \leq e^2$$

using that $\lim_{x\to\infty}(1+d/x)^x=e^d$. □

Now we turn to bound the ratio $\frac{P_E[\mathcal{E}_{X,W}|case\ 2]}{P_E[\mathcal{E}_{X,W}|case\ 1]}$ which is the actual ratio that we need to conclude the proof.

Lemma 4.2.6. *Let q be a k-level (X,W)-multiplier, $k\leq n/10$. Then $P_E[\mathcal{E}_{X,W} \mid case2] \leq 10P_E[\mathcal{E}_{X,W} \mid case1]$.*

Proof. Recall that $\mathcal{E}^*_{X,W}$ is the event that results from $\mathcal{E}_{X,W}$ if we ignore costly facility opening variables in W. If W contains t costly facility variables then $P_E[\mathcal{E}_{X,W} \mid case\ 1] \geq \frac{n-t}{n}P_E[\mathcal{E}^*_{X,W} \mid case\ 1]$ as in Case 1 exactly one costly facility is opened uniformly at random. So for $k\leq 10/n$ we have $P_E[\mathcal{E}_{X,W} \mid case\ 1] \geq \frac{9}{10}P_E[\mathcal{E}^*_{X,W} \mid case\ 1]$. As in Case 2 no costly facility is opened we have $P_E[\mathcal{E}_{X,W} \mid case\ 2] = P_E[\mathcal{E}^*_{X,W} \mid case\ 2]$. So by 4.2.5 we have $P_E[\mathcal{E}_{X,W} \mid case2] \leq 10P_E[\mathcal{E}_{X,W} \mid case1]$. □

By combining all the previous we have:
$\sum_{j \in C} P_E[x_{i'j} = 1 \wedge \mathcal{E}_{X,W}] =$
$\sum_{j \in C}(P_E[case\ 1]P_E[x_{i'j} = 1 \wedge \mathcal{E}_{X,W} \mid case\ 1] + P_E[case\ 2]P_E[x_{i'j} = 1 \wedge \mathcal{E}_{X,W} \mid case\ 2]) \leq$
$(U + 1/n)P_E[case\ 2]P_E[\mathcal{E}_{X,W} \mid case\ 2] + (U - n^2)P_E[case\ 1]P_E[\mathcal{E}_{X,W} \mid case\ 1])$

where the last inequality follows from the fact that $U + 1/n$ clients are assigned to a cheap facility in Case 2 while $(U - n^2)$ clients are assigned to it in Case 1. The last expression is equal to:

$U(P_E[case\ 2]P_E[\mathcal{E}_{X,W} \mid case\ 2] + P_E[case\ 1]P_E[\mathcal{E}_{X,W} \mid case\ 1])$
$-n^2 P_E[case\ 1]P_E[\mathcal{E}_{X,W} \mid case\ 1] + 1/nP_E[case\ 2]P_E[\mathcal{E}_{X,W} \mid case\ 2] =$
$UP_E[\mathcal{E}_{X,W}] - n^2\frac{10}{n}P_E[\mathcal{E}_{X,W} \mid case\ 1] + 1/n(1 - \frac{10}{n})P_E[\mathcal{E}_{X,W} \mid case\ 2] \leq$

which by Lemma 4.2.6 is at most

$UP_E[\mathcal{E}_{X,W}] - n^2\frac{10}{n}P_E[\mathcal{E}_{X,W} \mid case\ 1] + 10/n(1 - \frac{10}{n})P_E[\mathcal{E}_{X,W} \mid case\ 1] \leq$
$UP_E[\mathcal{E}_{X,W}]$

Thus we get the main result of this section which is stated in the following Theorem.

Theorem 4.2.1. *There is a family of instances $\mathcal{I}$ of uniform* CFL*, such that for every sufficiently large n, there is an instance in the family with $2n$ facilities and n^4+1 clients for which the relaxations obtained from (LP-classic) at $\Omega(n)$ levels of the Sherali-Adams hierarchy have an integrality gap of $\Omega(n)$.*

We remark that even though in the proofs of Theorems 4.1.1 and 4.2.1 we show the survival of the same fractional solution s, the lifted solutions obtained in our proofs are in generally different. Consider the situation at level $t \geq 1$ of both hierarchies. The product $Y_{\{y_i,y_i'\}} = y_i y_{i'}$ for any two distinct costly facilities i and i', is zero as the random experiment we define in the SA proof opens exactly one costly facility. In contrast, the associated entry of the protection matrix we construct for the LS proof is in general nonzero, because as long as none of the two facilities have been touched as Type 2, their product is a small but nonzero quantity.

4.2.3 Robustness of the SA gap

In this section we explain how adding simple valid inequalities to (LP-classic) does not affect our arguments on the SA hierarchy.

As an example we address the valid inequality (4.22), which was discussed also in the context of LS in Subsection 4.1.3. Of course, this inequality is rendered useless by slight modifications to the instance and the bad solution. Identifying "areas" of a fractional solution where the demand exceeds the available capacity seems impossible for polynomially sized relaxations without some yet unknown form of preprocessing. In fact, part of the motivation behind Theorem 4.2.1 was to demonstrate that the SA hierarchy is inadequate for such preprocessing purposes.

We modify the family $\mathcal{I}$ as follows. To every instance with n cheap and n costly facilities and $Un+1$ clients we add a single *dummy* facility a, all with 0 opening cost, on some point at distance 1 from the rest. The underlying metric space is a line, and thus we have an instance of the so called *facility location on a line*. We define bad solution s^a to agree with s on the old variables, additionally we set $y_a = 1$ and $x_{aj} = 0$ for all clients j. Inequality (4.22) is obviously satisfied by s^a.

In the definition of distribution E everything remains except for the new facility a: a is always opened and is not assigned any demand. Obviously the inequality (4.22) is satisfied and thus Theorem 4.2.1 holds for the strengthened relaxation as well.

We will also give a second, perhaps more generic proof that relies on basic properties of the SA system. This proof was proposed by an anonymous reviewer.

Lemma 4.2.7. *Consider some polytope $P \subseteq [0,1]^n$, and $\bar{x} \in P$ with $\bar{x}_i = b$, where $b \in \{0,1\}$. Let $P' \subseteq [0,1]^{n-1}$ be a polytope defined by all constraints*

$$\sum_{j:j\neq i} \alpha_i x_j + \alpha_i \cdot b \leq b$$

where $\alpha^T x \leq b$ is a constraint of P. Let also x' be the projection of $\bar{x}$ onto coordinates $j \neq i$ Then $\bar{x} \in S^k(P)$ if and only if $x' \in S^k(P')$.

Proof. Let z' be a solution of the lifted P' that projects to x'. Consider the solution $\overline{z}$ for the lifted P obtained by setting for all sets $I \not\ni i$ $\overline{z}_{I\cup\{i\}} = b \cdot z'_I$, and $\overline{z}_I = z'_I$ - $\overline{z}$ projects to $\overline{x}$. By Lemma 4.2.3 it is immediate that $\overline{z}$ is feasible for the lifted P if and only if z' is feasible for the lifted P'. □

Given the above facts, it is now easy to conclude the argument.

Theorem 4.2.2. *Let P be the polytope defined on instance I_a of family $\mathcal{I}_a$ by (LP-classic) augmented by constraint (4.22). For sufficiently large n and $k = \Omega(n)$, vector $s^a \in \mathrm{SA}^k(P)$.*

Proof. Let P_1 be the polytope defined by (LP-classic) where variable y_a has been replaced by the constant 1 and for all j x_{aj} has been replaced by the constant 0. Observe that the system defining P_1 is simply (LP-classic) on instance I augmented by the redundant inequality $0 \leq 1$. By Theorem 4.2.1, $s \in SA^k(P_1)$.

Let P' be the polytope whose description results from the system defining P after replacing y_a and x_{aj} for all j, with their values in s^a, namely, 1 and 0 respectively. By an inductive application of Lemma 4.2.7, $s^a \in \mathrm{SA}^k(P)$ if and only if $s \in \mathrm{SA}^k(P')$. Constraint (4.22) appears in the description of P' as

$$\sum_{i\neq a} y_i + 1 \geq \lceil \frac{n^4+1}{n^3} \rceil \Leftrightarrow \sum_{i\neq a} y_i \geq n \tag{4.23}$$

Inequality (4.23) is redundant for P' as it results from the summation of the $2n$ capacity constraints for the cheap and costly facilities. By Lemma 4.2.2, $\mathrm{SA}^k(P') = \mathrm{SA}^k(P_1)$ and since $s \in \mathrm{SA}^k(P_1)$ it follows that $s^a \in \mathrm{SA}^k(P)$. □

4.2.4 SA gaps for approximate LBFL

In this section we shall prove, similarly to the previous section, that the integrality gap of the classic relaxation for LBFL obtained after a number of SA levels that is linear in the number n of facilities of the instance, is not bounded by any constant, while after at most n levels the relaxation obtained is a description of the integer LBFL polytope.

The instance that we will consider is the following: there are n facilities, the lower bounds are uniform $B = n^3$ and there is a number of $n(B-1)$ clients of unit demand. The metric space here is more intriguing than the one for the CFL case. Consider a regular $(n-1)$-dimensional simplex with edge length 1. On each of the n vertices of the simplex a facility along with some $B-1$ clients are located. All opening costs are 0. Clearly every integer solution has a cost of at least $B-1$ since we can open at most $n-1$ of the facilities, and so at least $B-1$ clients will have to be assigned to some facility other than the close one, incurring a connection cost of $\Theta(B)$. We call a client j that is located on the same vertex with facility i an *exclusive* client of i. We denote by $Exclusive(i)$ the set of clients that are exclusive to facility i. We will show that the following bad solution s is feasible for the relaxation obtained after $\Omega(n)$ levels of the SA hierarchy. For all $i \in F$, $y_i = 1 - n^{-2}$; for a client $j \in C$, $x_{ij} = 1 - n^{-2}$, if $j \in Exclusive(i)$, and $x_{i'j} = \frac{n^{-2}}{n-1}$ for $j \notin Exclusive(i')$. Solution s incurs a cost of $o(B)$.

As in the case of CFL, we will give a distribution E with the help of which we shall define the solution z for the lifted relaxation obtained for $k = n/10$ levels of the SA lift-and-project method applied to the classic relaxation of LBFL - z is such that it projects to s. With probability $1 - 1/n$ all the facilities are opened and each one is assigned its exclusive clients. Call this *Case* 1. This case slightly violates the lower bound constraints. With the remaining probability $1/n$ do the following - call this *Case* 2: select one facility i' uniformly at random. Open all the other facilities and assign to each one its exclusive clients. Randomly assign the exclusive clients of i' to the open facilities so that each is assigned at least 1 of those clients. It is easy to see that bad solution s is indeed the average (y, x) vector induced by distribution E: any facility i is opened for a $1-1/n+1/n-1/n^2 = 1-n^{-2} = y_i$ fraction of the time, a client j that is exclusive to i is assigned to it a fraction $1-1/n+1/n-1/n^2 = 1-n^{-2} = x_{ij}$ of the time while it is assigned to any other facility i' a fraction $\frac{n^{-2}}{n-1} = x_{i'j}$ of the time. We define the solution z of the lifted polytope as follows: for any set $\mathcal{E}$ of variables, $z_{\mathcal{E}} = P_E[$ all variables in $\mathcal{E}$ have value $1]$.

Observe that each outcome of the random experiment that defines E satisfies all the inequalities of LP-classic except the lower bound constraints (2.16). The latter inequality is violated in Case 1 but is satisfied with positive slack in Case 2 and it is also satisfied by the average vector s. But for a constraint of the lifted relaxation obtained from the multiplication of such a constraint $\pi : \sum_{j \in C} x_{i'j} \geq B y_{i'}$ with a k-level (X, W)-multiplier p, which we denote as $p_{(X,W)}$ from now on, depending on the elementary events in the corresponding event $\mathcal{E}_p$, the slackness of Case 2 might decrease. We prove that, even in the worst case for the choice of p, the lifted inequality is still satisfied by z for the number of levels we consider.

For an expression T of probabilities we denote for convenience by $T \mid Cond$ the expression where all the probabilities of T are conditioned by the condition $Cond$. For an expression L containing (X, W)-multipliers, we denote by $[L]$ the expression on probabilities that results by replacing any (X, W)-multiplier p in L by $P_E[\mathcal{E}_{(X,W)}]$. Observe that, since we defined the lifted SA solution z via the distribution E, this substitution is by Lemma 4.2.4 valid for our calculations. Below we calculate the violation V of the constraint of the lifted polytope, if we condition on Case 1 - this is how much larger is the rhs from the lhs of the inequality of the lifted relaxation in Case 1. For convenience, we slightly abuse notation and we denote $Z \cup \{z_i\}$ by $Z \cup z_i$.

$V := [p_{(X,W)}(By_{i'} - \sum_{j\in C} x_{i'j})] \mid$ $case$ $1 = [Bp_{(X\cup y_{i'},W)} - \sum_{j\in C} p_{(X\cup x_{i'j},W)}] \mid$ $case$ $1 = BP_E[\mathcal{E}_p \mid case\ 1] - (B-1)P_E[\mathcal{E}_p \mid case\ 1] = P_E[\mathcal{E}_p \mid case\ 1]$

The second equality follows from Lemma 4.2.4. The last equality follows from linearity of expectation and by the fact that exactly $B-1$ clients are assigned to i' in Case 1, and by the fact that i' is always opened in that case.

So we have a violation of $V = P_E[\mathcal{E}_p \mid \text{case } 1]$ in Case 1. Conditioning on Case 2, similarly, we have a slackness of at least $S = (n^2-2)P_E[\mathcal{E}_p \mid case\ 2]$ - this is how much smaller is the rhs from the lhs in Case 2:

$S := [(\sum_{j\in C} x_{i'j} - By_{i'})p_{(X,W)}] \mid$ $case$ $2 = [\sum_{j\in C} p_{(X\cup x_{i'j},W)} - Bp_{(X\cup y_{i'},W)}] \mid$ $case$ $2 =$
$\frac{n-1}{n}((B-1)(1+\frac{1}{n-1})P_E[\mathcal{E}_p \mid case\ 2] - BP_E[\mathcal{E}_p \mid case\ 2]) =$
$(\frac{B}{n-1} - 1)P_E[\mathcal{E}_p \mid case\ 2] \geq (n^2-1)P_E[\mathcal{E}_p \mid case\ 2]$

The second equality above follows from the fact that in Case 2 facility i' is opened a fraction $\frac{n-1}{n}$ of the time and when it does it is assigned a number of $(B-1)(1+\frac{1}{n-1})$ clients. What is left is to relate $P_E[\mathcal{E}_p \mid case1]$ to $P_E[\mathcal{E}_p \mid case\ 2]$.

Lemma 4.2.8. *For any k-level (X,W) multiplier p, $k \leq \frac{n}{10}$, $P_E[\mathcal{E}_p \mid case\ 2] \geq \frac{9}{10}P_E[\mathcal{E}_p \mid case\ 1]$.*

Proof. Observe that we can ignore all elementary events $(x_{ij}=0) \in \mathcal{E}$ for $j \in Exclusive(i)$ since this never happens in Case 1 and the inequality of the Lemma trivially holds. We can also ignore $y_i = 0 \in \mathcal{E}$ for the same reasons, as well as $x_{ij} = 1 \in \mathcal{E}$ for $j \notin Exclusive(i)$. So we only have to consider elementary events of type $y_i = 1$, $x_{ij} = 1 \mid j \in Exclusive(i)$ and $x_{ij} = 0 \mid j \notin Exclusive(i)$. Each one of those events is always true in Case 1. In Case 2 each such elementary event eliminates a fraction of at most $1/n$ of the measure of the probability space in Case 2: the negation of such elementary event has a probability of at most $1/n$ and thus the probability of their intersection is lower bounded by 1 minus the sum of the probabilities of the negations of the elementary events, i.e. $P_E[\mathcal{E}_p \mid case\ 2] \geq 1 - (\sum_{t\in X} P_E[t=0 \mid case\ 2] + \sum_{t\in W} P_E[t=1 \mid case\ 2])$. So, for the number of at most $n/10$ levels we consider and thus for at most that many elementary events in $\mathcal{E}_p$, $P_E[\mathcal{E}_p \mid case\ 2] \geq \frac{9}{10}P_E[\mathcal{E}_p \mid case\ 1]$. □

By putting all the above together we can now prove the satisfaction of the constraint of the lifted relaxation:

$$[p_{(X,W)}(\sum_{j\in C} x_{i'j} - By_{i'})] =$$

$$P_E[case\ 1]([p_{(X,W)}(\sum_{j\in C} x_{i'j} - By_{i'})] \mid case\ 1) + P_E[case\ 2]([p_{(X,W)}(\sum_{j\in C} x_{i'j} - By_{i'})] \mid case\ 2) =$$

$$P_E[case\ 2]S - P_E[case\ 1]V = \frac{S}{n} - (1-1/n)V \geq$$

$$\frac{n^2-1}{n} P_E[\mathcal{E}_p \mid case\ 2] - (1-1/n) P_E[\mathcal{E}_p \mid case1] \geq$$

$$\frac{n^2-1}{n}\frac{9}{10} P_E[\mathcal{E}_p \mid case\ 1] - (1-1/n) P_E[\mathcal{E}_p \mid case1] \geq 0$$

where the second inequality follows from Lemma 4.2.8. Thus we have proved the following theorem.

Theorem 4.2.3. *The* LBFL *relaxation obtained from (LP-classic) at* $\Omega(n)$ *levels of the Sherali-Adams hierarchy has an integrality of* $\Omega(n)$*, where* n *is the number of facilities of the instance.*

4.3 Fooling the submodular inequalities for CFL

In this section we show that the relaxation obtained from (LP-classic) with the addition of the submodular inequalities proposed by Aardal et al. [1] has a gap not bounded by any constant. Levi et al. [46] conjectured that the addition of a subset of those inequalities, called the flow-cover inequalities, to the classic relaxation would improve the integrality gap to constant. We note that it is not known how to separate in polynomial time even the flow-cover inequalities. We wish to clarify, that the inequalities introduced by Carr et al. [20] and extended by Carnes and Shmoys [73] are referred to also as "flow-cover inequalities". Our proof does not apply to the latter.

We proceed to define first the flow-cover inequalities and then generalize the definition to the submodular ones. Consider the general case where facility i has capacity u_i and client j has demand d_j. For a set J of clients, we denote their total demand by $d(J) = \sum_{j\in J} d_j$. Let $J \subseteq C$ be a set of clients, let $I \subseteq F$ be a set of facilities, and let $J_i \subseteq J$ be a set of clients for each facility $i \in I$. Given a facility i, we denote the *effective capacity* of i with respect to J_i by $\bar{u}_i = \min\{u_i, d(J_i)\}$. I is a *cover* with respect to J if $\sum_{i\in I} \bar{u}_i = d(J) + \lambda$ with $\lambda > 0$. λ is called the *excess capacity*. Let $(x)^+ = \max\{x, 0\}$. In the case where $J_i = J$ for all $i \in I$ the following inequalities called *flow-cover* inequalities were introduced for CFL in [1].

$$\sum_{i\in I}\sum_{j\in J} d_j x_{ij} + \sum_{i\in I} (u_i - \lambda)^+ (1 - y_i) \leq d(J)$$

If $\max_{i\in I}(\bar{u}_i) > \lambda$, the following inequalities, called the *effective capacity inequalities* are valid and strengthen the flow-cover inequalities [1]. Note that we no longer assume that $J_i = J$.

$$\sum_{i\in I}\sum_{j\in J_i} d_j x_{ij} + \sum_{i\in I} (\bar{u}_i - \lambda)^+ (1 - y_i) \leq d(J)$$

The submodular inequalities introduced in [1] are even stronger than the effective capacity inequalities. From now on we limit our discussion to uniform CFL with *all clients having unit demands.*

Choose a subset $J \subseteq C$ of clients, and let $I \subseteq F$ be a subset of facilities. For each facility

$i \in I$ choose a subset $J_i \subseteq J$. Consider a 3-level network G with a source s, a set of nodes corresponding to the facilities, a set of nodes corresponding to the clients and a sink t. The source s is connected by an edge of capacity $\min\{U, |J_i|\}$ to each facility node i. That node is connected by an edge of unit capacity to each node corresponding to client $j, j \in J_i$. Each node corresponding to some client is connected by an edge of unit capacity to the sink t.

Define $f(I)$ as the maximum s-t flow value in G. Define $f(I \setminus \{i\})$ as the maximum flow when facility i is closed, i.e., when the capacity of edge (s, i) is set to zero. The difference in maximum flow when all facilities in I are open, and when all facilities except facility i are open, is called the *increment* function and is defined as $\rho_i(I \setminus \{i\}) = f(I) - f(I \setminus \{i\})$.

For any choice of $I \subseteq F$, $J \subseteq C$, and $J_i \subseteq J$, for all i, the following inequalities, called *the submodular inequalities,* are valid for CFL [1]. The name reflects the fact that the function $f(I)$ is submodular.

$$\sum_{i\in I}\sum_{j\in J_i} x_{ij} + \sum_{i\in I} \rho_i(I \setminus \{i\})(1 - y_i) \leq f(I)$$

The intuitive explanation for an integer solution is the following: if some set $S \subseteq I$ of facilities is closed then the loss in the total flow through G is at least $\sum_{i\in S} \rho_i(I \setminus \{i\})$. Thus the total assignments wrt the selected client sets J_i's cannot be greater than the maximum possible flow $f(I)$ minus the flow loss due to the closed facility of S which is at least $\sum_{i\in S} \rho_i(I \setminus \{i\})$. The proof of the following theorem uses some of the ideas we introduced earlier for Theorem 4.2.1.

Consider an inequality π and suppose that there is some costly facility i_b not in the support of π. We construct a solution $s_\pi = (y', x')$ by slightly modifying bad solution s: set $y'_{i_b} = 1 - \sum_{i\in Costly-\{i_b\}} y_i$ and leave all other variables the same as in the original bad solution s. We say that facility i_b *takes the blame for* π. When π is implied from the context, we simply say that i_b *takes the blame*. We will prove that s_π can be obtained as a convex combination E_d of a set of integer solutions satisfying $\sum_{i\in Costly} y_i = 1$.

Before proving the main Theorem of this section, we will first provide a Lemma which essentially shows how to fool inequalities that are local in the sense that not all the y variables of the costly facilities are in the support of π. The goal is to show that bad solution s is feasible for the LP obtained after adding to the classic relaxation the submodular inequalities. In the following Lemma we show that we can obtain s_π as the average vector of a distribution over feasible integer solutions E_d.

Lemma 4.3.1. *Solution $s_{\pi,z}$ is the expected (y, x) vector obtained from a E_d over feasible integer solutions.*

Proof. We describe a probabilistic experiment which induces the distribution E_d over integer solutions satisfying $\sum_{i\in Costly} y_i = 1$. Let i_b be the facility that takes the blame for π. We begin by defining some parameters for the integer solutions that will make up the support of the distribution E_d.

Let $w_{i_{co}} = \frac{\sum_{j\in C} x'_{i_{co}j}}{y'_{i_{co}}}$ be the number of clients assigned to facility i_{co} in the integer solutions in the support of E_d, when facility $i_{co} \in Costly$ is opened. To simplify the presentation let us assume for now that $w_{i_{co}}$ and the values we subsequently define are integers (we discuss at the end of the proof how to handle fractional values). Let $w^1_{i_{ch}} = \frac{|C| - w_{i_b}}{|Cheap|}$ be the number of clients assigned to facility $i_{ch} \in Cheap$ when i_b is the only opened costly facility. Likewise, for any costly facility $i_{co} \neq i_b$, let $w^2_{i_{ch}} = \frac{|C|-w_{i_{co}}}{|Cheap|}$ be the number of clients assigned to facility $i_{ch} \in Cheap$ in each integer solution in E_d where facility i_{co} is the only opened costly facility. Observe that all the defined values are less than U. The following procedure produces the distribution E_d.

Pick costly facility i_{co} with probability $y'_{i_{co}}$. If $i_{co} = i_b$ ($i_{co} \neq i_b$) then consider n bins corresponding to the n cheap facilities each one having w^1_{ch} (w^2_{ch}) slots and 1 bin corresponding to i_{co} having $w_{i_{co}}$ slots. Randomly distribute $|C|$ balls to the slots of the $n+1$ bins, with exactly one ball in each slot. Note that the above experiment induces a distribution over feasible integer solutions satisfying $\sum_{i \in Costly} y_i = 1$ since all the defined bin capacities are less than U and every client is assigned to exactly one opened facility in each outcome and exactly 1 costly facility is opened. It remains to show that the expected (y, x) vector with respect to E_d is solution $s_{\pi,z}$.

We argue now that s_π is the convex combination induced by E_d. The cheap facilities are always open, and the costly are open a fraction of the time that is equal to the value of their corresponding y' variable. The expected demand assigned to each $i_{co} \in Costly$ is $y'_{i_{co}} w_{i_{co}}$ which is the total demand assigned to i_{co} by $s_{\pi,z}$. Since the clients have the same probability of being tossed in the bin corresponding to i_{co}, the expected assignment of each client j to i_{co} is the same as in s_π.

Similarly, we can prove that the expected assignments to the cheap facilities are as required. Observe that in every outcome of the experiment the demand not assigned to costly facilities is exactly the demand assigned to cheap facilities. Since we have proved that the expected assignments to the costly facilities are those of the bad solution s_π, by linearity of expectation we get that the total assignments to all cheap facilities are $\sum_{i \in Cheap} \sum_j x'_{ij}$ (the total assignment of each client adds up to 1). By the symmetric way the cheap facilities are handled in the experiment, we have that the total expected demand assigned to each $i \in Cheap$ is $\sum_{j \in C} x'_{ij}$. Moreover, by the symmetric way the clients are assigned to i in the experiment, we get that the expected assignment of each $j \in C$ to i is $\frac{\sum_j x'_{ij}}{|C|} = x'_{ij}$.

To handle the case where the $w_{i_{co}}, w^1_{i_{ch}}, w^2_{i_{ch}}$ are not integers (which is actually always the case), we do the following: each time costly facility i_b ($i_{co} \neq i_b$) is picked, we set the number of slots of the corresponding bin to $\lfloor w_{i_b} \rfloor$ ($\lfloor w_{i_{co}} \rfloor$) with probability $1 - (w_{i_b} - \lfloor w_{i_b} \rfloor)$ ($1 - (w_{i_{co}} - \lfloor w_{i_{co}} \rfloor)$), otherwise set the slots to $\lceil w_{i_b} \rceil (\lceil w_{i_{co}} \rceil)$; this ensures that the expected number of slots is w_{i_b} ($w_{i_{co}}$). The same rationale applies to the remaining cases of the construction. If the number of slots of i_b (i_{co}) is set to $\lfloor w_{i_b} \rfloor$ ($\lfloor w_{i_{co}} \rfloor$), then we pick some $n(\frac{|C|-\lfloor w_{i_b} \rfloor}{n} - \lfloor(\frac{|C|-\lfloor w_{i_b} \rfloor}{n})\rfloor)$ ($n(\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n} - \lfloor(\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n})\rfloor)$) cheap facilities at random and set their corresponding number of slots to $\lceil\frac{|C|-\lfloor w_{i_b} \rfloor}{n}\rceil$ ($\lceil\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n}\rceil$) and the number of slots of the rest of the cheap facilities to $\lfloor\frac{|C|-\lfloor w_{i_b} \rfloor}{n}\rfloor(\lfloor\frac{|C|-\lfloor w_{i_{co}} \rfloor}{n}\rfloor)$. Otherwise, pick some $n(\frac{|C|-\lceil w_{i_b} \rceil}{n} - \lfloor(\frac{|C|-\lceil w_{i_b} \rceil}{n})\rfloor)$ ($n(\frac{|C|-\lceil w_{i_{co}} \rceil}{n} - \lfloor(\frac{|C|-\lceil w_{i_{co}} \rceil}{n})\rfloor)$) cheap facilities at random and set their corresponding number of slots to $\lceil\frac{|C|-\lceil w_{i_b} \rceil}{n}\rceil$ ($\lceil\frac{|C|-\lceil w_{i_{co}} \rceil}{n}\rceil$) and the number of slots of the rest to $\lfloor\frac{|C|-\lceil w_{i_b} \rceil}{n}\rfloor(\lfloor\frac{|C|-\lceil w_{i_{co}} \rceil}{n}\rfloor)$. Once again the capacities are respected since $w_{i_{co}} = \Theta(n)$ (if $i_{co} = i_b$), or $w_{i_{co}} = \Theta(n^3)$ (if $i_{co} \neq i_b$). In every case the expected number of slots per facility is the same as in the initial description of the experiment where we assumed that $w_{i_{co}}, w^1_{i_{ch}}, w^2_{i_{ch}}$ are integers.

□

Now we are ready to state and prove the main result of this section.

Theorem 4.3.1. *There is a family of instances $\mathcal{I}$ of uniform* CFL*, such that for every sufficiently large n, there is an instance in the family with $2n$ facilities and n^4+1 clients for which the integrality gap of the relaxation produced from (LP-classic) with the addition of the submodular inequalities is $\Omega(n)$.*

Proof. Fix n. Consider the resulting instance $\mathcal{I}(n)$ from the family $\mathcal{I}$ and the bad solution s that we used in Theorem 4.2.1 for the SA result. To prove that s is feasible for the classic relaxation strengthened by the submodular inequalities we take the idea of fooling local constraints a little further: either the constraint is local enough that we can use the ideas from our previous proofs (define s' that is a convex combination of integer solutions and the values of the variables in the support of the constraint agree with s), or we can define another instance $\mathcal{I}'(n)$ and solution s' for which the inequality in question is true with respect to s' and again s' has the same visible part as s with respect to the constraint. Note that our arguments include two different instances as opposed to all our other proofs so far.

Consider the submodular inequality π for some I, J and some selection of J_i's. If not all the costly facilities appear in the constraint the proof is similar to that of Lemma 4.3.1. If at least n assignment variables to cheap facilities do not appear in π we do the following: we add one more facility b to the instance. We construct a solution s' for the new instance $\mathcal{I}'(n)$ as follows. We transfer the demand corresponding to the missing assignments of the cheap to b, and we set $y'_b = 1$. Observe that π is valid for $\mathcal{I}'(n)$. Now we can show that s' is a convex combination of integer solutions, again similarly to the proof of Lemma 4.3.1. Again, we call facility b *active* if it is assigned some nonzero demand. Facility b will be open 100% of the time but it will be active only when no costly facilities are open, i.e., a fraction $1-\sum_{i\in Costly} y'_i$ of the time. When it is active, it is assigned $\frac{\sum_j x'_{bj}}{1-\sum_{i\in Costly} y'_i}$ amount of demand. For simplicity, assume that this quantity is an integer. If it is not, we can take the remedial action described in detail in the proof of Lemma 4.3.1.

In the fractional solution s', b is assigned a total demand of at least $1-1/n^2$, therefore, in each outcome of the random experiment in which b is active, it will be assigned at least one client. Thus, in each such outcome we obtain a feasible integer solution. By the convex combination produced, the inequality is satisfied by s'. Thus the same inequality for the original instance is satisfied by s, since the exact same values of variables are involved in both cases.

Now consider the case where less than n assignments to cheap facilities are missing from π. We will show that it cannot be the case that all y_i variables of costly facilities appear in the constraint as well. Consider the quantity $\rho_i(I\setminus\{i\})$ for some costly facility i. If $\rho_i(I\setminus\{i\})>0$, then J_i is not empty. We will show that the set of nodes $(Cheap\cap I)\cup\{i\}$ in G has enough incident edges so that the flow originating from them is equal to the total client demand $|J|$ in G. We first give some properties of graph G.

Claim 4.3.1. *If less than n assignments to cheap facilities are missing from π, then $(Cheap\cap I)=Cheap$ and $J=C$.*

Proof of Claim. To see that $(Cheap\cap I)=Cheap$, notice that if a cheap facility is missing from I, at least $|C|=n^4+1$ assignment variables will be missing from π, a contradiction. For the second part of the claim, if a client j is missing from J, then all the corresponding n edges that would connect j to a cheap facility cannot be in G. Therefore at least n assignment-to-cheap variables are missing from π, a contradiction. The proof of the claim is complete.

We return to proving that $Cheap \cup \{i\}$ has enough incident edges so that the flow originating from them is equal to the total client demand $|C|$ in G. "Assign" one client $j \in J_i$ to facility i and for the remaining $|C| - 1$ clients do the following: assign each client j' involved in the set of variables of assignments-to-cheap that are missing from π to a cheap facility i' such that $j' \in J_{i'}$. There is always such a cheap facility i' since the missing edges from the client-nodes in G to the cheap-facility nodes are less than n. Assign the remaining clients arbitrarily to the cheap facilities respecting the capacities, since all the edges from cheap to those clients are included in the network. Thus it must be the case that $\rho_{i'}(I \setminus \{i'\}) = 0$ for any other costly facility $i' \neq i$. Since the $y_{i'}$ variable of such a facility i' has 0 coefficient in the constraint, it can take the blame and the proof is similar to that of Lemma 4.3.1.

□

5. A NEW METHOD FOR LOWER BOUNDING THE SIZE OF EXTENDED FORMULATIONS

In this chapter we will present a new method for lower bounding the extension complexity. We start with some basic extended formulation definitions in Subsection 5.1, then in Subsection 5.2 we prove the main contribution which is also the central idea of this chapter: strong extended formulations with a specific set of variables and specific encodings of solutions have enough expressive strength to simulate any extended formulation without increasing its size. This idea leads to a method for lower bounding the combinatorial parameter of the *extension complexity,* i.e., the minimum size of an extended formulation of a polytope since we can focus on these particular formulations. The method also yields a characterization of the extension complexity in graph-theoretic terms which is given in Subsection 5.4. In Subsection 5.5, as a warm up, we give an example of lower bounding the size of a formulation with known variables and encoding: we show exponential lower bounds on the size of natural formulations for approximate CFL – we do not consider extended formulations here. Then we conclude this chapter by giving an example of lower bounding a specific type of extended formulation for mixed integer programs in Subsection 5.6. While we left as an open question whether this type captures all the extended formulations as we show in the fully integer case, our result is strong enough to imply optimal SA gaps *regardless* of the initial relaxation as long as it has subexponential size and uses the classic encoding.

5.1 Preliminaries on Extended Formulations

Given a polyhedron $K(x,y) = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^{d_y} \mid Ax + By \leq b\}$ the *projection to the x-space* is defined as $\{x \in \mathbb{R}^d \mid \exists y \in \mathbb{R}^{d_y}\colon Ax + By \leq b\}$, denoted as $\mathrm{proj_x}(K(x,y))$. An *extended formulation* of a polyhedron $P(x) \subseteq \mathbb{R}^d$ is a linear system $K(x,y) = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^{d_y} \mid Ax + By \leq b\}$ such that $\mathrm{proj_x}(K(x,y)) = P(x)$. The *size* of a polyhedron $P(x)$ is the minimum number of inequalities in its halfspace description. The *extension complexity* of $P(x)$ is the minimum size of an extended formulation of $P(x)$.

We define now ρ-approximate formulations as in [15]. Given a combinatorial optimization problem $T(S,f)$, a *linear encoding* of T is a pair (L,O) where $L \subseteq \{0,1\}^*$ is the set of encodings of *feasible solutions* to the problem and $O \subset \mathbb{R}^*$ is the set of encodings of the *admissible objective functions.* An instance of the linear encoding is a pair (d,w) where d is a positive integer defining the dimension of the instance and $w \subseteq O \cap \mathbb{R}^d$ is the set of admissible cost functions for instances of dimension d. Solving the instance (d,w) means finding $x \in L \cap \{0,1\}^d$ such that $w^T x$ is either maximum or minimum, according to the type of problem T. Let $P = \mathrm{conv}(\{x \in \{0,1\}^d \mid x \in L\})$ be the corresponding 0-1 polytope of dimension d. Given a linear encoding (L,O) of a maximization problem, the corresponding polytope P, and $\rho \geq 1$, a *ρ-approximate extended formulation of P* is an extended relaxation $Ax + By \leq b$ of P with $x \in \mathbb{R}^d, y \in \mathbb{R}^{d_y}$ such that

$$\max\{w^T x \mid Ax + By \leq b\} \geq \max\{w^T x \mid x \in P\} \qquad \text{for all } w \in \mathbb{R}^d \text{ and}$$
$$\max\{w^T x \mid Ax + By \leq b\} \leq \rho \max\{w^T x \mid x \in P\} \qquad \text{for all } w \in O \cap \mathbb{R}^d.$$

For a minimization problem, we require

$$\min\{w^T x \mid Ax + By \leq b\} \leq \min\{w^T x \mid x \in P\} \qquad \text{for all } w \in \mathbb{R}^d \text{ and}$$
$$\min\{w^T x \mid Ax + By \leq b\} \geq \rho^{-1} \min\{w^T x \mid x \in P\} \qquad \text{for all } w \in O \cap \mathbb{R}^d.$$

The *ρ-approximate extension complexity* of 0-1 integer polytope $P(x) \subseteq [0,1]^d$ is the minimum size of a ρ-approximate extended formulation of P.

One can prove a bound on the extension complexity of polytope P via an "LP-reduction" from a polytope P' for which a lower bound on $xc(P')$ is known. Such reductions, which were initiated in [66], prove that the extension complexity of polytope P is at least as high as the complexity of polytope P', usually by showing that P' is a face of P and thus a formulation of P gives a formulation of P' of the same size. In general we seek to prove that P is an *extension* of P', more formally:

Definition 5.1.1. *Let $P' \subseteq \mathbb{R}^d$ be a polytope. A polytope $P \subseteq \mathbb{R}^n$ is an extension of P', if there exists a linear map $\pi : \mathbb{R}^n \to \mathbb{R}^d$ so that $\pi(P) = P'$.*

We turn now to define two generic extended formulations that will play a central role in this chapter.

Definition 5.1.2. *Given a 0-1 integer polytope $P(x) \subseteq [0,1]^d$, a product formulation $D(z)$ of $P(x)$ is an extended formulation $D(z)$ of $P(x)$, where $z \in \mathbb{R}^{2^d-1}$ and for every nonempty subset $\mathcal{E} \subseteq \{x_1, x_2, \ldots, x_d\}$ of the original variables, we have a variable $z_{\mathcal{E}}$, (where $z_{\{x_i\}}$ denotes x_i, $i = 1, \ldots, d$). For any feasible integer solution $x^s \in P(x) \cap \{0,1\}^d$ the vector z^s, whose components are defined as $z^s_{\mathcal{E}} = 1$ iff all variables in $\mathcal{E}$ have value 1 in x^s and $z^s_{\mathcal{E}} = 0$ otherwise, is feasible for any product formulation $D(z)$ of $P(x)$. We will refer to z^s as the encoding of the feasible integer solution x^s in the product variables.*

Another type of extended formulations that we shall consider is that of *distributional formulations*. In distributional formulations we have one variable d_s for each solution $x^s \in P(x) \cap \{0,1\}^d$. The encoding of a solution $x^{s'}$ is quite natural: $d_s = 1$ if $s = s'$, otherwise $d_s = 0$. Intuitively the feasible vectors of such formulations represent distributions over feasible integer solutions, hence the name. Note that a trivial (large) distributional exact formulation of $P(x)$ is $\sum_s d_s = 1, d_s \geq 0 \; \forall s$ and the projection $x = \sum_s x^s d^s$. More formally:

Definition 5.1.3. *Given a 0-1 integer polytope $P(x) \subseteq [0,1]^d$, a distributional formulation $R(d)$ of $P(x)$ is an extended formulation $R(d)$ of $P(x)$, where $d \in \mathbb{R}^{|\{s | x^s \in P(x) \cap \{0,1\}^d\}|}$ and for every $x^s \in P(x) \cap \{0,1\}^d$ we have a variable d_s. For any feasible integer solution $x^{s^*} \in P(x) \cap \{0,1\}^d$ the vector d^{s^*}, whose components are defined as $d^{s^*}_s = 1$ iff $s = s^*$ and $d^{s^*}_s = 0$ otherwise, is feasible for any distributional formulation $D(z)$ of $P(x)$. We will refer to d^{s^*} as the encoding of the feasible integer solution x^{s^*} in the distributional variables.*

For a mixed integer set $M(x, w) \subseteq \{0,1\}^{d_x} \times \mathbb{R}^{d_w}$ the corresponding mixed integer polytope $P(x, w)$ is $\text{conv}(M(x, w))$. Let (x^s, w^s) be the encoding of a feasible mixed integer solution $s \in M$. In case one starts from a mixed integer polytope, the additional z variables of the product relaxation correspond to sets that contain at most one fractional variable. Including only one fractional variable in each product, mimics the variable space of the final-level SA relaxation.

Definition 5.1.4. *Let $P(x,w) \subseteq [0,1]^{d_x} \times \mathbb{R}^{d_w}$ be a mixed integer polytope. A mixed product relaxation $D(z)$ of $P(x,w)$ is an extended relaxation $D(z)$ of $P(x,w)$, where $z \in \mathbb{R}^{(d_w+1)2^{d_x}-1}$, with $z_{\{w_j\}} = w_j$, $j = 1,\ldots,d_w$, and*
(i) for every set $\varnothing \neq \mathcal{E} \subseteq \{x_1, x_2, \ldots, x_{d_x}\}$ we define $d_w + 1$ variables: one that we denote $z_{\mathcal{E}}$ and, for each fractional variable w_j, $j = 1,\ldots,d_w$, one that we denote $z_{\mathcal{E}w_j}$. Moreover $z_{\{x_i\}}$ denotes x_i, $i = 1,\ldots,d_x$.
(ii) the encoding of a feasible mixed-integer solution s is $z_{\mathcal{E}} = \prod_{x_i \in \mathcal{E}} x_i^s$ and $z_{\mathcal{E}w_j} = \prod_{x_i \in \mathcal{E}} x_i^s \cdot w_j^s$.

Note that the lifted polytope obtained from some specific linear relaxation of the 0-1 polytope $P(x)$ (mixed integer $P(x,w)$), at any level of the SA hierarchy, after linearization and before projection to the original variables, is a (mixed) product relaxation.

5.2 The expressive power of product relaxations

In this section we show the following. For every 0-1 polytope $P(x)$ and every (approximate) extended formulation $Q(x,y) = \{(x,y) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \mid Ax + By \leq b\}$ of $P(x)$ there is a product formulation T_Q which has the size of $Q(x,y)$ and is at least as strong in terms of approximability. Similarly, we show that there is a distributional formulation R_Q of the same as $Q(x,y)$ and at least as strong.

A *substitution T for the product space* is a linear map of the form $y = Tz$ where T is a $d_y \times (2^{d_x} - 1)$ matrix and z is a $2^{d_x} - 1$ dimensional vector having a coordinate $z_{\mathcal{E}}$ for each nonempty set $\mathcal{E}$ of the form $\{x_i \mid i \in S \subseteq 2^{\{1,\ldots,d_x\}}\}$. For any substitution T, the *translation* of $Q(x,y)$, denoted T_Q, the formulation resulting by substituting $T_{(i)}z$, for y_i, $i = 1, ..., d_y$. Here $T_{(i)}$ denotes the ith row of T. We require that T_Q is a product formulation (see Definition 5.1.2) and we say that we have a *translation of Q to product formulations* (recall that the original variables x_i coincide with the variables $z_{\{x_i\}}$). Observe that the number of inequalities of T_Q is the same as in $Q(x,y)$. The translation may heighten exponentially the dimension, but, since our methodology will give lower bounds on the size of the product formulations, those bounds apply to the size of $Q(x,y)$ as well. A *substitution T for the distributional space* and a *translation to distributional formulations* is defined similarly.

Theorem 5.2.1. *Given a 0-1 polytope $P(x) \subseteq [0,1]^{d_x}$, for every polytope $Q(x,y)$ such that $P(x) \subseteq \mathrm{proj_x}(Q(x,y))$ there is a translation T_Q to product formulations such that $P(x) \subseteq \mathrm{proj_x}(T_Q) \subseteq \mathrm{proj_x}(Q(x,y))$.*

Proof. We shall give a substitution T for the variables $y \in \mathbb{R}^{d_y}$ of $Q(x,y)$ so that the theorem holds. Let $g(x)$ be a section of $Q(x,y)$ (recall that a section associates every feasible 0-1 vector x of $P(x)$ to a specific y such that $(x,y) \in Q(x,y)$). We denote by $(p,1) \in \mathbb{R}^{n+1}$ the vector resulting from $p \in \mathbb{R}^n$ by appending the scalar 1 as an extra coordinate.

Observe that a product variable $z_{\mathcal{E}}$ behaves, as far as the encondings z^s of solutions $x^s \in P(x) \cap \{0,1\}^{d_x}$ to product variables are concerned, like the monomial $\prod_{x_i \in \mathcal{E}} x_i$ would. Those monomials plus the constant 1 form the *Fourier basis*. Likewise we can see a variable y_i, as far as the encondings x^s, y^s of solutions $x^s \in P(x) \cap \{0,1\}^{d_x}$ are concerned, as a boolean function $y_i(x) : \{0,1\}^{d_x} \to \mathbb{R}$ such that $y_i(x^s) = y_i^s$. By basic functional analysis (see, e.g.,[36]), we have that every boolean function $y_i(x)$ has a unique Fourier representation $y_i(x) = \sum_{\mathcal{E}\subseteq\{x_i|i=1,\dots,d_x\}} a_{\mathcal{E}}^{y_i} \prod_{x_i\in\mathcal{E}} x_i$. The intuition is that we will use the encodings z^s to product variables to simulate the encodings y^s. So we define the substitution T_i for a variable y_i as follows:

$$y_i = \sum_{\mathcal{E}\subseteq\{x_i|i=1,\dots,d_x\}} a_{\mathcal{E}}^{y_i} z_{\mathcal{E}} \tag{5.24}$$

In the above expression we assume, for notational convenience that, $z_\emptyset = 1$. Recall that product variables are defined for nonempty sets.

Obviously $\mathrm{proj}_\mathrm{x}(T_Q) \subseteq \mathrm{proj}_\mathrm{x}(Q(x,y))$: from any feasible solution (x^0, z^0) of T_Q we can derive a feasible solution (x^0, y^0) of $Q(x,y)$ by setting y^0 equal to Tz^0.

We will now show that $P(x) \subseteq \mathrm{proj}_\mathrm{x}(T_Q)$ or, more specifically, that the encodings z^s of solutions to product variables are feasible for T_Q as required by the definition of product relaxations. Observe that by letting the z vector take the value z^s for some $s \in P\cap\{0,1\}^{d_x}$, by (5.24) we get that the quantities involved in the inequalities of T_Q are the exact same quantities involved in the corresponding inequalities of $Q(x,y)$ for $(x,y) = (x^s, y^s)$. By definition (x^s, y^s) is feasible for $Q(x,y)$ and thus z^s is feasible for T_Q. □

Corollary 5.2.1. *A lower bound b on the size of any product relaxation D which is a ρ-approximate extended formulation of the 0-1 polytope $P(x)$, for $\rho \geq 1$, implies a lower bound b on the size of any ρ-approximate extended formulation $Q(x,y)$ of $P(x)$.*

We prove similar statements for the distributional formulations.

Theorem 5.2.2. *Given a 0-1 polytope $P(x) \subseteq [0,1]^{d_x}$, for every polytope $Q(x,y)$ such that $P(x) \subseteq \mathrm{proj}_\mathrm{x}(Q(x,y))$ there is distributional formulation R_Q such that $P(x) \subseteq \mathrm{proj}_\mathrm{x}(R_Q) \subseteq \mathrm{proj}_\mathrm{x}(Q(x,y))$.*

Proof. We shall, once again, give a substitution T for the variables $y \in \mathbb{R}^{d_y}$ of $Q(x,y)$, effectively "translating" them to distributional variables. Like before, we will use the encodings d^s of solutions $x^s \in P(x) \cap \{0,1\}^{d_x}$ to distributional variables to simulate the encodings y^s. We emphasize that d^s is the encoding of s and it is a fixed vector of the form $(0,..,0,1,0,..,0)$, not to be confused with d_s which is the variable (component of vector d) that corresponds to solution s. The substitution T is defined as follows: for a variable y_i we have

$$y_i = \sum_{s|x^s\in P(x)\cap\{0,1\}^{d_x}} y_i^s d_s \tag{5.25}$$

and for a variable x_j we have:

$$x_j = \sum_{s|x^s\in P(x)\cap\{0,1\}^{d_x}} x_j^s d_s \tag{5.26}$$

By substituting each occurrence of a variable y_i or x_j with the corresponding expression of distributional variables we get a formulation R_Q. Obviously $\mathrm{proj}_\mathrm{x}(R_Q) \subseteq \mathrm{proj}_\mathrm{x}(Q(x,y))$: from any feasible solution (d^0) of R_Q we can derive a feasible solution (x^0, y^0) of $Q(x,y)$ by setting (x^0, y^0) equal to Td^0.

We will now show that $P(x) \subseteq \mathrm{proj}_{\mathrm{x}}(R_Q)$ or, more specifically, that the encodings d^s of solutions to distributional variables are feasible for R_Q as required by the definition of product relaxations. Observe that by letting the d vector take the value d^s for some s such that $x^s \in P(x) \cap \{0,1\}^{d_x}$ by (5.25) and (5.26) we get that the quantities involved in the inequalities of R_Q are the exact same quantities involved in the corresponding inequalities of $Q(x,y)$ for $(x,y) = (x^s, y^s)$. By definition (x^s, y^s) is feasible for $Q(x,y)$ and thus d^s is feasible for R_Q. □

Corollary 5.2.2. *A lower bound b on the size of any distributional formulation D which is a ρ-approximate extended formulations of the 0-1 polytope $P(x)$, for $\rho \geq 1$, implies a lower bound b on the size of any ρ-approximate extended formulation $Q(x,y)$ of $P(x)$.*

Let $P(x,w)$ be a mixed integer polytope. The notion of the encoding of solutions of P for some extended relaxation $Q(x,w,y)$ of P is more challenging: intuitively, the solutions are characterized by two parts – a boolean part of the 0-1 assignments on the integer variables x and a "linear" part of the real variables w in the following sense: once the boolean part (the "hard" one) is fixed, the linear part can be obtained as the feasible region of a (usually small) system of inequalities, possibly empty.

Motivated by the above we define the following type of encoding for an extended formulation $Q(x,w,y)$ of a mixed-integer polytope. A *mixed-linear encoding* of EF Q consists of a function g_i for every auxiliary variable y_i such that the value $g_i(x', w)$ for a given integer vector x' is an affine function on w denoted $g_i^{x'}(w)$. If there is such a mixed-linear encoding for $Q(x,w,z)$, we say that Q is an *extended formulation with a mixed linear encoding.* An example of EFs with a mixed linear encoding are formulations arising from the SA procedure where y is the vector of the new variables corresponding to the linearized products. The following theorem can be proved similarly to Theorem 5.2.1.

Theorem 5.2.3. *Given a mixed integer polytope $P(x,w) \subseteq [0,1]^{d_x} \times \mathbb{R}^{d_w}$, for every ρ-approximate, $\rho \geq 1$, extended formulation $Q(x,w,y)$ with a mixed linear encoding, there is a translation $T[Q(x,w,y)]$ to mixed product relaxations such that*

$$P(x,w) \subseteq \mathrm{proj}_{\mathrm{x,w}}(T[Q(x,w,y)]) \subseteq \mathrm{proj}_{\mathrm{x,w}}(Q(x,w,y)).$$

Proof. Let the dimension of $P(x,w)$ be $d = d_x + d_w$. We shall give for the variables $y \in \mathbb{R}^{d_y}$ of $Q(x,w,y)$ a substitution T so that the theorem holds.

Consider a variable y_i and the corresponding coordinate of the mixed linear encoding, $g_i^{x'}(w) = \sum_j b_i^{x'} w_j + c_{x'}$ for each $x' \in \{0,1\}^{d_x}$ and $i = 1, \ldots, d_y$.

First, we will prove a helpful claim which states a fact from elementary Fourier analysis in our setting. For $x, s \in \mathrm{proj}_{\mathrm{x}}\left(P(x,w) \cap (\{0,1\}^{d_x} \times \mathbb{R}^{d_w})\right)$, define the boolean indicator operator $\chi_s(x)$ to be 1 when $s = x$ and 0 otherwise. First, we will show that this operator can be expressed as a linear combination of the product encodings constrained to monomials with only boolean variables. In other words, we determine coefficients $a_{\mathcal{E}}^s$, $\mathcal{E} \subseteq \{x_1, \ldots, x_{d_x}\}$, such that $\chi_s(x) = \sum_{\mathcal{E}} a_{\mathcal{E}}^s f_{\mathcal{E}}(x)$. The translation of the indicator operator $i_s(x)$ of an integer solution s is a linear expression of the form $T_{i_s} = \sum a_{\mathcal{E}}^s P[\mathcal{E}](x)$. We shall iteratively generate the coefficients $a_{\mathcal{E}}^s$. The only nonzero coefficients will be those corresponding to sets of variables that are supersets of the set of variables being 1 in s – let that set be $\mathcal{E}_1^s$. We give the construction iteratively starting from $|\mathcal{E}_1^s|$ to d_x, defining in step k the coefficients of such sets of size k.

In the first iteration simply set $a_{\mathcal{E}_1^s} = 1$. At step $k > |\mathcal{E}_1^s|$, for each set $\mathcal{E}'$ of size k that is a superset of $\mathcal{E}_1^s$, set $a_{\mathcal{E}'}^s = -\sum_{\mathcal{E} \subset \mathcal{E}'} a_{\mathcal{E}}^s$. This concludes the definition of the coefficients.

Claim 5.2.1. *For each integer solution* $s' \in \mathrm{proj}_{\mathrm{x}}\left(P(x,w) \cap (\{0,1\}^{d_x} \times \mathbb{R}^{d_w})\right)$, $\chi_s(s') = \sum_{\mathcal{E}} a^s_{\mathcal{E}} f_{\mathcal{E}}(s')$.

Proof of the claim. By overloading the notation, we denote by s both the integer solution and the support of that integer solution, that is the set $\{x_i \mid s_i = 1\}$. If $s' \supseteq s$ then the nonzero terms of the sum $\sum_{\mathcal{E}} a^s_{\mathcal{E}} f_{\mathcal{E}}(s')$ are exactly those that correspond to sets $\mathcal{E}$ such that $s \subseteq \mathcal{E} \subseteq s'$. We have that $\sum_{\mathcal{E}} a^s_{\mathcal{E}} f_{\mathcal{E}}(s') = \sum_{\mathcal{E} \subseteq s'} a^s_{\mathcal{E}}$ which, by the construction of the coefficients, is 1 if $s = s'$ and 0 if $s' \supset s$, as required. Otherwise, if $s - s' \neq \varnothing$, then all the $f_{\mathcal{E}}(s')$ with nonzero coefficients are 0, so $\sum_{\mathcal{E}} a^s_{\mathcal{E}} f_{\mathcal{E}}(s') = 0$.

By Claim 5.2.1 we have that for an integer vector $s \in \{0,1\}^{d_x}$ the indicator operator $\chi_s(x)$ is equal to $\sum_{\mathcal{E} \subseteq \{x_1,\ldots,x_{d_x}\}} a^s_{\mathcal{E}} f_{\mathcal{E}}(x)$. For each set of integer variables $\mathcal{E}$ and each fractional variable w_j let $z_{\mathcal{E}w_j}$ denote the corresponding mixed product variable and $f_{\mathcal{E}w_j}(x,w)$ the corresponding coordinate of the mixed product encoding. It is now easy to show the following.

Claim 5.2.2. *For each mixed integer solution* (x', w'), *and for* $i = 1, \ldots, d_y$,
$g^{x'}_i(w') = \sum_j \sum_{\mathcal{E}} a^s_{\mathcal{E}} b^{x'}_i f_{\mathcal{E}w_j}(x', w') + \sum_{\mathcal{E}} a^s_{\mathcal{E}} c_{x'} f_{\mathcal{E}}(x')$.

To conclude the definition of T, set

$$y^i = \sum_{x'} \sum_j \sum_{\mathcal{E}} a^s_{\mathcal{E}} b^i_{x'} z_{\mathcal{E}w_j} + \sum_{x'} \sum_{\mathcal{E}} a^s_{\mathcal{E}} c_{x'} z_{\mathcal{E}}, \quad i = 1, \ldots, d_y.$$

which implies

$$y^i = \sum_{x'} \sum_{\mathcal{E}} a^s_{\mathcal{E}} (\sum_j b^i_{x'} z_{\mathcal{E}w_j} + c_{x'} z_{\mathcal{E}}), \quad i = 1, \ldots, d_y$$

By Claim 5.2.2, using arguments similar to the ones in the proof of Theorem 5.2.1, it follows that that $P(x,w) \subseteq \mathrm{proj}_{\mathrm{x,w}}(T[Q(x,w,y)]) \subseteq \mathrm{proj}_{\mathrm{x,w}}(Q(x,w,y))$. □

Corollary 5.2.3. *A lower bound b on the size of any mixed product relaxation D which is a ρ-approximate extended formulation of the 0-1 mixed integer polytope $P(x,w)$ implies a lower bound b on the size of any ρ-approximate extended formulation $Q(x,w,y)$ of $P(x,w)$ with a mixed linear encoding.*

5.3 The model of Linear Formulations

Recently concerns were raised on whether the extension complexity depends on the encoding of the solutions of a combinatorial problem in the feasible integer points of the polytope $P(x)$, to which we require a formulation Q to project (see also [18]). We frequently call the variable space of $P(x)$ as the *original space* and we call the encoding of solutions in the feasible integer points of $P(x)$ as the *basic encoding*. More specifically the concern was that the underlying optimization problem Π, whose solutions are encoded in the feasible integer points of a polytope $P(x)$, might have a different formulation as a linear program $K(t)$ such that the extension complexity of $K(t)$ is lower that the extension complexity of $P(x)$. In [18] the theory of extended formulations was generalized to be independent of the encoding of a problem Π as linear program and they showed that this was not the case. Here, after presenting the general setting of [18], we show that the distributional and product formulations capture the strength of this model as well. Then, by using the distributional formulations, we give a short and elegant independent proof of the independence of approximate extension complexity from the basic encoding. Then we show that any encoding in which the objective functions can be faithfully represented, has the maximum strength with respect to extension complexity. Although this result can already be inferred from [66] we give an independent proof via distributional formulations.

We use the (very natural) definition of an optimization problem of [18].

Definition 5.3.1. *[18] (Optimization problems). An optimization problem $P = (S, F)$ consists of a set S of feasible solutions and a set F of objective functions. An approximation problem $P^* = (P, F^*)$ is an optimization problem P together with a family $F^* := \{f^* \mid f \in F\}$ of real numbers called approximation guarantees for P, so that $f^* \geq \max_{s\in S} f(s)$ if P is a maximization problem, and $f^* \leq \min_{s\in S} f(s)$ if P is a minimization problem. An algorithm approximately solves P^* if it provides an approximation $\hat{f}$ of the optimum $\max_{s\in S} f(s)$ (resp. $min_{s\in S} f(s)$) for all $f \in F$ satisfying*

$$\max_{s\in S} f(s) \leq \hat{f} \leq f^* (resp. \min_{s\in S} f(s) \geq \hat{f} \geq f^*). \tag{5.27}$$

In order to determine the exact maximum of a maximization problem, one chooses $f^* := \max f$. To determine the maximum of nonnegative objective functions within an approximation factor $0 < \rho \leq 1$, one chooses $f^* = (\max f)/\rho$. This choice is motivated so as to be comparable with factors of approximation algorithms finding a feasible solution s with $f(s) \geq \rho \max f$. For a minimization problem, $f^* := \min f$ in the exact case, and $f^* = (\min f)/\rho$ for an approximation factor $\rho \geq 1$ provided f is nonnegative.

Definition 5.3.2. *[18] (LP formulation of an optimization problem). An LP formulation of an approximation problem $P^* = (P, F^*)$ with $P = (S, F)$ is a linear program $Ax \leq b$ with $x \in \mathbb{R}^d$ together with the following realizations:*

(i) Feasible solutions as vectors $x^s \in \mathbb{R}^d$ for every $s \in S$ so that $Ax^s \leq b$ for all $s \in S$, i.e., the system $Ax \leq b$ is a relaxation (superset) of $\mathrm{conv}(x^s \mid s \in S)$.

(ii) Objective functions via affine functions $w^f : \mathbb{R}^d \to \mathbb{R}$ for every $f \in F$ such that $w^f(x^s) = f(s)$ for all $s \in S$, i.e., we require that the linearization w^f of f is exact on all x^s with $s \in S$.

(iii) Achieving approximation guarantee f^* via requiring

$$\max\{w^f(x) \mid Ax \leq b\} \leq f^* \text{for all } f \in F, \tag{5.28}$$

for maximization problems (resp. $\min\{w^f(x) \mid Ax \leq b\} \geq f^*$ for minimization problems).

The size of the formulation is the number of inequalities in $Ax \leq b$. Finally, the *LP formulation complexity* $\mathrm{fc}_+(P)$ of the problem P is the minimal size of all its LP formulations.

The distributional and product relaxations, similar to the case of extended formulations, can also simulate any LP formulation with respect to Definition 5.3.2 without increasing its size. The two theorems that follow are proved similarly to Theorems 5.2.1 and 5.2.3 respectively.

Theorem 5.3.1. *Let $Ay \leq b$, $y \in \mathbb{R}^n$, $b \in \mathbb{R}^t$, $A \in \mathbb{R}^t \times \mathbb{R}^n$, be an LP formulation of size t of an approximation problem $P^* = (P, F^*)$ for some optimization problem $P = (S, F)$. Let w^f be the vector of the formulation that corresponds to the objective function $f \in F$. There* *is a distributional formulation $R \subseteq \mathbb{R}^{|S|}$ of the optimization problem P, defined by some substitution T of the y variables with the d_s variables, which has size t and* $\max\{w^f(Td) \mid d$ *feasible for* $R\} \leq \max\{w^f(y) \mid Ay \leq b\} \leq f^*$.

Proof. We define the substitution T as follows. For each variable y_i: $y_i = \sum_s y_i^s d_s$

Now substitute each occurrence of a variable y_i in $Ay \leq b$ with $\sum_s y_i^s d_s$ - the resulting formulation is R. For each objective $f \in F$, the encoding of f for R is simply $r^f = w^f(T)$. In the latter expression we treat the substitution T as being a function mapping points from the distributional space to points in the space of the LP formulation. The affine function r^f is a faithful encoding of f since for all $s \in S$ we have $r^f(d^s) = w^f(Td^s) = w^f(y^s) = f(s)$.

The defined formulation R is a valid distributional formulation that has size t and is at least as strong as $Ay \leq b$ because:

- the encoding d^{s^*} of any solution $s^* \in S$ to distributional variables is by construction feasible for R since, for $d = d^{s^*}$, we get the same quantities in the lhs of the inequalities of R that we would get if we have let the variables y take the values y^{s^*} in $Ay \leq b$, which is by definition feasible.
- for any objective $f \in F$ we have $\max\{r^f(d) \mid d$ feasible for $R\} \leq \max\{w^f(y) \mid Ay \leq b\} \leq f^*$ since for any feasible solution d of R we can get a feasible solution y for $Ay \leq b$ of the same cost by letting the y_i variables take the value of the expression $\sum_s y_i^s d_s$ for that particular d.

□

Although the equivalence between product and distributional formulations can be inferred from Theorems 5.2.1 and 5.2.3, we will give an independent proof.

Theorem 5.3.2. *Let $Ay \leq b$, $y \in \mathbb{R}^n$, $b \in \mathbb{R}^t$, $A \in \mathbb{R}^t \times \mathbb{R}^n$, be an LP formulation of size t of an approximation problem $P^* = (P, F^*)$ for some optimization problem $P = (S, F)$. There is a product formulation $D \subseteq \mathbb{R}^{2^{\lceil \log |S| \rceil} - 1}$ of the optimization problem P, defined by some substitution T of the y variables with the z variables, which has size t and* $\max\{w^f(Tz) \mid z$ *feasible for* $D\} \leq \max\{w^f(y) \mid Ay \leq b\} \leq f^*$.

Proof. Name the members of S with binary strings of length $\lceil \log |S| \rceil$, giving each $s \in S$ a distinct label. Let $x \in \mathbb{R}^{\lceil \log |S| \rceil}$ be a vector whose ith component x_i represents the ith bit of the label of $s \in S$. Let x^s be the encoding of $s \in S$ in the x variables. We define the variables of the product formulation D with respect to the x variables, and the encoding z^s in product variables with respect to x^s, $s \in S$.

We define the substitution T as follows. For each variable y_j: $y_j = \sum_{\mathcal{E} \subseteq \{x_i | i=1,\ldots,\lceil \log |S| \rceil\}} a_{\mathcal{E}}^{y_j} z_{\mathcal{E}}$. Here, as in the proof of Theorem 5.2.3, $a_{\mathcal{E}}^{y_j}$ are the coefficients of the unique Fourier representation $y_j(x) = \sum_{\mathcal{E} \subseteq \{x_i | i=1,\ldots,\lceil \log |S| \rceil\}} a_{\mathcal{E}}^{y_j} \prod_{x_i \in \mathcal{E}} x_i$. Once again we assume, for notational convenience, that $z_\emptyset = 1$.

Now substitute each occurrence of a variable y_j in $Ay \leq b$ with $\sum_{\mathcal{E}} a_{\mathcal{E}}^{y_j} z_{\mathcal{E}}$ - the resulting formulation is D. For each objective $f \in F$, the encoding of f for R is simply $r^f = w^f(T)$. The vector r^f is a faithful encoding of f since for all $s \in S$ we have $r^f(z^s) = w^f(Tz^s) = w^f(y^s) = f(s)$.

R is a valid product formulation that has the size t and is at least as strong as $Ay \leq b$ because:

- the encoding z^{s^*} of any solution $s^* \in S$ to product variables is by construction feasible for D since, for $z = z^{s^*}$, we get the same quantities in the lhs of the inequalities of R that we would get if we have let the variables y take the values y^{s^*} in $Ay \leq b$, which is by definition feasible.
- for any objective $f \in F$ we have $\max\{r^f(z) \mid z$ feasible for $D\} \leq \max\{w^f(y) \mid Ay \leq b\} \leq f^*$ since for any feasible solution z of D we can get a feasible solution y for $Ay \leq b$ of the same cost by letting the y variables take the value of the expression $y_j = \sum_{\mathcal{E} \subseteq \{x_i | i=1,\ldots,\lceil \log |S| \rceil\}} a_{\mathcal{E}}^{y_j} z_{\mathcal{E}}$ for that particular z.

$\square$

To display the usefulness of using distributional (product) formulations when arguing about the power of extended formulations or about the power of the more general LP formulations we will prove a perhaps surprising result: the (approximate) extension complexity *does not* depend on the encoding. In other words problems that are hard for extended formulations with respect to some encoding are also hard for extended formulations with respect to some other encoding and moreover of the exact same hardness. In the next theorem we prove the independence of the classic (approximate) extension complexity from the encoding used or, equivalently, from the polytope $P(x)$ that traditionally we require the extended formulation to project to (or to include in its projection in the case of the approximate extended formulations).

Theorem 5.3.3. *Let $Ay \leq b$, $y \in \mathbb{R}^n$, $b \in \mathbb{R}^t$, $A \in \mathbb{R}^t \times \mathbb{R}^n$ be an LP formulation of size t of an approximation problem $P^* = (P, F^*)$ of some optimization problem $P = (S, F)$. Let $x \in \mathbb{R}^{d_x}$ be the vector of a set of variables, let the encodings of solutions be x^s for $s \in S$ and assume that there is an encoding $h^f \in \mathbb{R}^{d_x}$ for every objective function $f \in F$ of the problem P^* for the x variables such that $h^f(x^s) = f(s)$. Then there is an approximate extended formulation $R^+(x, d)$ with respect to x^s and h^f that has size t and is at least as strong, i.e., $\max\{h^f(x) \mid \exists d$ s.t. (x, d) is feasible for $R^+\} \leq \max\{w^f(y) \mid Ay \leq b\} \leq f^*$ for any f.*

Proof. Let R be the distributional formulation that we constructed in the proof of Theorem 5.3.1. Add to R the following set of equalities (defining the projection to the x variables): $x_j = \sum_{s \in S} x_j^s d_s$. Let R^+ be the resulting extended formulation. It is easy to see that the projection of R^+ contains $\mathrm{conv}(x^s \mid s \in S)$ since the vector (x^s, d^s) is feasible for R^+. What is left is to show that for any objective $f \in F$, $\max\{h^f(x) \mid \exists z$ s.t. (x, z) is feasible for $R^+\} \leq \max\{w^f(y) \mid Ay \leq b\}$. For simplicity we will assume that the objective functions are linear – the case of affine objective functions is no harder to prove.

Let (x', d') be a feasible solution to R^+. We will show that for any objective f there is a solution y' to $Ay \leq b$ such that $h^f(x') = w^f(y') \leq f^*$. Let $y' = Td'$ be the vector defined by the substitution T, where T is defined as in the proof of Theorem 5.3.1. Recall that y' is feasible for $Ay \leq b$ since d' is feasible for R. We have that $h^f(x') = \sum_{s \in S} h^f(x^s d'_s) = \sum_{s \in S} f(s) d'_s = \sum_{s \in S} w^f(y^s) d'_s = w^f(\sum_{s \in S} y^s d'_s) = w^f(y')$.

Thus the strength of classic extended formulations is independent of the encoding and it is equal to the strength of LP formulations. $\square$

We conclude this section by commenting on the importance of the existence of faithfull encodings of the objective functions for some encoding of a combinatorial problem. For example one could associate the feasible solutions $s \in S$ of the Matching problem (on a clique K_n) with the vertices x^s of a hypercube, which, of course, has low extension complexity. Let us consider the following set of objective functions for Matching: we have one objective function f_G for every graph G on n vertices and the value of a matching s for f_G is the number of edges of s that are also edges of G. Is the description of the aforementioned hypercube a small formulation for this optimization version of Matching? The answer lies in the fact that there is no way to encode the objective functions in faithful way, that is to define for each f_g a vector w^{f_G} such that $w^{f_G} x^s = f_G(s)$. The latter can already be inferred from the result of Rothvoss [57] (see also [18]).

5.4 A method for lower bounding the size of LPs with known encodings

Here we present a methodology to lower bound the size of relaxations that achieve a desired integrality gap. For simplicity we do not deal in this section with mixed integer sets.

Our method can be summarized as follows. Let $G(z) \subseteq [0,1]^d$ be a 0-1 polytope. We design a family $\mathcal{I}$ of instances parameterized by the dimension d. For each instance $I \in \mathcal{I}$ of dimension d we define a set of points $\mathcal{C}_I \subseteq [0,1]^d \setminus G(z)$ which we call the *core of* I *with respect to* G. Note that the points of the core must be infeasible for G. To prove a lower bound $r(n)$ on the size of G it suffices to show that at least that many inequalities are needed to separate $\mathcal{C}_I$ from G. Additionally, for a minimization problem with O being the set of admissible objective functions, if for some $z \in \mathcal{C}_I$ there is an admissible cost function w_z such that $w_z^T z < \rho^{-1}\text{Opt}_{I,w_z}$, $0 < \rho \leq 1$, where Opt_{I,w_z} is the cost of the optimal integer solution with respect to w_z, we call z *ρ-gap inducing wrt O*. If we design the core so that all its members are ρ-gap inducing, the lower bound will hold for ρ-approximate formulations.

To define constructively the core for a specific family of extended formulations of a polytope P the encoding of solutions in the variables z must be known. This requirement is fulfilled by the product (distributional) relaxations we will focus on. By Theorem 5.2.1 above, proving a lower bound on the size for an arbitrary extended relaxation $Q(x,y)$ of a polytope $P(x)$ can be reduced to a proof of the same bound on the size of a corresponding product relaxation $D(z)$. The following meta-theorem shows that such a proof can always be obtained by proving the existence of a suitable core for the product relaxation. Let the "tightest" product relaxation of $P(x)$ be $\hat{D} = \text{conv}(z^s \mid s \in P \cap \{0,1\}^d)$. We say that a set of vectors $s \subseteq [0,1]^d \setminus \hat{D}$ is *conflicting* if $\text{conv}(s) \cap \hat{D} \neq \varnothing$. Any single valid inequality of $\hat{D}$ cannot separate all points of a conflicting set. Given a set $O_d \subseteq \mathbb{R}^d$ of admissible objective functions associated with a 0-1 polytope $P(x) \subseteq [0,1]^d$, we define $\tilde{O}_d \subseteq \mathbb{R}^{2^d-1}$, to contain the vectors in O_d extended with zeroes in the coordinates corresponding to the non-singleton product variables.

Theorem 5.4.1. *Given a 0-1 polytope $P(x) \subseteq [0,1]^d$, and an associated set of admissible objective functions $O_d \subseteq \mathbb{R}^d$, the ρ-approximate extension complexity, $\rho \geq 1$, of $P(x)$ is at least $r(n)$, iff there exists a family of instances $\mathcal{I}(n)$ and, for every $I \in \mathcal{I}$, a core $\mathcal{C}_I$ wrt $\hat{D}$, which consists of ρ-gap inducing vectors wrt $\tilde{O}_d$, with the following property: for any partition of $\mathcal{C}_I$ into less than $r(n)$ parts there must be a part containing a set of conflicting vectors.*

Proof. Assume first that the ρ-approximate extension complexity is at least $r(n)$. Define $\mathcal{C}_I$ to be the set of all ρ-gap inducing product vectors. If we can partition $\mathcal{C}_I$ into less than $r(n)$ parts so that there is no conflicting subset s in any part, then we can define an inequality for each part of the partition that separates the vectors of at least that part from $\hat{D}$. But we know that less than $r(n)$ inequalities cannot separate all the ρ-gap inducing product vectors. Thus we have that for any decomposition of those vectors into less than $r(n)$ parts there must be a part containing a set of conflicting vectors.

Conversely, assume we can find a core $\mathcal{C}_I$ wrt $\hat{D}$ consisting of ρ-gap inducing vectors such that for any partition of $\mathcal{C}_I$ into less than $r(n)$ sets there must be a part containing a set of conflicting vectors. Then the size of $\hat{D}$ is at least $r(n)$. If not, there is a decomposition into less than $r(n)$ parts where each part consists of the core members separated by each inequality – in case a member is separated by more than one inequality, we arbitrarily include it into just one of the resulting parts. Observe that $\mathcal{C}_I$ is not only a core wrt $\hat{D}$ but also is a core wrt any ρ-approximate product relaxation of P. By Theorem 5.2.1, the lower bound $r(n)$ applies to the size of any ρ-approximate extended formulation of P. □

Let $\mathcal{H}(\mathcal{C}_I)$ be the, possibly infinite, hypergraph with vertices the members of $\mathcal{C}_I$ and hyperedges the conflicting subsets of $\mathcal{C}_I$. Theorem 5.4.1 can be restated more conveniently:

Theorem 5.4.2. *Given a 0-1 polytope $P(x) \subseteq [0,1]^d$, and an associated set of admissible objective functions $O_d \subseteq \mathbb{R}^d$, the ρ-approximate extension complexity, $\rho \geq 1$, of $P(x)$ is at least $r(n)$, iff there exists a family of instances $\mathcal{I}(n)$ and, for every $I \in \mathcal{I}$, a core $\mathcal{C}_I$ wrt $\hat{D}$, which consists of ρ-gap inducing vectors wrt $\tilde{O}_d$, such that $\mathcal{H}(\mathcal{C}_I)$ has chromatic number $r(n)$.*

Theorem 5.4.1 suggests that the best possible lower bound on the extension complexity can always be achieved by proving the existence of an appropriate core in the product space. In the applications in this chapter we implement a version of the method that imposes stronger requirements on the decomposition, namely the constructed hypergraph will be a clique.

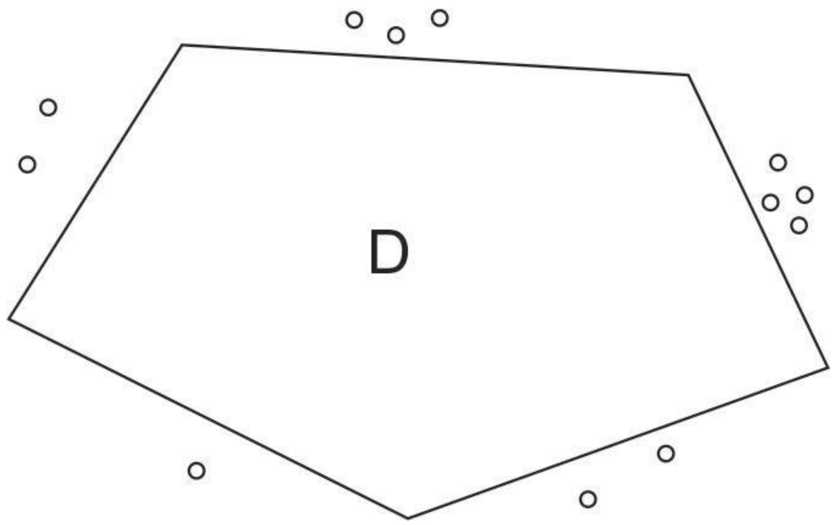

Figure 5.3: Example of a valid partition of a set of gap inducing vectors. In this example the core $\mathcal{C}_I$ consists of all the points in the figure.

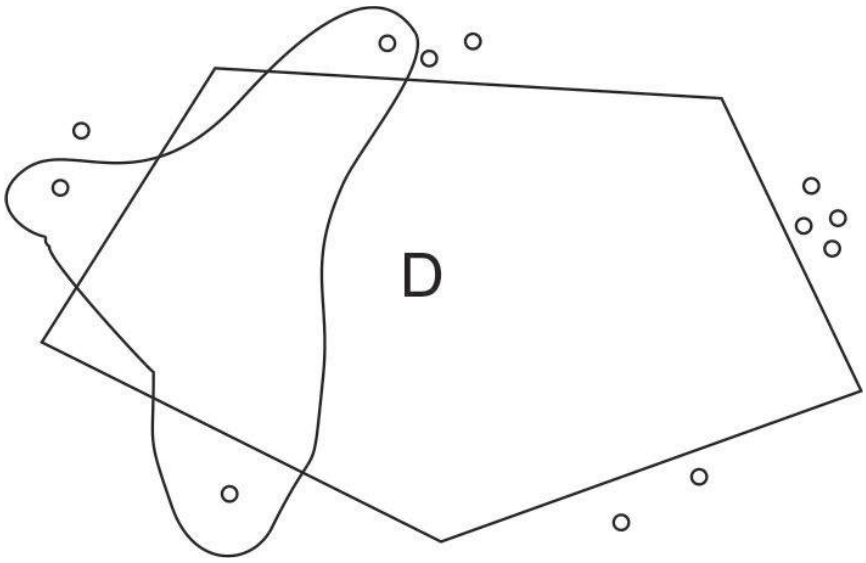

Figure 5.4: One of the conflicting sets of $\mathcal{C}_I$.

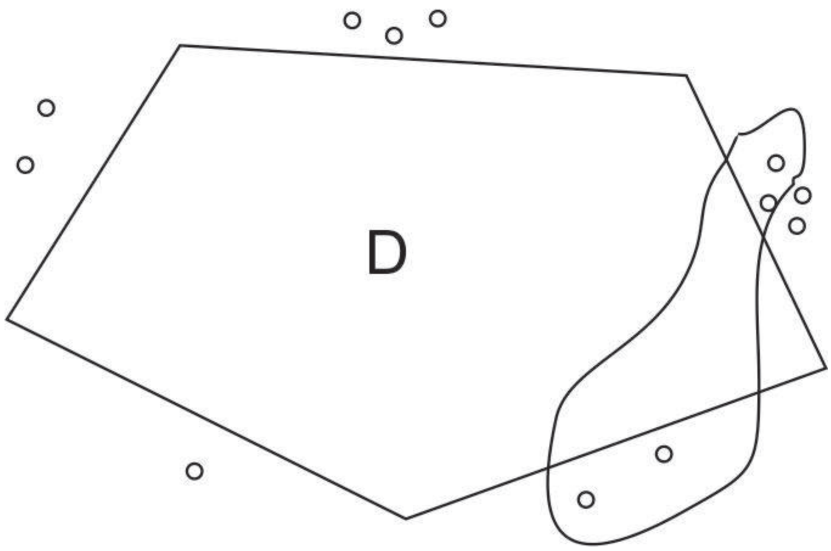

Figure 5.5: One more conflicting set of $\mathcal{C}_I$.

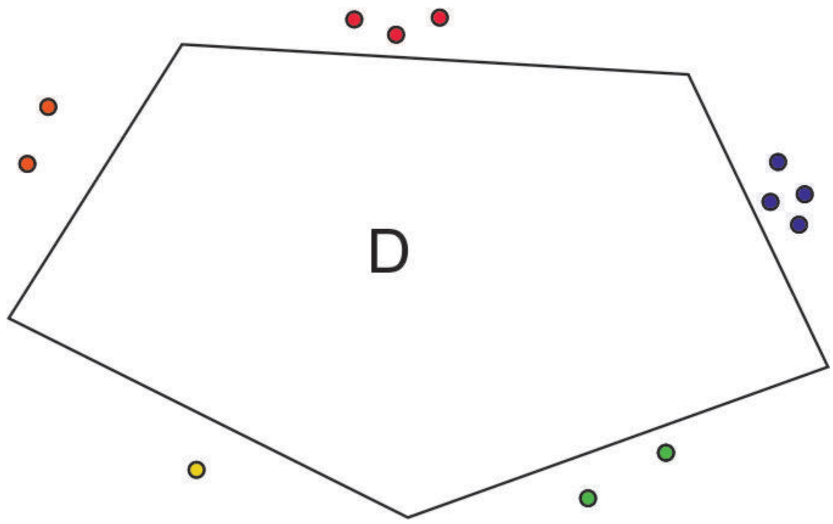

Figure 5.6: A proper coloring of the underlying hypergraph $\mathcal{H}(\mathcal{C}_I)$.

5.5 Warmup: bounds on the size of approximate CFL relaxations that use the classic variables

In the case of CFL, the linear encoding (L, O) is defined as follows. For a CFL instance, given the number n of facilities, the number m of clients, the capacities $K \in \mathbb{R}^n_+$ and the demands $D \in \mathbb{R}^m_+$, we use the variables y_i, $i = 1, \ldots, n$, x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ with the usual meaning of facility opening and client assignment respectively. The set of feasible solutions (y, x) is defined in the obvious manner. Thus for dimension $d = n + nm$, $L \cap \{0, 1\}^d$ is completely determined by the quadruple (n, m, K, D). The set of admissible objective functions $O \cap \mathbb{R}^{n+nm}$ is the set of pairs $(\mathbf{f}, \mathbf{c})$ where $\mathbf{f} \in \mathbb{R}^n_+$ are the facility opening costs and $\mathbf{c} = [c_{ij}] \in \mathbb{R}^{nm}_+$ are connection costs that satisfy $c_{ij} \leq c_{i'j} + c_{i'j'} + c_{ij'}$.

In our proof we will consider feasible sets of the form $(n, m, U\mathbf{1}, \mathbf{1})$, i.e., with uniform capacities $U > 0$, and unit demands. Therefore the triple (n, m, U) is sufficient description. Furthermore, it will be convenient to deviate from the convention that the number of facilities is n - this is to simplify the expressions appearing through the proof. Let the number of facilities be n^2, the number of clients be an^4 for some integer $a \geq 2$ and the capacity U of each facility be n^3. Thus for a given n, the feasible set is uniquely determined by the triple (n^2, an^4, n^3). To avoid cumbersome expressions, we slightly abuse terminology and refer to such a triple as an *instance* $I(n^2, an^4, n^3)$. We denote for the instance in question the set of facilities by F and the set of clients by C.

We first describe the core $\mathcal{C}_I$ of the instance $I(n^2, an^4, n^3)$.

Definition 5.5.1. *The core $\mathcal{C}_I$ of the instance $I(n^2, an^4, n^3)$ is the following set of (y, x) vectors. $\forall k, l \subset F$ with $|k|, |l| = n$ and $k \cap l = \emptyset$ and for a set $C_{k,l}$ of clients with $|C_{k,l}| = Un+1$ we define a vector $s_{k,l}$ such that: (1) $y_i = 1, \forall i \in k$, $y_i = \frac{10}{n^2}, \forall i \in l$, $y_i = 1, \forall i \notin k \cup l$. (2) For a client $j \in C_{k,l}$ we have $x_{ij} = \frac{1-1/n^2}{n}, \forall i \in k$, $x_{ij} = \frac{1}{n^3}, \forall i \in l$ and $x_{ij} = 0, \forall i \notin k \cup l$. (3) For a client $j \notin C_{k,l}$ we have $\forall i \in k \cup l$, $x_{ij} = 0$ and $\forall i \notin k \cup l$, $x_{ij} = \frac{1}{n^2-2n}$.*

We say that two vectors $s_{k,l}, s_{k',l'} \in \mathcal{C}_I$ *collide* with each other if $l \setminus (k' \cup l') \neq \emptyset$ and $l' \setminus (k \cup l) \neq \emptyset$. We proceed by proving that for each $s \in \mathcal{C}_I$ the ratio of the number of the members of $\mathcal{C}_I$ that do not collide with s to the number of the colliding members is exponentially small.

Lemma 5.5.1. *For each $s_{k,l} \in \mathcal{C}_I$ let $\mathcal{U} \subseteq \mathcal{C}_I$ be the set of vectors in the core that collide with $s_{k,l}$. Then $\frac{|\mathcal{C}_I| - |\mathcal{U}|}{|\mathcal{C}_I|} = 2^{-\Omega(n \log n)}$.*

Proof. We lower-bound the ratio in question by upper bounding the probability that a member of $\mathcal{C}_I$ picked uniformly at random does not collide with $s_{k,l}$. Consider the event $\mathcal{E}_1$ that $l' \setminus (k \cup l) = \emptyset$. It must be the case that $l' \subseteq k \cup l$. The probability $P[\mathcal{E}_1]$ is at most $(\frac{2n}{n^2})^n = (2/n)^n$ - this is the probability that all members of l' are in $k \cup l$ if we were to pick them with repetition and the probability of the actual $\mathcal{E}_1$ is less since we do not allow repetitions in the set l. Likewise the probability of the event $\mathcal{E}_2$ that $l \setminus (k' \cup l') = \emptyset$ is the same. So, by the union bound, the probability that a randomly picked element of $\mathcal{C}_I$ does not collide with $s_{k,l}$ is $P[\mathcal{E}_1 \cup \mathcal{E}_2] \leq 2(2/n)^n$. □

Next we show that for any two colliding vectors $s_{k,l}$ and $s_{k',l'}$ in $\mathcal{C}_I$ there is a convex combination s' of them that is contained in the integer polytope.

Lemma 5.5.2. *For any two colliding vectors* $s_{k,l}, s_{k',l'} \in \mathcal{C}_I$, $\mathrm{conv}(\{s_{k,l}, s_{k',l'}\}) \cap \mathrm{conv}(P) \neq \emptyset$.

Proof. We will actually show that the average vector $s' = \frac{s_{k,l}+s_{k',l'}}{2}$ is a convex combination of integer solutions. We will do show by giving a distribution $\mathcal{D}$ over integer solutions whose expected vector $(y^{\mathcal{D}}, x^{\mathcal{D}})$ with respect to $\mathcal{D}$ is s'. The intuition behind the proof is that each one of the vectors $s_{k,l}, s_{k',l'}$, in order to become a convex combination of integer solutions, needs what the other has in abundance: some measure for the y variable of a facility in l and some measure for the y variable of a facility in l' respectively. With probability $1/2$ we choose to perform experiment A and with probability $1/2$ we perform experiment B described below.

Suppose that experiment A is chosen. We will describe the random solution in two steps: A_1 and A_2. We describe first step A_1. Let f be a member of the set $l \setminus (k' \cup l')$, which is non-empty by the choice of k, l, k', l'. We select exactly one facility to be opened from the set l according to to the following probabilities: for $i \in l - \{f\}$ the probability is equal to $y_i^{s_{k,l}} = \frac{10}{n^2}$, while for facility f the probability is equal to $1 - \sum_{i \in l - \{f\}} y_i^{s_{k,l}}$. Facilities in k are always opened in the experiment step A_1. When some facility $i \in l - \{f\}$ is chosen we randomly select $w^i_{A_1} = \frac{\sum_{j \in C_{k,l}} x_{ij}^{s_{k,l}}}{y_i^{s_{k,l}}}$ clients from $C_{k,l}$ and assign them to i - we assume without loss of generality that $w^i_{A_1}$ is an integer, in the Appendix we show how to handle fractional w's. Assign the remaining clients in $C_{k,l}$ randomly to the facilities $i' \in k$ so that each one is assigned exactly $w^{i'}_{A_1} = \frac{|C_{k,l}| - w^i_{A_1}}{n}$ clients (again we assume w.l.o.g. that this is an integer). When facility f is chosen we randomly select $w^f_{A_1} = \frac{\sum_{j \in C_{k,l}} x_{fj}^{s_{k,l}}}{1 - \sum_{i \in l - \{f\}} y_i^{s_{k,l}}}$ clients from $C_{k,l}$ and assign them to f. Assign the remaining clients in $C_{k,l}$ randomly to the facilities $i' \in k$ so that each one is assigned exactly $w^{i'}_{A_1} = \frac{|C_{k,l}| - w^f_{A_1}}{n}$ clients (again we assume w.l.o.g. that the w's are integers).

For the second step A_2 of the experiment, let g be a facility in $l' \setminus (k \cup l)$. We select facility g to be opened with a probability $\sum_{i \in l} y_i^{s_{k,l}}$, the other facilities in $F - (k \cup l)$ are always opened in the experiment step A_2. If g is opened, it is assigned $w^g_{A_2} = \frac{\sum_j x_{gj}^{s_{k,l}}}{\sum_{i \in l} y_i^{s_{k,l}}}$ clients randomly chosen from $C - C_{k,l}$ and the remaining clients of $C - C_{k,l}$ are assigned randomly to the facilities i' in $F - (k \cup l) - \{g\}$ so that each one is assigned exactly $w^{i'}_{A_2} = \frac{|C - C_{k,l}| - w^g_{A_2}}{|F - (k \cup l) - \{g\}|}$. If g is not opened, all the clients in $C - C_{k,l}$ are assigned randomly to the facilities i' in $F - (k \cup l) - \{g\}$ so that each one is assigned exactly $w^{i'}_{A_2} = \frac{|C - C_{k,l}|}{|F - (k \cup l) - \{g\}|}$.

Now suppose that experiment B is chosen. This case is symmetric to the previous experiment by exchanging sets k, l with k', l' respectively but we give the full description for the sake of completeness. Again we will describe the random solution in two steps B_1 and B_2.

We descibe first step B_1. Let g be the member of $l' \setminus (k \cup l)$ that we used in step A_2. We select exactly one facility to be opened from the set l' with respect to the following probabilities: for $i' \in l' - \{g\}$ the probability is equal to $y_{i'}^{s_{k',l'}} = \frac{10}{n^2}$, while for facility g the probability is equal to $1 - \sum_{i \in l' - \{g\}} y_i^{s_{k',l'}}$. Facilities in k' are always opened in the experiment step B_1. When some facility $i' \in l' - \{g\}$ is chosen we randomly select $w^{i'}_{B_1} = \frac{\sum_{j \in C_{k',l'}} x_{i'j}^{s_{k',l'}}}{y_{i'}^{s_{k',l'}}}$ clients from $C_{k',l'}$ and assign them to i' - assume again w.l.o.g. that $w^{i'}_{B_1}$ is an integer. Assign the rest of the clients in $C_{k',l'}$ randomly to the facilities $i'' \in k'$ so that each one is assigned exactly $w^{i''}_{B_1} = \frac{|C_{k',l'}| - w^{i'}_{B_1}}{n}$ clients (again we assume w.l.o.g. that this is an integer). When facility g is chosen we randomly select $w^g_{B_1} = \frac{\sum_{j\in C_{k',l'}} x^{s_{k',l'}}_{gj}}{1-\sum_{i\in l'-\{g\}} y^{s_{k',l'}}_i}$ clients from $C_{k',l'}$ and assign them to g. Assign the rest of clients in $C_{k',l'}$ randomly to the facilities $i'' \in k'$ so that each one is assigned exactly $w^{i''}_{B_1} = \frac{|C_{k',l'}|-w^g_{B_1}}{n}$ clients (again we assume w.l.o.g. that the w's are integers).

For the second step B_2 of the experiment, let f be the facility in $l \setminus (k' \cup l')$ used in step A_1. We select facility f to be opened with a probability $\sum_{i\in l} y^{s_{k,l}}_i$, the other facilities in $F - (k' \cup l')$ are always opened in the experiment step B_2. If f is opened, it is assigned $w^f_{B_2} = \frac{\sum_j x^{s_{k',l'}}_{fj}}{\sum_{i\in l} y^{s_{k,l}}_i}$ clients randomly chosen from $C-C_{k',l'}$ and the remaining clients of $C-C_{k',l'}$ are assigned randomly to the facilities i' in $F - (k' \cup l') - \{f\}$ so that each one is assigned exactly $w^{i'}_{B_2} = \frac{|C-C_{k',l'}|-w^f_{B_2}}{|C-(k'\cup l')-\{f\}|}$. If f is not opened, all the clients in $C - C_{k',l'}$ are assigned randomly to the facilities i' in $F - (k' \cup l') - \{f\}$ so that each one is assigned exactly $w^{i'}_{B_2} = \frac{|C-C_{k',l'}|}{|F-(k'\cup l')-\{f\}|}$.

It is easy to see that the outcome of each experiment is always a feasible integer solution, since all clients are assigned to opened facilities and the capacities are respected by the choice of w's. It is also easy to verify that s' is the expected vector of the distribution $\mathcal{D}$ defined above. A facility $i \in F - (k \cup l \cup k' \cup l')$ is always opened in both experiments and thus $y^{\mathcal{D}}_i = y'_i = 1$. A facility $i' \in (k\cup l\cup k'\cup l') - (\{f, g\})$ is opened in experiment A a fraction $y^{s_{k,l}}_{i'}$ of the time and is opened in experiment B a fraction $y^{s_{k',l'}}_i$ of time, and since each experiment is selected with $1/2$ probability, we have $y^{\mathcal{D}}_{i'} = \frac{y^{s_{k,l}}_{i'}+y^{s_{k',l'}}_{i'}}{2} = y^{s'}_{i'}$. For facility f we have that in experiment A it is opened $1-\sum_{i\in l-\{f\}} y^{s_{k,l}}_i$ while in experiment B it is opened $\sum_{i\in l} y^{s_{k,l}}_i = \sum_{i\in l'} y^{s_{k',l'}}_i$ of the time and so $y^{\mathcal{D}}_f = 1/2 + y^{s_{k,l}}_f/2 = y^{s'}_f$. Similarly for facility g. By similar arguments the desired properties can be shown for the assignment variables. Let us consdier for example, for facility $f \in F - (k' \cup l')$ we have that in A_1 the expected total demand assigned to it is $w^f_{A_1} P[y_f = 1] = \sum_{j\in C_{k,l}} x^{s_{k,l}}_{fj}$ and by the symmetric way that clients in $C_{k,l}$ are assigned to f in A_1 we have that $x^{A_1}_{fj} = x^{s_{k,l}}_{fj}$ for all j. In experiment step B_2, the expected total demand assigned to f is $w^f_{B_2} P[y_f = 1] = \sum_{j\in C-C_{k',l'}} x^{s_{k',l'}}_{fj}$ and by the symmetric way that clients in $C - C_{k',l'}$ are assigned to f in B_2 we have that $x^{B_2}_{fj} = x^{s_{k',l'}}_{fj}$ for all j. Thus $x^{\mathcal{D}}_{fj} = \frac{x^{s_{k,l}}_{fj}+x^{s_{k',l'}}_{fj}}{2}$. □

Theorem 5.5.1. *Every approximate formulation for metric* CFL *that uses the natural encoding and has integrality gap at most g for some constant $g > 0$, has $2^{\Omega(n\log n)}$ constraints.*

Proof. We first prove that for every vector $s_{k,l} \in \mathcal{C}_I$ there is an admissible cost function $w_{k,l}$ such that $w^T_{k,l} s_{k,l} = o(w^T_{k,l} s^{opt}_{k,l})$ where $s^{opt}_{k,l}$ is an optimal integer solution of $I(n^2, an^4, n^3)$ with respect to $w_{k,l}$. Consider two points p_1, p_2 in some Euclidean space at distance 1 from each other. At the first point p_1, the facilities of $k \cup l$ and the clients of $C_{k,l}$ are co-located and the remaining facilities and clients are all co-located at p_2. Additionally the facilities in l have all opening cost of 1 and the rest have 0 opening cost. It is easy to see that every integer solution has a cost of at least 1: either some client $j \in C_{k,l}$ is assigned to some facility located at p_2 and thus incurs a connection cost of 1, or some costly facility in l must be opened at p_1, incurring a facility cost of 1. On the other hand $w^T_{k,l} s_{k,l} = o(1)$.

Consider some inequality π of a g-approximate relaxation Q, where $g > 0$ is a constant. (In fact the proof holds for $g = o(n)$). Suppose there is some $s_{k,l} \in \mathcal{C}_I$ that violates π. Then, for every $s'_{k',l'} \in \mathcal{C}_I$ which collides with $s_{k,l}$, π must be satisfied otherwise by Lemma 5.5.2

we have violation of validity. By Lemma 5.5.1 we have that π eliminates $2^{-\Omega(n \log n)}|\mathcal{C}_I|$ members of the core, and by using the union bound the theorem is proved. We note that for the sake of simplicity the parameters are not optimized - by using a different core we can get tighter bounds. $\square$

5.6 Lower bounds for approximate mixed product relaxations for CFL

For CFL, the linear encoding $\mathcal{N}_{\text{CFL}} = (L, O)$ is defined as follows. For a CFL instance, given the number n of facilities, the number m of clients, the capacities $K \in \mathbb{R}^n_+$ and the demands $D \in \mathbb{R}^m_+$, we use the classic variables $y_i, i = 1, \ldots, n, x_{ij}, i = 1, \ldots, n, j = 1, \ldots, m$ with the usual meaning of facility opening and client assignment respectively. The set of feasible solutions (y, x) is defined in the obvious manner. Thus for dimension $d = n + nm$, $L \cap \{0,1\}^d$ is completely determined by the quadruple (n, m, K, D). The set of admissible objective functions $O \cap \mathbb{R}^{n+nm}$ is the set of pairs $(\mathbf{f}, \mathbf{c})$ where $\mathbf{f} \in \mathbb{R}^n_+$ are the facility opening costs and $\mathbf{c} = [c_{ij}] \in \mathbb{R}^{nm}_+$ are connection costs that satisfy $c_{ij} \leq c_{i'j} + c_{i'j'} + c_{ij'}$.

The capacitated facility location problem with general capacities and demands is a mixed integer optimization problem where the facilities are opened integrally but the clients are allowed to be assigned fractionally to the set of opened facilities. Before we proceed with the main application of our lower bounding methodology for extended formulations which concerns lower bounds for mixed product relaxations of approximate CFL, we first give a small proof of exponential extension complexity of the CFL polytope (the exact, not the approximate version). We will use a simple "LP-reduction" from the minimum Knapsack polytope: we show that the CFL polytope is an extension of the minimum Knapsack polytope.

The 0-1 Knapsack problem is one of the "easiest" NP-hard optimization problem as it admits a pseudo-polynomial time algorithm which yields a simple FPTAS. The classic integer program formulation of Knapsack is a very simple one having a single non-trivial linear constraint. The Max 0-1 Knapsack polytope associated with weights $a \in \mathbb{R}^n$ and right-hand side $b > 0$ is defined as $\text{MAX}\, K(n, a, b) := \text{conv}\{t \in \mathbb{R}^n \cap \{0,1\}^n \mid a^T t \leq b\}$ while the Min version is defined as $\text{MIN}\, K(n, a, b) := \text{conv}\{t \in \mathbb{R}^n \cap \{0,1\}^n \mid a^T t \geq b\}$

As noted in [54], when discussing the extension complexity, $\text{MAX}\, K(n, a, b)$ and $\text{MIN}\, K(n, a, b)$ are equivalent since one is obtained from the other by complementing variables: we can take a formulation for the min version from the max version as $a^T(1-t) \leq \sum_i a_i - b$ and using the map $y_i = 1 - t_i$. In [54] the following theorem was proved via LP-reduction: a face of the Knapsack polytope is an extension of the correlation polytope.

Theorem 5.6.1. *[54] For any $n \in \mathbb{N}$ there exists a 0-1 Knapsack polytope $P \subseteq \mathbb{R}^n$ so that $xc(P) = \Omega(2^{\sqrt{n}})$.*

Historically, the above theorem was the first example of a polytope with high extension complexity which, however, admits a fully polynomial-size LP relaxation scheme (recall that this is the LP equivalent of an FPTAS): as it was shown in [14] that, for fixed ϵ, the $(1+\epsilon)$ approximate Knapsack polytope admits a formulation of polynomial size. By reducing Knapsack to CFL with general capacities and client demands we get the following.

Theorem 5.6.2. *The CFL polytope has an extension complexity of $\Omega(2^{\sqrt{n}})$ where n is the number of facilities.*

Proof. Consider an instance of the Min Knapsack with n items, weights a_i and total weight target b. We construct an instance of single source CFL with a single client (source) with demand b and n facilities with capacities a_i. It is easy to see that the linear map $t_i = y_i$ certifies that the minimum Knapsack polytope is indeed the projection of a face of the CFL polytope (or, equivalently, a face of the CFL polytope is an extension of the minimum Knapsack polytope): for any feasible $\bar{t}$ solution to the Knapsack instance there is a feasible solution (y, x) that maps to $\bar{t}$: just assign the demand in any feasible way to the opened facilities. For the opposite direction, any feasible $(\bar{y}, \bar{x})$ to the CFL instance maps to a feasible solution for the Knapsack instance: by adding 2.10 for all facilities and using 2.15 we get $\sum_i a_i y_i \leq \sum_i b x_{ij} = b$. □

For the rest of this section, we show an exponential lower bound on the size of any mixed product relaxation of the *approximate* CFL polytope. We note that, while the extension complexity of the CFL polytope is expected to be exponential since the CFL problem is NP-hard, at the same time it can be approximated within a factor of $o(n)$ of the optimal and moreover by a constant factor (see the relevant discussion in the introduction).

In our proof we will consider a parameterized instance $I = I(3n, m, U, d)$ with uniform capacities U and uniform unit demands $d = 1$, where $3n$ is the number of facilities, and m the number of clients. Furthermore we will have that the number of clients is $m = n^4 + 1$ and the capacities and demands are such that $(n^4 + 1) - nU = 2^{-n^2}$. Observe that $n^3 < U < (n^3 + 1)$. In order to define the core $\mathcal{C}_I$ of the instance I we first describe a random experiment based on whose outcome we will later define the members of the core. Given disjoint sets $k, l \subseteq F$ of size n each, the random experiment defines a distribution $\mathcal{D}_{k,l}$ over mixed integer vectors in the classic encoding. These vectors correspond in general to pseudo-solutions. The following experiment defines the distribution $\mathcal{D}_{k,l}$. The quantities $\bar{x}_{ij}$ are defined in Lemma 5.6.1 below.

RANDOM EXPERIMENT

Facilities in k are always opened.

Case 1. With probability $1 - \frac{20}{n^2(1+1/n)}$ all facilities in $F - l$ are opened and those of l are closed. Distribute evenly the client demand to facilities in k. Note that this outcome of the experiment does not respect the capacities.

Case 2. Otherwise, i.e., with probability $\frac{20}{n^2(1+1/n)}$, pick at random a subset q of the facilities in $F - k$ with at least one facility from l and open them. Assign randomly demand to each facility i in $q \cap l$ so that i takes $\frac{\sum_j \bar{x}_{ij}}{10/n^2}$ units and the rest of the demand is equally distributed to the facilities in k.

Lemma 5.6.1. *The expected vector $(\bar{y}, \bar{x})$ wrt $\mathcal{D}_{k,l}$ is the following: $\bar{y}_i = 1$ for $i \in k$, $\bar{y}_i = 1 - \frac{10}{n^2(1+1/n)}$ for $i \in F - k - l$, $\bar{y}_i = \frac{20(2^{n-1})}{n^2(1+1/n)(2^n-1)}$ for $i \in l$. For all $j \in C$, $\bar{x}_{ij} = \frac{1-n^{-2}}{|k|}$ for $i \in k$, $\bar{x}_{ij} = 0$ for $i \in F - \{k \cup l\}$, $\bar{x}_{ij} = \frac{n^{-2}}{|l|}$ for $i \in l$.*

Proof. For $i \in k$ we have that i is always open in $\mathcal{D}_{k,l}$ so $\bar{y}_i = P_{\mathcal{D}_{k,l}}[i \text{ opened}] = 1$. For $i \in l$, note that it is opened only in case 2 when $i \in q$. The set $l \cap q$ is a randomly selected nonempty subset of l. Thus $\bar{y}_i = P_{\mathcal{D}_{k,l}}[i \text{ opened}] = P_{\mathcal{D}_{k,l}}[i \in q \text{ of case 2}] = \frac{20}{n^2(1+1/n)} \frac{2^{n-1}}{2^n-1} = \frac{20(2^{n-1})}{n^2(1+1/n)(2^n-1)}$. Similarly for the $\bar{y}$ variables of facilities in $F - (k \cup l)$. As for the assignment variables, for each j and each facility $i \in l$, each time i is opened it is assigned $\frac{\sum_j \bar{x}_{ij}}{\bar{y}_i}$ demand at random and since it is opened a $\bar{y}_i$ fraction of the time, the total expected

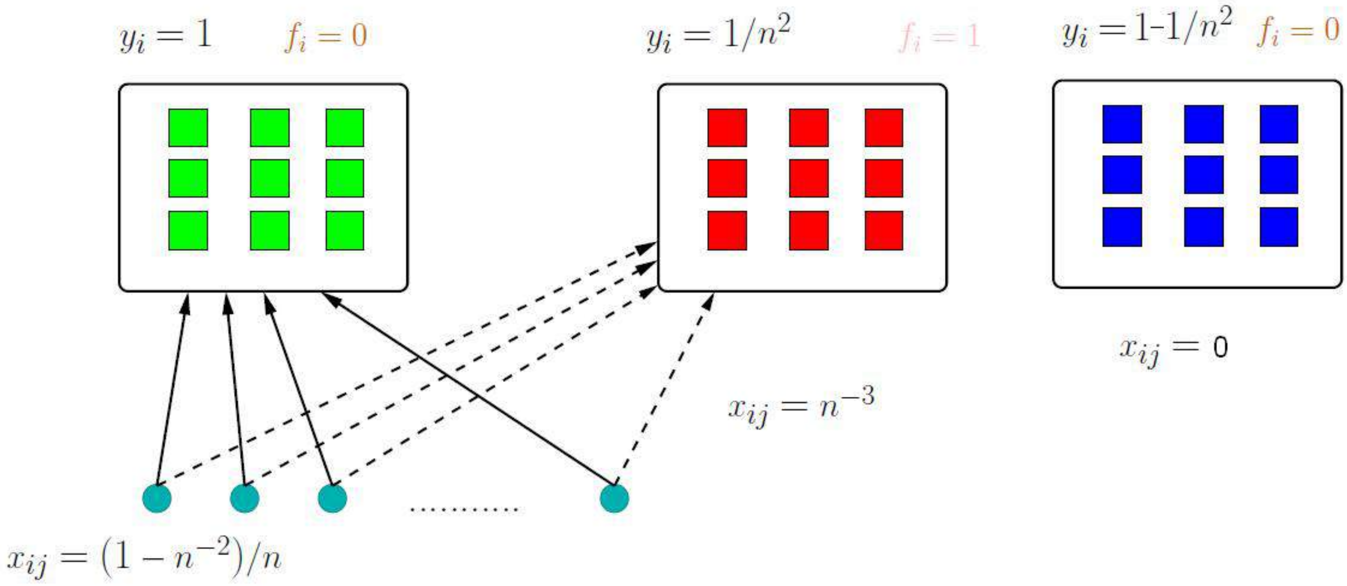

Figure 5.7: Depiction of a member of the core $\mathcal{C}_I$. The green facilities are those of the set k, the red facilities belong to l and the rest are the blue facilities in $F - k - l$

demand assigned to it is $\frac{\sum_j \bar{x}_{ij}}{\bar{y}_i}\bar{y}_i = \sum_j \bar{x}_{ij}$. Since the assignments are random each client is assigned to i with the same fraction in expectation, so $P_{\mathcal{D}_{k,l}}[i \text{ assigned to } j] = \bar{x}_{ij}$. Facilities in $F - \{k \cup l\}$ are never assigned any demand. By the construction of the distribution the demand not assigned to l, is assigned to the facilities in k randomly and so the expected $\bar{x}_{ij}$ have their intended values.

□

The distribution $\mathcal{D}_{k,l}$ will be subsequently used to define the members of the core $\mathcal{C}_I$. Let $\mathcal{E}$ be a subset of integer variables in the original space, i.e., $\mathcal{E} \subseteq \{y_1, \ldots, y_{3n}\}$. We denote by $E_{\mathcal{D}_{k,l}}[\mathcal{E}]$ the expectation of the event where all the variables in $\mathcal{E}$ have value 1, i.e., the expectation of the product $\prod_{y_{i_k} \in \mathcal{E}} y_{i_k}$. Similarly, we denote $E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij}]$ the expectation of the product $(\prod_{y_{i_k} \in \mathcal{E}} y_{i_k}) \cdot x_{ij}$. Let $\chi(case1), \chi(case2)$ be the 0-1 random variables that indicate whether Case 1 and Case 2 occur, respectively. We denote by $E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1]$ the expectation of the product $(\prod_{y_{i_k} \in \mathcal{E}} y_{i_k}) \cdot \chi(case1)$ and by $E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case1]$ the expectation of the product $(\prod_{y_{i_k} \in \mathcal{E}} y_{i_k}) \cdot x_{ij} \cdot \chi(case1)$. Similarly for Case 2. Intuitively, $E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case1]$ is the "mass" that $\mathcal{D}_{k,l}$ assigns to x_{ij} over all outcomes of case 1 where the variables of $\mathcal{E}$ have value 1.

To simplify notation, we use $z(i)$ instead of z_i to refer to a coordinate of vector z indexed by i. From now on, P denotes the CFL polytope and $\hat{D}$ its canonical product relaxation.

Definition 5.6.1. *Fix a set $k \subset F$ of size n. The core $\mathcal{C}_I$ of the instance $I(3n, n^4 + 1, U, 1)$ wrt $\hat{D}$ is the following set of product vectors: $\forall l \subset F$ with $|l| = n$ and $k \cap l = \varnothing$ and for every set $\mathcal{E}$ of integer variables and for every fractional variable x_{ij} we define $z_{k,l}(\mathcal{E}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}]$ and $z_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij}]$.*

Now we are ready to state the key Lemma 5.6.2 from which our main theorem will be

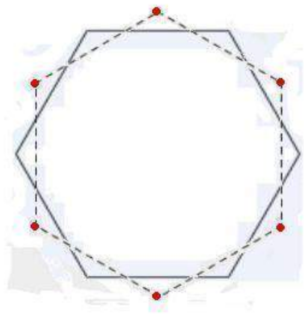

Figure 5.8: A depiction of relation of the members of the core with respect to $\hat{D}$: the segment defined by any two members of $\mathcal{C}_I$ intersect $\hat{D}$.

derived. The proof of the lemma is quite technical and for organization purposes we dedicated Subsection 5.6.1 to the exposition of its proof.

Lemma 5.6.2. *For any two $z_{k,l}, z_{k,l'} \in \mathcal{C}_I$ such that $l - l' \neq \varnothing$ there is some $z \in \mathrm{conv}(z_{k,l}, z_{k,l'})$ which is feasible for $\hat{D}$.*

Theorem 5.6.3. *Given the family of* CFL *instances $I(3n, n^4 + 1, U, 1)$, each member of $\mathcal{C}_I$ is $\Omega(n)$-gap inducing and $\chi(\mathcal{H}(\mathcal{C}_I)) = 2^{\Omega(n)}$. Therefore, there is a constant $c > 0$, s.t. any cN-approximate EF for* CFL *with a mixed linear encoding has size $2^{\Omega(N)}$, where N is the number of facilities.*

Proof. Since we proved in Lemma 5.6.2 that any two members of the core $\mathcal{C}_I$ form a conflicting set, $\mathcal{H}(\mathcal{C}_I)$ is a clique and thus its chromatic number is $|\mathcal{C}_I| = \binom{2n}{n} = 2^{\Theta(n)}$. For each member of the core $z_{k,l}$ there is an admissible cost function $w_{k,l}$ inducing $\Theta(n)$ gap: facilities in l have unit opening costs and every other facility has 0 opening cost. The facilities in $k \cup l$ and all the clients are co-located, and the rest of the facilities are co-located at distance 2^{n^2} from the former. Observe that each feasible mixed integer solution has a cost of at least 1 since either some facility in l must be opened integrally or at least 2^{-n^2} client demand has to be assigned to some facility in $F - k - l$. On the other hand the cost of $z_{k,l}$ wrt $w_{k,l}$ is $\Theta(n^{-1})$ since the (y, x) projection of $z_{k,l}$ is the expected vector $(\bar{y}, \bar{x})$ of $\mathcal{D}_{k,l}$. □

On the other hand, it is easy to see that for every instance I of CFL there is an exact formulation of size $2^N p$ where p is a polynomial expression in the size of the instance. Moreover this formulation can be written as a mixed product relaxation. The idea is to simply define a formulation for each choice of the opened facilities and then take the convex hull of those polytopes.

Observation 5.6.1. *There is an exact mixed product relaxation of the* CFL *polytope of size $2^N p$, where $p = \Theta(mN)$, N and m being the number of facilities and clients respectively.*

Proof. For each choice of opened facilities $O \subseteq F$ consider the following polytope P^O:

$$x_{ij}^O \leq y_i^O \qquad \forall i \in F, \forall j \in C \tag{5.29}$$
$$\sum_{i \in F} x_{ij}^O = 1 \qquad \forall j \in C \tag{5.30}$$
$$0 \leq x_{ij}^O \leq 1 \qquad \forall i \in F, \forall j \in C \tag{5.31}$$
$$\sum_{j \in C} x_{ij}^O \leq U_i y_i^O \qquad \forall i \in F \tag{5.32}$$
$$y_i^O = 0 \qquad \forall i \in F - O \tag{5.33}$$
$$y_i^O = 1 \qquad \forall i \in O \tag{5.34}$$

It obviously an exact formulation of the (possibly empty) polytope when we fix the values of the y_i's wrt O. Then by introducing "selection" variables s^O with the meaning of whether the set of opened facilities is O or not we get the following convex combination of polytopes:

$$\sum_{O \subseteq F} s^O = 1 \tag{5.35}$$
$$s^O \geq 0 \qquad \forall O \subseteq F \tag{5.36}$$
$$x_{ij}^O \leq y_i^O \qquad \forall i \in F, \forall j \in C, \forall O \subseteq F \tag{5.37}$$
$$\sum_{i \in F} x_{ij}^O = s^O \qquad \forall j \in C, \forall O \subseteq F \tag{5.38}$$
$$0 \leq x_{ij}^O \leq s^O \qquad \forall i \in F, \forall j \in C, \forall O \subseteq F \tag{5.39}$$
$$\sum_{j \in C} x_{ij}^O \leq U_i y_i^O \qquad \forall i \in F, \forall O \subseteq F \tag{5.40}$$
$$y_i^O = 0 \qquad \forall O \subseteq F, \forall i \in F - O \tag{5.41}$$
$$y_i^O = s^O \qquad \forall O \subseteq F, \forall i \in O \tag{5.42}$$

The proof is concluded by observing that the above linear program has a mixed linear encoding and by using Theorem 5.2.3. □

Theorem 5.6.4. *Let P be any linear relaxation of the* CFL *polytope for the family of instances $I(3n, n^4 + 1, U, 1)$ that uses the encoding $\mathcal{N}_{\text{CFL}}$ and has size $2^{o(n)}$. There is a constant $c > 0$, such that for all $t \leq cn$, the integrality gap of $\text{SA}^t(P)$ is $\Omega(n)$.*

Proof. Observe that for every level of SA there is a suitable projection of $\mathcal{C}_I$ s.t we obtain a legal core with respect to the product variables used in that level. Therefore, the lower bound on the size implied by Theorem 5.6.3 holds at all levels. The number of the inequalities of the t-level SA relaxation after the lifting and linearization stages, and before projection, obtained from any starting relaxation P of size r is less than $r\binom{n}{t}2^t$. By choosing $t \leq cn$, with c sufficiently small, we obtain that $r\binom{n}{t}2^t \leq r2^{\delta n}$ for a small $\delta > 0$. By Theorem 5.6.3 we get that for this value of t, the integrality gap on the given family of instances is $\Omega(n)$. This is asymptotically tight since SA is known to produce an exact formulation after $3n$ levels □

5.6.1 Proof of Lemma 5.6.2

In the first part of the proof we will show that by exchanging some measure of some components of the two product vectors $z_{k,l}, z_{k,l'}$ of the core, we can construct two new product vectors $z^*_{k,l}, z^*_{k,l'}$ each of which is feasible for $\hat{D}$. To establish this feasibility we will show for each of them that it is a convex combination over vectors of the form $f(y, x)$ where (y, x) are feasible mixed integer solutions in the CFL polytope P and f is the mixed product encoding.

Consider the two sets of facilities $l - l'$ and $l' - l$. Clearly $|l - l'| = |l' - l| > 0$, since $l \neq l'$ and $|l| = |l'| = n$. We construct a product vector $z^*_{k,l}$ based on $z_{k,l}$ and making some alterations and, symmetrically, a product vector $z^*_{k,l'}$ based on $z_{k,l'}$.

Construction of $z^*_{k,l}$

For any set $\mathcal{E}$ containing only facilities from $F - l'$ with at least one from $l - l'$: $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1]$ (Similarly, for any i, j, $z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1]$). In the case set $\mathcal{E}$ contains only facilities from $F - l$ with at least one from $l' - l$ we have $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) - E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1]$. (Similarly, for any i, j, $z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij}) - E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case1]$). In any other case and for any i, j let $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E})$ and $z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij})$.

The construction of $z^*_{k,l'}$ is symmetric but we give the details for the sake of completeness.

Construction of $z^*_{k,l'}$

For any set $\mathcal{E}$ containing only facilities from $F - l$ with at least one from $l' - l$: $z^*_{k,l'}(\mathcal{E}) = z_{k,l'}(\mathcal{E}) + E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1]$ (Similarly, for any i, j, $z^*_{k,l'}(\mathcal{E}x_{ij}) = z_{k,l'}(\mathcal{E}x_{ij}) + E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case1]$). In the case set $\mathcal{E}$ contains only facilities from $F - l'$ with at least one from $l - l'$ we have $z^*_{k,l'}(\mathcal{E}) = z_{k,l'}(\mathcal{E}) - E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1]$. (Similarly, for any i, j, $z^*_{k,l'}(\mathcal{E}x_{ij}) = z_{k,l'}(\mathcal{E}x_{ij}) - E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1]$.) In any other case and for any i, j let $z^*_{k,l'}(\mathcal{E}) = z_{k,l'}(\mathcal{E})$ and $z^*_{k,l'}(\mathcal{E}x_{ij}) = z_{k,l'}(\mathcal{E}x_{ij})$.

Next we show that the constructed $z^*_{k,l}$ and $z^*_{k,l'}$ are indeed the expected vectors of distributions $\mathcal{D}^*_{k,l}$ and $\mathcal{D}^*_{k,l'}$, respectively, over feasible mixed integer product solutions. We will only give the proof for $z^*_{k,l}$ since the other case is similar.

Before we continue the proof, we first explain the intuition behind the construction above. Both $z_{k,l}$ and $z_{k,l'}$ are not derived from distributions over feasible solutions because in any such feasible solution at least one facility from l and l' respectively has to be opened and assigned some demand. By assigning demand only to the set of facilities in k we cannot satisfy the total demand without violating the capacities. This is actually the main difference between the distributions $\mathcal{D}^*_{k,l}$ and $\mathcal{D}_{k,l}$. In Case 1^* there is at least one from l opened but, if we were to explain $z_{k,l}$ and $z_{k,l'}$ as resulting distributions over feasible solutions, the total measure of the opening of facilities in l and l' respectively is too small to have some facility from any of those sets opened 100% of the time when Case 1 happens. So in the construction of $z^*_{k,l}$ we will have all the facilities in $l - l'$ opened in Case 1^* and in the construction of $z^*_{k,l'}$ we will have all the facilities in $l' - l$ opened in

Case 1^*. Since those facilities are now opened a greater fraction of the time, many events involving them will have their probabilities increased. But where did we find the measure to increase the probability of those events? We construct $z^*_{k,l}$ "from" $z_{k,l}$ by increasing the probability of those events in $z^*_{k,l}$ by the same amount that we decrease their probability in the construction of $z^*_{k,l'}$ from $z_{k,l'}$. Similarly we increase the probability of those events in the construction of $z^*_{k,l'}$ from $z_{k,l'}$ by the same amount that we decrease their probability in the construction of $z^*_{k,l}$ from $z_{k,l}$. If we prove the validity of the above and moreover prove that $\mathrm{conv}(z^*_{k,l}, z^*_{k,l'}) \cap \mathrm{conv}(z_{k,l}, z_{k,l'}) \neq \varnothing$, then Lemma 5.6.2 follows.

Claim 5.6.1. *There is a distribution $\mathcal{D}^*_{k,l}$ over mixed integer product vectors which are feasible for $\hat{D}$ such that for any $\mathcal{E}$ and any i, j, $z^*_{k,l}(\mathcal{E}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$ and $z^*_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$.*

Proof. Let $\mathcal{D}^*_{k,l}$ be the distribution defined by the following experiment. The vector $(\bar{y}, \bar{x})$ is the one defined in Lemma 5.6.1.

Facilities in k are always opened.

Case 1^*

With probability $1 - \frac{20}{n^2(1+1/n)}$ all facilities in $F - l'$ are opened - all the facilities in l' are closed. Evenly assign client demand to facilities in k so each one takes exactly U, and assign the remaining 2^{-n^2} demand evenly to the facilities in $l - l'$.

Case 2^*

With probability $\frac{20}{n^2(1+1/n)}$ pick at random a subset q of the facilities in $F - k$ with at least one facility from l and open them.

Case $2.a^*$

If $q \neq F - k - l'$ assign randomly demand to facilities in $q \cap l$ so that each one of them takes $\frac{\sum_j \bar{x}_{ij}}{y_i}$ demand and the rest of the demand is equally distributed to the facilities in k.

Case $2.b^*$

Otherwise, when $q = F - k - l'$, assign randomly a quantity $\frac{\sum_j \bar{x}_{ij} d}{\bar{y}_i} - 2^{-n^2} \frac{P[\chi(case1^*)]}{P[\chi(case2.b^*)]}$ to each of the facilities in $l - l'$ and assign the remaining demand evenly to the facilities in k.

It is easy to see that $\mathcal{D}^*_{k,l}$ is a distribution over feasible solutions for the instance.

Proposition 5.6.1. *Each outcome of the experiment defining $\mathcal{D}^*_{k,l}$ is a feasible solution to the instance.*

Proof. In every outcome of the probabilistic experiment which induces the distribution $\mathcal{D}^*_{k,l}$ all the client demand is assigned to the opened facilities. It remains to show that the capacities U are respected. Consider case 1: the capacities of the facilities in k are respected by construction (recall that in this case those facilities are saturated) and facilities in $l - l'$ share a total demand of $2^{-n^2} < U$. Now consider case 2: a facility $i \in q$ is assigned at most $\frac{\sum_j \bar{x}_{ij}}{10/n^2} < U$ demand and since the rest of the demand is equally distributed among the facilities in k, each facility in k takes at most (equality when $q = F - k - l'$) $\frac{(n^4+1) - [(|l-l'|) \frac{\sum_j \bar{x}_{ij}}{\bar{y}_i} - (2^{-n^2}) \frac{E[\chi(case1^*)]}{E[\chi(case2.b^*)]}]}{|k|} < U$. □

To get Claim 5.6.1, we prove the following proposition.

Proposition 5.6.2. *For every set $\mathcal{E}$ and any i, j, we have that $z^*_{k,l}(\mathcal{E}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$ and $z^*_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$.*

Proof. Consider the following cases.

Case A.1

Assume that set $\mathcal{E}$ contains facilities only from $F - l'$ with at least one from $l - l'$. We have that $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1]$. By the definition of $z_{k,l}(\mathcal{E})$ we have $z_{k,l}(\mathcal{E}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}] = E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case2]$ since in case 1 of $\mathcal{D}_{k,l}$ none of the facilities in l are opened. We also have that $E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case2] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case2^*]$ since no assignment variable appears in $\mathcal{E}$ and all the other elements of the experiments of cases 2 and 2^* induce the exact same distribution. We also have that $E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case1^*]$ by the fact that when case 1 happens in $\mathcal{D}_{z_{k,l'}}$ the facilities in $l - l'$ are opened 100% and the same happens in $\mathcal{D}^*_{k,l}$, while the 2 distributions agree on everything except the assignments in that case by construction (recall again that no assignment variable appears in $\mathcal{E}$). So we have $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case2^*] + E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case1^*] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$ since cases 1^* and 2^* partition the probability space.

Case A.2.a

Consider the case where $\mathcal{E}$ contains facilities only from $F - l'$ with at least one from $l - l'$ and let x_{ij} be an assignment to some facility $i \in k$. We have that

$z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1] =$
$E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case2] + E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1]$(none of the facilities in l are opened in case 1 of $\mathcal{D}_{k,l}$).

Note that in case 1 of $\mathcal{D}_{k,l}$ all the client demand is assigned to the facilities in k while in case 1 of $\mathcal{D}^*_{k,l}$ the demand assigned to facilities in k is the total demand minus 2^{-n^2}, which is assigned to the facilities in $l - l'$. Also note that assignments in k are done evenly for all clients. Thus the last equation yields:

$(E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij} \cap case2^*] - p) + (E_{\mathcal{D}^*_{k,l'}}[\mathcal{E}x_{ij} \cap case1^*] + p) =$(here $p = 2^{-n^2} E_{\mathcal{D}_{k,l}}[\chi(case1)]/|k||C|$)
$E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$.

Case A.2.b

Now, if set $\mathcal{E}$ contains facilities only from $F - l'$ and x_{ij} is an assignment variable of $l - l'$ then, again, $z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1]$. But now $E_{\mathcal{D}_{k,l'}}[\mathcal{E}x_{ij} \cap case1] = 0$ since none of $l - l'$ are assigned any demand when case 1 happens in $\mathcal{D}_{k,l'}$. From the definition of the core we have $z_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij}] = E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case2]$ (in case 1 all assignments to l are zero). We have:

$z_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij} \cap case2]$.

Note that in case $2.b*$ of $\mathcal{D}^*_{k,l}$ the total demand assigned to i is $2^{-n^2} \frac{P[\chi(case1^*)]}{P[\chi(case2.b^*)]}$ less than the total demand assigned to it in the corresponding wrt to q case of $\mathcal{D}_{k,l}$. Thus, again by symmetry of the assignments, the last expression is equal to:

$E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij} \cap case2^*] + r =$(here $r = 2^{-n^2} E_{\mathcal{D}_{k,l}}[\chi(case1)]/|l - l'||C|$)

$E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$.
So once again $z^*_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$.

Case B

Consider the case where $\mathcal{E}$ contains facilities in $l' - l$ and so $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) - E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1]$. By definition of the core $z_{k,l}(\mathcal{E}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}]$. So $z^*_{k,l}(\mathcal{E}) = E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case2]$. On the other hand, $E_{\mathcal{D}^*_{k,l}}[\mathcal{E}] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case1^*] + E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case2^*] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case2^*]$ since $E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case1^*] = 0$ (the facilities in $l' - l$ are always closed in case 1 of $\mathcal{D}^*_{k,l}$). Case 2 of $\mathcal{D}_{k,l}$ and case 2^* of $\mathcal{D}^*_{k,l}$ differ only on the case where $q = F - k - l'$. Since $\mathcal{E}$ contains facilities in $l' - l$ it cannot be the case $q = F - k - l'$. So we have that $E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case2] = E_{\mathcal{D}^*_{k,l}}[\mathcal{E} \cap case2^*]$. So once again we have $z^*_{k,l}(\mathcal{E}) = E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$. The exact same arguments are valid in case we consider $\mathcal{E}x_{ij}$ for any assignment variable x_{ij}.

Case C

For every other set $\mathcal{E}$ and any (i, j) we have $z^*_{k,l}(\mathcal{E}) = z_{k,l}(\mathcal{E}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}]$($z^*_{k,l}(\mathcal{E}x_{ij}) = z_{k,l}(\mathcal{E}x_{ij}) = E_{\mathcal{D}_{k,l}}[\mathcal{E}x_{ij}]$) which is equal to $E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$($E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$) by construction of the distributions $\mathcal{D}^*_{k,l}$ and $\mathcal{D}^*_{k,l}$.

The proof of Proposition 5.6.2 is complete. □

To complete the proof of the claim note that $\mathcal{D}^*_{k,l}$ can also be seen as a distribution over product vectors by associating each mixed integer outcome (y, x) of the experiment with the product vector $f(y, x)$ - recall that $f(y, x)$ is the mixed product encoding. Observe that the expectations $E_{\mathcal{D}^*_{k,l}}[\mathcal{E}]$ and $E_{\mathcal{D}^*_{k,l}}[\mathcal{E}x_{ij}]$ are exactly the expectations of the corresponding components of the product vectors $f(y, x)$.

The proof of Claim 5.6.1 is complete. □

Finally, we show the following

Claim 5.6.2. $\frac{1}{2}(z^*_{k,l} + z^*_{k,l'}) \in \mathrm{conv}(z_{k,l}, z_{k,l'})$.

Proof. We shall actually show that $1/2(z^*_{k,l} + z^*_{k,l'}) = 1/2(z_{k,l} + z_{k,l'})$. Let $\mathcal{E}$ be a set containing facilities only from $F - l'$ with at least one from $l - l'$. Then $z^*_{k,l}(\mathcal{E}) + z^*_{k,l'}(\mathcal{E}) = z_{k,l}(\mathcal{E}) + E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1] + z_{k,l'}(\mathcal{E}) - E_{\mathcal{D}_{k,l'}}[\mathcal{E} \cap case1] = z_{k,l}(\mathcal{E}) + z_{k,l'}(\mathcal{E})$.

Let $\mathcal{E}$ be a set containing facilities only from $F - l$ with at least one from $l' - l$. We have $z^*_{k,l}(\mathcal{E}) + z^*_{k,l'}(\mathcal{E}) = z_{k,l}(\mathcal{E}) - E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1] + z_{k,l'}(\mathcal{E}) + E_{\mathcal{D}_{k,l}}[\mathcal{E} \cap case1] = z_{k,l}(\mathcal{E}) + z_{k,l'}(\mathcal{E})$.

In the remaining cases for the set $\mathcal{E}$ we simply have $z^*_{k,l}(\mathcal{E}) + z^*_{k,l'}(\mathcal{E}) = z_{k,l}(\mathcal{E}) + z_{k,l'}(\mathcal{E})$ by construction. The exact same arguments are valid in case we consider $\mathcal{E}x_{ij}$ for any assignment variable x_{ij}. □

The proof of Lemma 5.6.2 is complete.

6. CONCLUSION

In the context of this thesis we exposed the limitations of linear programming methods for providing satisfactory approximations to assignments problem with restrictions such as capacities. In particular we showed that the unboundedness of the integrality gap of CFL or LBFL relaxations persists even after applying the tightenings of the LS and SA hierarchies. We did so by proving the feasibility of a bad fractional solution for an asymptotically tight number of levels. We also proved that the submodular inequalities do not reduce the integrality gap to constant. Then, while turning our attention to the more general model of extended formulations, we devised a methodology for lower bounding the extension complexity which also serves as a characterization of the extension complexity. We applied our method to derive tight bounds on the size of mixed product relaxations which result also implies tight SA gaps regardless of the initial relaxation. Lastly, we proved similar negative results for families of proper relaxations that capture general configuration LPs. The obtained results answered a number of interesting open questions and conjectures from the relevant literature.

In the recent work of An et al. [7] the first constant factor LP-based approximation algorithm for CFL was given. However, the proposed relaxation is exponential in size and, according to the authors, it is not known to be separable in polynomial time. A natural question that arises is whether there is a polynomially-sized relaxation achieving a constant integrality gap. An interesting direction is that of determining the minimum size of an approximate extended formulation of the CFL polytope, which our results arguably suggest to be exponential. We leave this as an open problem.

Regarding our lower bounding methodology for extended formulations, the proof of our result for mixed product relaxations for CFL made use of a core whose underlying hypergraph is actually a graph and moreover a clique. To generalize this result to product formulations or distributional formulations, or to prove bounds on the extension complexity of other polytopes, we believe that the power of general hypergraphs needs to be exploited. Observe that our methodology requires only the existence of a suitable core, and thus, one could possibly employ probabilistic arguments to prove the existence of suitable hypergraphs of high chromatic number.

In the case of mixed integer polytopes, we leave as an open problem whether the mixed product relaxations are strong enough to simulate any extended formulation, as is the case for product relaxations and 0-1 polytopes.

We also believe that it would be interesting to revisit polytopes, whose extension complexity has been shown to be large, and provide independent proofs using our method, ideally by improving on the known bounds. Moreover, as we showed in Chapter 4.3 for CFL using product or distributional formulations one can provide lower bounds as well and this can be of help in settling the extension complexity.

References

[1] Karen Aardal, Yves Pochet, and Laurence A. Wolsey. Capacitated facility location: Valid inequalities and facets. *Mathematics of Operations Research*, 20:562–582, 1995.

[2] Karen Aardal, Yves Pochet, and Laurence A. Wolsey. Erratum: Capacitated facility location: Valid inequalities and facets. *Mathematics of Operations Research*, 21:253–256, 1996.

[3] Zoë Abrams, Adam Meyerson, Kamesh Munagala, and Serge Plotkin. The integrality gap of capacitated facility location. Technical Report CMU-CS-02-199, Carnegie Mellon University, 2002.

[4] Ankit Aggarwal, Anand Louis, Manisha Bansal, Naveen Garg, Neelima Gupta, Shubham Gupta, and Surabhi Jain. A 3-approximation algorithm for the facility location problem with uniform capacities. *Mathematical Programming*, 141(1-2):527–547, 2013.

[5] Sara Ahmadian and Chaitanya Swamy. Improved approximation guarantees for lower-bounded facility location. *CoRR*, abs/1104.3128, 2011.

[6] Hyung-Chan An, Aditya Bhaskara, and Ola Svensson. Centrality of trees for capacitated k-center. *CoRR*, abs/1304.2983, 2013.

[7] Hyung-Chan An, Mohit Singh, and Ola Svensson. LP-based algorithms for capacitated facility location. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 256–265. IEEE Computer Society, 2014.

[8] Sanjeev Arora, Béla Bollobás, and László Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 313–322, Washington, DC, USA, 2002. IEEE Computer Society.

[9] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(1):19–51, 2006.

[10] Egon Balas, Sebastián Ceria, and Gérard Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58(3):295–324, February 1993.

[11] Manisha Bansal, Naveen Garg, and Neelima Gupta. A 5-approximation for capacitated facility location. In Leah Epstein and Paolo Ferragina, editors, *Algorithms { ESA 2012*, volume 7501 of *Lecture Notes in Computer Science*, pages 133–144. Springer Berlin Heidelberg, 2012.

[12] Nikhil Bansal and Maxim Sviridenko. The Santa Claus problem. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 31–40, New York, NY, USA, 2006. ACM.

[13] Mohammadhossein Bateni and Mohammadtaghi Hajiaghayi. Assignment problem in

content distribution networks: Unsplittable hard-capacitated facility location. *ACM Transactions on Algorithms*, 8(3):20:1–20:19, July 2012.

[14] Daniel Bienstock. Approximate formulations for 0-1 knapsack sets. *Oper. Res. Lett.*, 36(3):317–320, 2008.

[15] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation limits of linear programs (beyond hierarchies). In *FOCS*, pages 480–489, 2012.

[16] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. In *FOCS*, pages 688–697. IEEE Computer Society, 2013.

[17] Gábor Braun and Sebastian Pokutta. The matching polytope does not admit fully-polynomial size relaxation schemes. *CoRR*, abs/1403.6710, 2014.

[18] Gábor Braun, Sebastian Pokutta, and Daniel Zink. LP and SDP inapproximability of combinatorial problems. *CoRR*, abs/1410.8816, 2014.

[19] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proc. STOC*, pages 161–170, 2013.

[20] Robert D. Carr, Lisa Fleischer, Vitus J. Leung, and Cynthia A. Phillips. Strengthening integrality gaps for capacitated network design and covering problems. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2000, January 9-11, 2000, San Francisco, CA, USA*, pages 106–115, 2000.

[21] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 350–359, 2013.

[22] Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 378–388, 1999.

[23] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing*, STOC '09, pages 283–292, New York, NY, USA, 2009. ACM.

[24] Fabián A. Chudak and David B. Shmoys. Improved approximation algorithms for a capacitated facility location problem. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '99, pages 875–876, Philadelphia, PA, USA, 1999. Society for Industrial and Applied Mathematics.

[25] Fabián A. Chudak and David P. Williamson. Improved approximation algorithms for capacitated facility location problems. *Mathematical Programming*, 102(2):207–222, March 2005.

[26] Gérard Cornuéjols. Valid inequalities for mixed integer linear programs. *Mathematical Programming*, 112(1):3–44, 2008.

[27] Marek Cygan, MohammadTaghi Hajiaghayi, and Samir Khuller. LP rounding for k-centers with non-uniform hard capacities. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 273–282, 2012.

[28] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the eighteenth annual ACM-SIAM symposium*

on Discrete algorithms, SODA '07, pages 53–61, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[29] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds. In *Proceedings of the 44th Symposium on Theory of Computing*, STOC '12, pages 95–106, New York, NY, USA, 2012. ACM.

[30] Konstantinos Georgiou and Avner Magen. Limitations of the Sherali-Adams lift and project system: Compromising local and global arguments. Technical Report CSRG-587, University of Toronto, 2008.

[31] Michel X. Goemans and Levent Tunçel. When does the positive semidefiniteness constraint help in lifting procedures? *Mathematics of Operations Research*, 26(4):796–815, 2001.

[32] S. Guha, A. Meyerson, and K. Munagala. Hierarchical placement and network design problems. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 603–, Washington, DC, USA, 2000. IEEE Computer Society.

[33] Dorit S. Hochbaum. Heuristics for the fixed cost median problem. *Mathematical Programming*, 22:148–162, 1982.

[34] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6):795–824, November 2003.

[35] Kamal Jain and Vijay V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. *Journal of the ACM*, 48(2):274–296, March 2001.

[36] Stasys Jukna. *Boolean Function Complexity - Advances and Frontiers*, volume 27 of *Algorithms and combinatorics*. Springer, 2012.

[37] Volker Kaibel and Stefan Weltge. Lower bounds on the sizes of integer programs without additional variables. In Jon Lee and Jens Vygen, editors, *IPCO*, volume 8494 of *Lecture Notes in Computer Science*, pages 321–332. Springer, 2014.

[38] D. R. Karger and M. Minkoff. Building Steiner trees with incomplete global knowledge. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 613–, Washington, DC, USA, 2000. IEEE Computer Society.

[39] Stavros G. Kolliopoulos and Yannis Moysoglou. Extended formulation lower bounds via hypergraph coloring? *to appear in 32nd International Symposium on Theoretical Aspects of Computer Science, STACS 2015, Munich, Germany, March 4th -7th*, 2015.

[40] Stavros G. Kolliopoulos and Yannis Moysoglou. Exponential lower bounds on the size of approximate formulations in the natural encoding for capacitated facility location. *CoRR*, abs/1312.1819, 2013.

[41] Stavros G. Kolliopoulos and Yannis Moysoglou. Sherali-Adams gaps, flow-cover inequalities and generalized configurations for capacity-constrained facility location. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4-6, 2014, Barcelona, Spain*, pages 297–312, 2014.

[42] Madhukar R. Korupolu, C. Greg Plaxton, and Rajmohan Rajaraman. Analysis of a

local search heuristic for facility location problems. In *Proceedings of the ninth annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '98, pages 1–10, Philadelphia, PA, USA, 1998. Society for Industrial and Applied Mathematics.

[43] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th International IPCO Conference on Integer Programming and Combinatorial Optimization*, pages 293–303, London, UK, UK, 2001. Springer-Verlag.

[44] Monique Laurent. A Comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0–1 Programming. *Mathematics of Operations Research*, 28(3):470–496, July 2003.

[45] Janny M. Y. Leung and Thomas L. Magnanti. Valid inequalities and facets of the capacitated plant location problem. *Mathematical Programming*, 44(1-3):271–291, 1989.

[46] Retsef Levi, David B. Shmoys, and Chaitanya Swamy. LP-based approximation algorithms for capacitated facility location. *Mathematical Programming*, 131(1-2):365–379, 2012. Preliminary version in Proc. IPCO 2004.

[47] Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45–58, January 2013.

[48] Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Proc. 45th STOC*, pages 901–910. ACM, 2013.

[49] Lazlo Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[50] Mohammad Mahdian and Martin Pál. Universal facility location. In Giuseppe Battista and Uri Zwick, editors, *Algorithms - ESA 2003*, volume 2832 of *Lecture Notes in Computer Science*, pages 409–421. Springer Berlin Heidelberg, 2003.

[51] Mohammad Mahdian, Yinyu Ye, and Jiawei Zhang. Approximation algorithms for metric facility location problems. *SIAM Journal on Computing*, 36(2):411–432, August 2006.

[52] Éva Tardos Martin Pál and Tom Wexler. Facility location with nonuniform hard capacities. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pages 329–, Washington, DC, USA, 2001. IEEE Computer Society.

[53] Claire Mathieu and Alistair Sinclair. Sherali-Adams relaxations of the matching polytope. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 293–302, New York, NY, USA, 2009. ACM.

[54] Sebastian Pokutta and Mathieu Van Vyve. A note on the extension complexity of the knapsack polytope. *Oper. Res. Lett.*, 41(4):347–350, 2013.

[55] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 475–484, 1997.

[56] Thomas Rothvoß. The matching polytope has exponential extension complexity. *CoRR*, abs/1311.2369, 2013.

[57] Thomas Rothvoß. The matching polytope has exponential extension complexity. In *Proc. STOC*, pages 263–272, 2014.

[58] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of Vertex Cover and Max Cut. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 302–310, New York, NY, USA, 2007. ACM.

[59] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal of Discrete Mathematics*, 3(3):411–430, 1990.

[60] David B. Shmoys, Éva Tardos, and Karen I. Aardal. Approximation algorithms for facility location problems. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 265–274, 1997.

[61] Ola Svensson. Santa Claus schedules jobs on unrelated machines. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 617–626, New York, NY, USA, 2011. ACM.

[62] Z. Svitkina. Lower-bounded facility location. In *Proceedings of the 19th ACM-SIAM Symposium on Discrete Algorithms*, pages 1154–1163, 2008.

[63] Madhur Tulsiani. Lovász-Schrijver reformulation. In J. J. Cochran, Jr. L. Anthony Cox, P. Keskinocak, J. P. Kharoufeh, and J. Cole Smith, editors, *Encyclopedia of Operations Research and Management Science*. John Wiley and Sons, 2011.

[64] Jens Vygen. Approximation algorithms for facility location problems (Lecture Notes). Report 05950-OR, Research Institute for Discrete Mathematics, University of Bonn, 2005. URL: www.or.uni-bonn.de/∼vygen/files/fl.pdf.

[65] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2011.

[66] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441 – 466, 1991.

[67] Jiawei Zhang, Bo Chen, and Yinyu Ye. A multi-exchange local search algorithm for the capacitated facility location problem. In Daniel Bienstock and George Nemhauser, editors, *Integer Programming and Combinatorial Optimization*, volume 3064 of *Lecture Notes in Computer Science*, pages 219–233. Springer Berlin Heidelberg, 2004.

[68] James R. Lee and Prasad Raghavendra and David Steurer. Lower Bounds on the Size of Semidefinite Programming Relaxations. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC 2015, Portland, OR, USA, June 14-17, 2015.

[69] Subhash Khot. On the Power of Unique 2-Prover 1-Round Games. In *Proceedings of the 17th Annual IEEE Conference on Computational Complexity*, Montréal, Québec, Canada, May 21-24, 2002.

[70] Gábor Braun and Sebastian Pokutta and Daniel Zink. Inapproximability of Combinatorial Problems via Small LPs and SDPs. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC 2015, Portland, OR, USA, June 14-17, 2015.

[71] Christos H. Papadimitriou and Mihalis Yannakakis. The Complexity of Facets (and Some Facets of Complexity). In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, May 5-7, 1982, San Francisco, California, USA.

[72] Martin Henk and Jürgen Richter-Gebert and Günter M. Ziegler. Basic Properties of Convex Polytopes. Chapman and Hall/CRC, 2004, San Francisco, California, USA.

[73] Tim Carnes and David B. Shmoys. Primal-Dual Schema for Capacitated Covering Problems. In *Integer Programming and Combinatorial Optimization, 13th International Conference*, IPCO 2008, Bertinoro, Italy, May 26-28, 2008, Proceedings.

[74] Stavros G. Kolliopoulos and Yannis Moysoglou. Tight bounds on the Lovász-Schrijver rank for approximate Capacitated Facility Location. Manuscrip, 2013.

[75] Stavros G. Kolliopoulos and Yannis Moysoglou. Integrality gaps for strengthened linear relaxations of Capacitated Facility Location. To appear in *Mathematical Programming*, 43 pages, Springer, 2015.