

A Polyhedral Approach to Grouping Problems

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op
beveling van de Rector Magnificus, Prof. Mr. M. J. Cohen, volgens het besluit
van het College van Decanen, in het openbaar te verdedigen op vrijdag 9
december 1996 om 11.00 uur

door

Maarten Ooster

Promotores:

Prof.dr.ir.drs. O.J. Vrieze

Prof.dr.ir. A.W.J. Kolen

Prof.dr. Y. Crama (Université de Liège, B.)

Leden van de beoordelingscommissie:

Prof.dr. H.J.M. Peeters (voorzitter)

Prof.dr. E. Balas (Carnegie Mellon University, U.S.A.)

Dr. S. van Hoesel

Prof.dr. S.H. Tijs

Prof.dr. L.A. Wolsey (Université Catholique de Louvain, B.)

Preface

To many people, mathematics may look like series of true statements, with the corresponding irrefutable proofs, ordered in a perfectly logical manner. Perhaps one would expect that the process of writing a mathematical thesis reflects this image. My experiences were quite different. To me it was a time of guesses and doubts, of conjectures, for which proofs and refutations were proposed to shoot holes in them. A period of trying different approaches, and re-evaluations in order to give some coherence to the process. A process involving emotions ranging from victorious joy to levels that I do not wish to recall right now. Looking back, I conclude that acquiring mathematical knowledge is not very different from the rest of life. So it will come to no surprise that like in real life, for completing this thesis I owe a great deal to the support of many other people. I owe so much to so many people that it is impossible to tell to whom I owe most.

When I started working at Maastricht University, it was Yves Crama who introduced me to a mix of manufacturing problems and polyhedral theory. He succeeded in getting me interested, and together with the enthusiastic and optimistic support of Koos Vrieze, and the enthusiastic and critical support of Antoon Kolen, he created a research environment that has been very beneficial to me.

Furthermore, I wish to thank Egon Balas for our work on the Chapters 2 and 5, for introducing me to many aspects of polyhedral theory, and for triggering the development of my view on mathematics.

I also want to express my gratitude to the members of the assessment committee: Hans Peeters, Egon Balas, Stan van Hoesel, Stef Tijs, and Laurence Wolsey. Even I have to admit that in many cases you can not tell from reading my proofs that I always aim for elegance.

Many thanks are also due to my colleagues at the Department of Mathematics and the Department of Quantitative Economics of Maastricht University. In particular to my room-mates: Frits Spijksma, Jeroen Kuipers,

Pim Martens, and Jeroen Rutten. They have shown me many useful things. For example, how to derive interesting results in chaos, and in order; or the advantages of doing research with a more direct relevance for society. And they have shown me what talent is about.

I am aware that there has been much more to writing this thesis than mathematics. As a consequence, many more people have been involved in this thesis than I mention here. Likewise, I will show that there are many more ways to show gratitude.

To many people, mathematics looks like a series of abstract concepts, but the corresponding intuitive proofs, which are a perfectly logical manner. Perhaps one would expect that the process of writing a mathematical thesis reflects this image. My experience was quite different. To me it was a time of guesses and doubts, of conjectures, for which proofs and refutations were proposed to shoot holes in them. A period of trying different approaches, and re-evaluations in order to give some coherence to the process. A process involving emotions ranging from victories up to bewilderment that I do not want to recall right now. Looking back, I conclude that acquiring mathematical knowledge is not very different from the rest of life. So it will come to no surprise that life in real life, for completing this thesis I owe a great deal to the support of many other people. I owe so much to so many people that it is impossible to tell to whom I owe most.

When I started working at Maastricht University, it was Yves Cramis who introduced me to a mix of manufacturing problems and polyhedral theory. He succeeded in getting me interested, and together with the enthusiastic and optimistic support of Kees Vries, and the enthusiastic and critical support of Antoon Kolen, he created a research environment that has been very beneficial to me.

Furthermore, I wish to thank Egon Balas for our work on the Chapters 2 and 3, for introducing me to many aspects of polyhedral theory, and for triggering the development of my view on mathematics.

I also want to express my gratitude to the members of the assessment committee: Hans Posters, Egon Balas, Stas van Hoesel, Stet Tijl, and Jan-Thomas Wolsey. Even I have to admit that in many cases you can not tell from reading my proofs that I always aim for elegance.

Many thanks are also due to my colleagues at the Department of Mathematics and the Department of Quantitative Economics of Maastricht University. In particular to my room-mates: Frits Spielman, Jeroen Kuipers,

Contents

Preface

iii

1 Introduction

1

1.1 Example: The Vending Machine Payment	2
1.2 Definitions and Terminology	8
1.3 Outline	12

2 On the Dimension of Projected Polyhedra

13

2.1 Introduction	14
2.2 The Main Result	15
2.3 Consequences	19

3 Sequential Lifting Procedures

23

3.1 Introduction and Terminology	24
3.2 Comparison of the Procedures	27
3.3 The Dimension Gap and Equality Lifting	31
3.4 Extended Lifting	36
3.5 Complete lifting	40
3.6 Strict Sequential Lifting	44
3.7 An Example on the Cell Formation Polytope	46
3.8 Polarity, Duality, and Lifting	49

4 On the Strength of Inequalities

51

4.1 Introduction	52
4.2 Worst Case Analysis for Zero-Identical Relaxation Polyhedra	54
4.3 Some Examples	57

5	The Cycle Polytope of a Directed Graph	61
5.1	Introduction	62
5.2	The Cycle Polytope	64
5.3	The Main Result and Some of Its Consequences	69
5.4	Cloning	74
5.5	Direct Lifting Procedures	76
5.6	Some Complete Descriptions for Small n	83
6	Models for the Cell Formation Problem	89
6.1	Introduction	90
6.2	Review of Existing Mathematical Programming Formulations	94
6.3	A new Mathematical Programming Model	99
6.4	Numerical examples	107
6.5	Conclusions	114
7	The Polytope of the Cell Formation Problem	117
7.1	Introduction	118
7.2	Properties of facet defining inequalities	120
7.3	Facet Defining Inequalities	124
7.4	Lifting Theorems	139
8	Patching Procedures for the Polytope of the Cell Formation Problem	151
8.1	Introduction	152
8.2	Covering, Patching, and Recognizing	155
8.3	Intersection Patching Procedures	159
8.4	Disjoint Patching Procedures	163
8.5	A Recursive Class of Facet Defining Inequalities	169
	References	173
	Samenvatting	181
	Curriculum Vitae	183

Chapter 1

Introduction

When I started the quest for a PhD-thesis at the University of Limburg in June 1990, I joined a group of researchers focusing on combinatorial aspects of problems arising in flexible manufacturing systems. Their efforts culminated in a series of publications, and eventually theses ([Spieksma, 1992], [Oerlemans, 1992]), on different topics concerning scheduling and set-up problems in the highly automated environment of flexible manufacturing systems. In general, for production systems like these, it is extremely difficult to schedule the processing of every part in the right time on the right machine with the right tools. Therefore, it could be advantageous to split up the production process into smaller and relatively independent subprocesses, just to make the tooling, loading, and scheduling problems more tractable. So, I started studying the possibilities of forming production cells. The choice of applying polyhedral techniques was based on the observation that these techniques are efficient for other partitioning problems. Furthermore, they were in line with ongoing research within our group.

However, the central topic of this study is not cell formation, but the relations between polyhedra. The reason is that the formation of cells is not just one grouping problem, but it entails a wide variety of slightly different problems. For example, sometimes we have to take into account additional constraints; limits on the number of groups, the number of machines in one group, the number of parts in one group; pairs of machines that have to be in the same group; the possibility to buy additional machines; or alternative ways to process a part. It is easy to extend this list. Clearly, the value of deriving theoretical results for a specific problem is dependent on the possibility to transform those results into useful information about variants

of the original problem. As a consequence the emphasis of the study shifted from the formation of production cells towards relating polyhedra.

To get acquainted with the basic ideas of applying polyhedral theory to combinatorial problems, we discuss a simple example in the next section. The most important concepts will be introduced formally in Section 1.2. In the final section of this introduction a brief outline of this thesis is presented.

1.1 Example: The Vending Machine Payment

In this section we discuss a small instance of a combinatorial problem, that in real life not many people would be inclined to tackle with polyhedral techniques. It is presented here just for the sake of illustrating the basic concepts of polyhedral theory applied to combinatorial problems.

Example 1.1.1 *In the dungeons of an american university a vending machine provides foods, among which are Swiss Rolls. Suppose we want to purchase a pack of Rolls. This purchase should cost us 40 cents (it does not, but that is a different discussion).*

In our wallet we have 3 quarters, 5 dimes, and 10 cents. A little red light indicates a lack of change. Not willing to tip for this bad service, we decide to pay exact. Is this possible? If so, we would prefer to use as few coins as possible. How many of each type of coins should we use?

This example is an instance of the abstract problem of paying exactly a certain amount of money given a set of coins and banknotes, such that the total number of coins and notes that are used to pay is minimal. In our case a solution of the instance consists of three numbers. Let us denote the number of quarters as $x_{quarters}$, the number of dimes as x_{dimes} , and the number of cents as x_{cents} . In a feasible solution, $x_{quarters}$ can only have the values 0, 1, 2, or 3. Similarly, $x_{dimes} \in \{0, 1, 2, 3, 4, 5\}$, and $x_{cents} \in \{0, 1, \dots, 10\}$. These integer constraints express that we can only use the coins we have. To be feasible, a solution also has to satisfy the following linear constraint. Let us call it the payment equality. It expresses that we are paying exactly 40 cents:

$$25x_{quarters} + 10x_{dimes} + x_{cents} = 40.$$

An optimal solution is one of the feasible solutions with the best objective value. In our case the objective is to minimize the number of coins used. That is:

$$\text{minimize } x_{quarters} + x_{dimes} + x_{cents}.$$

To determine an optimal solution we need to evaluate all solutions, preferably in the way requiring the least effort. Checking all possible 264 combinations of values for our three variables is an option, but it is not likely to be the most efficient one. A more promising method is to split up the instance into subinstances that we can analyze more easily. For example, we could split up our instance into four subinstances, one for each possible value of $x_{quarters}$. Obviously, the subinstances for which the value of the used quarters exceeds 40 cents do not have a feasible solution and can be discarded. Implicitly, we discard half of all possible combinations of values for the three variables.

Evaluating a subinstance is slightly more work than evaluating a solution, but it is worth the effort. We can cut back on the number of subinstances to be evaluated by improving the evaluations. Up to now the evaluation was a simple check of feasibility: if the remaining coins should add up to a negative amount, we discard the subinstance. We could do more, like deriving an upper bound on the optimal value of the instance and a lower bound on the optimal value of the subinstance. If for a subinstance the lower bound exceeds the upper bound, we can discard the subinstance. Usually, trivial bounds, like zero and the number of coins available, are not of much help. To derive useful bounds we have to make an extra effort.

As for the upper bounds, any feasible solution implies a nontrivial upper bound for all subinstances. So we could design a simple algorithm that provides us in reasonable time with a reasonable chance with a feasible solution that is fairly good (though maybe not optimal). Such an algorithm is called a heuristic. Deriving a nontrivial lower bound is harder, since not every infeasible solution implies a lower bound. What we can do is try to solve a relaxation of the instance. We should relax or even omit some of the constraints in such a way that we get an instance of a problem that can be solved more easily.

In our case I would suggest to relax the integer constraints on $x_{quarters}$, x_{dimes} , and x_{cents} , by replacing them with upper and lower bounds on these variables. Let P_{coins} denote the set of solutions satisfying the payment equality and the upper and lower bounds on the variables:

$$P_{coins} = \left\{ \begin{array}{l} x_{quarters} \in \mathbb{R} \\ x_{dimes} \in \mathbb{R} \\ x_{cents} \in \mathbb{R} \end{array} \middle| \begin{array}{l} 25x_{quarters} + 10x_{dimes} + x_{cents} = 40 \\ 0 \leq x_{quarters} \leq 3 \\ 0 \leq x_{dimes} \leq 5 \\ 0 \leq x_{cents} \leq 10 \end{array} \right\}.$$

The task of optimizing our objective function over this set is called the *linear relaxation* of the original instance. It is linear, because all the constraints as

well as the objective are linear expressions in the variables. In our case, the linear relaxation can be formulated as follows:

$$\begin{array}{ll} \text{minimize} & x_{quarters} + x_{dimes} + x_{cents} \\ \text{subject to} & (x_{quarters}, x_{dimes}, x_{cents}) \in P_{coins} \end{array}$$

The set P_{coins} is a *polytope*. Optimizing a linear function over a polytope implies that there is always at least one corner point of the polytope that is an optimal solution. There exists a wide variety of computer packages designed to select an optimal corner point efficiently.

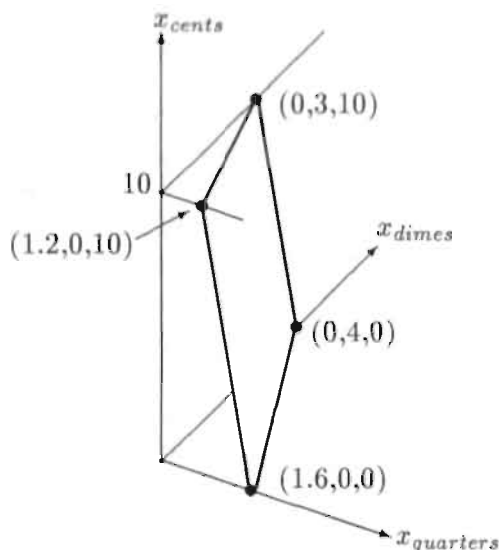


Figure 1.1: A geometrical representation of the polytope P_{coins} .

The set P_{coins} is represented geometrically in Figure 1.1. Since P_{coins} is a polytope, and the objective function is linear in the variables, we only need to check the corner points of P_{coins} to find an optimal solution. In this case inspection is sufficient to find an optimal solution: the corner point for which $x_{quarters} = 1.6$ and $x_{dimes} = x_{cents} = 0$. The corresponding objective value, the sum of the variables, is 1.6. This solution is optimal for the linear relaxation, but it is not a feasible solution of the original instance since it is not integral. Still, it gives us some nontrivial information: if there exists a feasible solution, then it involves at least two coins.

Evaluating an instance by solving the linear relaxation is slightly more work than before, but maybe it is worth it. We could cut back on the number of instances to be evaluated by improving the linear relaxations. This is where polyhedral theory comes in. From now on we are not going to bother about smart strategies to split up instances, or about clever heuristics, but instead we are going to focus on deriving lower bounds for an instance applying polyhedral theory.

Often it is easier to solve a problem defined in two dimensions than one defined in three. Notice that if we know two of our numbers $x_{quarters}$, x_{dimes} , and x_{cents} , we can easily compute what the third number should be. Let us get rid of the cents by rewriting the payment equality to $x_{cents} = 40 - 25x_{quarters} - 10x_{dimes}$ and by replacing x_{cents} with $40 - 25x_{quarters} - 10x_{dimes}$ in the other expressions. Now if we solve the following instance:

$$\begin{array}{ll} \text{minimize} & 40 - 24x_{quarters} - 9x_{dimes} \\ \text{subject to} & x_{quarters} \in \{0, \dots, 3\} \\ & x_{dimes} \in \{0, \dots, 5\} \\ & 25x_{quarters} + 10x_{dimes} \in \{30, \dots, 40\} \end{array}$$

we implicitly solve the original instance, since, given the optimal solution of the simplified instance, it is easy to construct the optimal solution of the original instance.

Now let us study the linear relaxation of the simplified instance. We denote the set of feasible solutions of this linear relaxation of our instance as Q_{coins} . Clearly, Q_{coins} is a polytope (see Figure 1.2):

$$Q_{coins} = \left\{ \begin{array}{l} x_{quarters} \in \mathbb{R} \\ x_{dimes} \in \mathbb{R} \end{array} \left| \begin{array}{l} 0 \leq x_{quarters} \leq 3 \\ 0 \leq x_{dimes} \leq 5 \\ 30 \leq 25x_{quarters} + 10x_{dimes} \leq 40 \end{array} \right. \right\}.$$

In fact, Q_{coins} is a *projection* of P_{coins} on the plane where $x_{cents} = 0$. And it is a very special projection, since for each corner point in Q_{coins} there is exactly one corner point in P_{coins} , and the other way round. Among other things, necessary and sufficient conditions for this property are discussed in the next chapter of this thesis. Obviously, projective relations can be useful to abstract information about one polytope from the other.

The linear relaxation of the simplified instance can be formulated as follows:

$$\begin{array}{ll} \text{minimize} & 40 - 24x_{quarters} - 9x_{dimes} \\ \text{subject to} & (x_{quarters}, x_{dimes}) \in Q_{coins} \end{array}$$

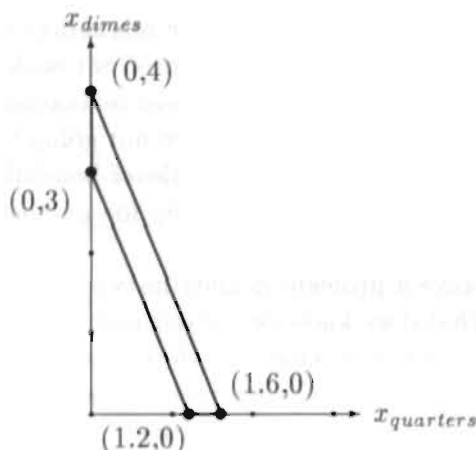


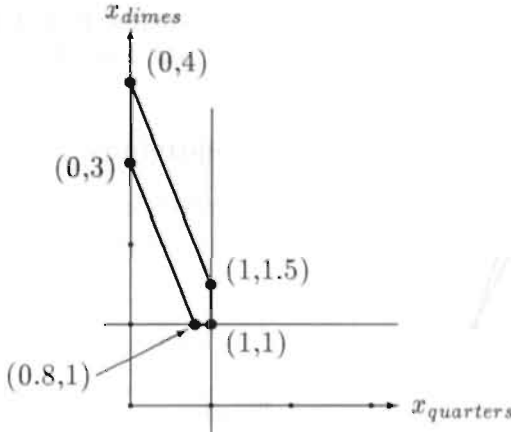
Figure 1.2: A geometrical representation of the polytope Q_{coins} .

Clearly, the objective function is linear in the variables. So, we only need to check the corner points of Q_{coins} to find an optimal solution: the corner point for which $x_{\text{quarters}} = 1.6$ and $x_{\text{dimes}} = 0$. The corresponding objective value is 1.6.

How can we improve this result? Looking at Figure 1.2 it is obvious that we could chop off pieces of Q_{coins} without excluding a feasible integer solution. For example, we can reason that it is allowed to add the constraint $x_{\text{quarters}} \leq 1$ to the constraints that define Q_{coins} . All feasible integer solutions satisfy this inequality, but the solution that is optimal in Q_{coins} does not. Geographically, adding the inequality can be interpreted as cutting off a piece of Q_{coins} . Therefore, this inequality is called a *cutting plane*. But it is not the only cutting plane we could think of. We could also add the constraint $x_{\text{dimes}} \geq 1$ to the constraints that define Q_{coins} . Which cutting plane is better is hard to say. We discuss the strength of cutting planes in Chapter 4 of this thesis.

Let $\bar{Q}_{\text{coins}}$ be the set of solutions in Q_{coins} that satisfy both cutting planes discussed in the paragraph above; see Figure 1.3 for a geometrical representation. Optimizing over $\bar{Q}_{\text{coins}}$, we find as an optimal solution $x_{\text{quarters}} = 1$, $x_{\text{dimes}} = 1.5$ and the objective value is 2.5. This implies that we have to use at least three coins: adding the cutting planes improved our lower bound!

But how do we go from here? How to come up with new cutting planes? Do we just have to wait for inspiration to formulate new cutting planes?

Figure 1.3: The polytope $\bar{Q}_{coins}$.

No, not necessarily. For instance, sometimes we can derive a cutting plane by improving an inequality we already know. Consider for example the inequality $25x_{quarters} + 10x_{dimes} \geq 30$, or equivalently: $2\frac{1}{2}x_{quarters} + x_{dimes} \geq 3$. The variable $x_{quarters}$ is either 0 or 1. If $x_{quarters} = 0$, then the inequality implies $x_{dimes} \geq 3$. If on the other hand $x_{quarters} = 1$, then the inequality implies $x_{dimes} \geq \frac{1}{2}$. Can we change the inequality such that it stays valid for all integer solutions and implies $x_{dimes} \geq 1$ if $x_{quarters} = 1$? Sure, by lowering the coefficient of $x_{quarters}$ in the inequality from $2\frac{1}{2}$ to 2: $2x_{quarters} + x_{dimes} \geq 3$. Similarly, we can improve the inequality $25x_{quarters} + 10x_{dimes} \leq 40$, or equivalently: $2\frac{1}{2}x_{quarters} + x_{dimes} \leq 4$. We can lift $2\frac{1}{2}$ to 3 to get: $3x_{quarters} + x_{dimes} \leq 4$. This kind of improvement techniques are called *lifting procedures* and we discuss them in Chapter 3 of this thesis.

Actually, together with $x_{quarters} \geq 0$, the two improved inequalities describe the convex hull H_{coins} of the integer solutions of the simplified instance, as is illustrated in Figure 1.4. All corner points of H_{coins} are integer solutions. So, optimizing any linear function over H_{coins} yields an integer solution, which is therefore also an optimal solution of the simplified instance and implies an optimal solution for the original instance. No further analysis is needed. Only one instance has to be evaluated to solve the original instance.

There are different possibilities to describe H_{coins} . Preferably, we would like to use as few constraints as possible. Clearly, for H_{coins} the minimum number of inequalities to describe H_{coins} is three, one for each of the three

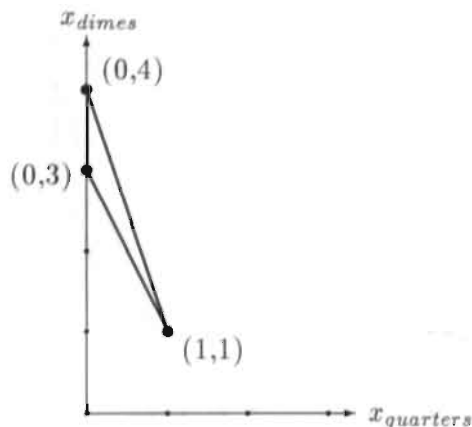


Figure 1.4: The convex hull of the integer solutions, H_{coins} .

line-pieces forming the boundary of H_{coins} . These line-pieces are called facets. Facets have our special attention, since for each facet we need to have an inequality in the minimal description of H_{coins} .

Deriving inequalities that define facets for H_{coins} is not very interesting unless we can generalize the inequalities for the problem of paying exactly a certain amount of money, given a set of coins and banknotes. For the cell formation problem such a generalization is possible. In Chapter 6 a general model for the cell formation problem is formulated and discussed. Valid and facet defining inequalities for the polytope that is the convex hull of all solutions to the model are studied in Chapter 7 and Chapter 8.

1.2 Definitions and Terminology

In the remaining chapters of this study the reader is assumed to be familiar with the basic concepts of polyhedral theory. In this section, only the most important concepts are defined. For a thorough discussion see for example [Rockafellar, 1970], [Schrijver, 1986], and [Nemhauser and Wolsey, 1988].

Definition 1.1 A **polyhedron** is the set of points that satisfy all the inequalities of a finite set of linear inequalities.

Definition 1.2 A polyhedron $P \subseteq \mathbb{R}^n$ is **bounded** if there exists a positive number $M \in \mathbb{R}$ such that $P \subseteq \{x \in \mathbb{R}^n \mid -M \leq x_i \leq M \text{ for all } i \in \{1 \dots n\}\}$.

Definition 1.3 A **polytope** is a bounded polyhedron.

Definition 1.4 A polyhedron is of **dimension** p , denoted as $\dim(P) = p$, if and only if the maximum number of affinely independent points in P is $p + 1$.

As a consequence, an empty polyhedron is of dimension -1 .

Definition 1.5 A constraint is called **valid** for a set S if all points in S satisfy the constraint.

Definition 1.6 A **face** of a polyhedron P is a subset F of P for which there is a linear inequality, valid for P , such that a point satisfying that inequality at equality is in P if and only if it is in F .

For every linear inequality that is valid for a polyhedron P , there is a set F of all points in P that satisfy the inequality at equality. This set F is a face of P and the inequality is said to *define* F . So, an inequality defines a face of P if and only if the inequality is valid for P .

The polyhedron P itself is a face of P , defined by inequalities equivalent to $0 \leq 0$. The empty set is a face of all polyhedra. To see this, notice that the inequality $0 \leq 1$ is valid for all polyhedra, while there is no point satisfying this inequality at equality.

Definition 1.7 A face of P is called a **proper face** of P if it is nonempty and not identical to P .

If a face of P is not proper, it is called **improper**. If F is a proper face of P , then $0 \leq \dim(F) \leq \dim(P) - 1$. A polyhedron P has no proper faces if and only if P is a subspace.

Definition 1.8 A nonempty face of lowest dimension is called an **extreme face**.

In optimizing a linear objective function over a polyhedron P , the existence of an optimal solution implies the existence of an extreme face such that all points in the face are optimal.

An extreme face is not necessarily a proper face. If a polyhedron consists of a subspace, then the polyhedron has only one extreme face, namely the polyhedron itself.

Definition 1.9 *A face of dimension 0 is called an extreme point, or equivalently a corner point.*

An extreme point is always an extreme face. If P is a polytope, then its extreme faces are always points. Optimizing a linear objective function over a polytope P , there exists an extreme point that is optimal.

Definition 1.10 *A face of dimension $\dim(P) - 1$ is called a facet.*

Definition 1.11 *A facial description of a polyhedron is a set of valid inequalities such that for every valid inequality, there is a linear combination of the inequalities in the set that is equal to the inequality up to multiplication by a nonnegative scalar.*

For each face containing P , two linear inequalities defining that face are needed in the facial description of P . For each facet of P , a linear inequality defining that face is needed in the facial description of P . In a facial description, for every face F all inequalities defining F can be generated by taking a nonnegative weighted sum of the inequalities in the facial description. In literature, a description of a polyhedron P by linear inequalities is often intuitively understood to be a set of valid inequalities such that every point not belonging to P violates at least one inequality. In many cases that boils down to a facial description. But not always. The difference is crucial for some of the theorems in Chapter 3. For example, there is a difference in case P is of dimension zero. Then, the empty set is its only facet, and an inequality defining the empty set is needed for a facial description of P ! More general, if P is a subspace, an inequality defining the empty set is necessary for a facial description of the subspace. Furthermore, in case P is the empty set, all inequalities define P as a face of P itself. To be able to generate all inequalities, at least $n + 2$ inequalities are needed.

Studying the description of a polyhedron $P \subseteq \mathbb{R}^n$ implies studying the set of all inequalities valid for P . This set can be represented in a $(n + 1)$ -dimensional space. There is one variable for every coefficient in the inequality, and a variable for the right hand side. This way, the set of inequalities valid for P is a polyhedron. A special property of this polyhedron is that it is closed under multiplication with a nonnegative scalar.

Definition 1.12 *A cone is a set which is closed under nonnegative scalar multiplication.*

Convex cones are often easier to analyse than arbitrary polyhedra. For example, the polar of a closed nonempty convex cone can be defined as follows.

Definition 1.13 *The polar K^0 of a closed convex cone $K \subset \mathbb{R}^n$ is the closed convex cone defined as:*

$$K^0 := \{(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i \alpha_i \leq 0 \quad \forall (x_1, \dots, x_n) \in K\}.$$

Let K be a closed nonempty convex cone, and K^0 the polar of K . Then, K is the polar of K^0 . Some characteristics of a polar pair of closed nonempty convex cones K and K^0 are "polar". The following two examples are important. To start with, the extreme face of K^0 has dimension $n - \dim(K)$. Secondly, consider a face F^0 of K^0 of one dimension higher than the extreme face of K^0 . The dimension of F^0 equals $n - \dim(K) + 1$. An element on F^0 corresponds to an inequality defining a face F of K of one dimension lower than K itself. The dimension of F equals $\dim(K) - 1$. This implies that F is a facet of K .

We denote the polyhedron representing all inequalities valid for a polyhedron P as $\Pi^0(P)$, and we will refer to it as the polar cone of P .

Definition 1.14 *The polar cone of a polyhedron P is the set denoted as $\Pi^0(P)$, defined as:*

$$\Pi^0(P) := \left\{ (\pi_1, \dots, \pi_n, \pi_0) \in \mathbb{R}^{n+1} \mid \begin{array}{l} \sum_{i=1}^n x_i \pi_i - \pi_0 \leq 0 \\ \forall (x_1, \dots, x_n) \in P \end{array} \right\}.$$

Now, consider the polar of the polar cone of a polyhedron $P \subseteq \mathbb{R}^n$:

$$\left\{ (x_1, \dots, x_n, x_0) \in \mathbb{R}^{n+1} \mid \begin{array}{l} \sum_{i=1}^n \pi_i x_i - \pi_0 x_0 \leq 0 \\ \forall (\pi_1, \dots, \pi_n, \pi_0) \in \Pi^0(P) \end{array} \right\}.$$

The relation with P is straightforward. The polyhedron P is the intersection of this cone with the subspace defined by $x_0 = 1$, projected on the subspace defined by $x_0 = 0$. We denote the polar of the polar cone of a polyhedron P as $\Pi(P)$, and we will refer to it as the induced cone of P .

Definition 1.15 *The induced cone $\Pi(P)$ of a polyhedron P is the polar of the polar cone of P .*

There is a one-to-one relationship between the facets of P and the facets of the induced cone of P . Thus, by polarity, an element on a face of the polar

cone $\Pi^0(P)$ of P whose dimension is one dimension higher than the extreme face of $\Pi^0(P)$, corresponds to an inequality, valid for P and defining a facet of P . This concept is used and extended in Chapter 3.

The concept of an induced cone is also of importance in Chapter 4. Intersecting the induced cone at different values of x_0 , and then projecting them on the subspace defined by $x_0 = 0$, yields a series of polyhedra $P(x_0)$ in $\mathbb{R}^n$. If these polyhedra contain the origin, than this series of polyhedra is nested. That gives us a tool to analyze the quality of a relaxation of a polyhedron, by enclosing the relaxation between $P(1)$ and $P(\alpha)$. The closer α can be to 1, the tighter the relaxation of P .

1.3 Outline

In the next chapter we focus on the relation between the dimension of a polyhedron and the dimension of its projection. We derive sufficient conditions for a one-to-one relation between facets of a polytope and the facets of its projection. Chapter 3 deals with sequential lifting procedures by establishing the relations between the description of a polyhedron and the description of a face of that polyhedron. In Chapter 4 a measure for the strength of classes of inequalities with respect to a linear relaxation is suggested: the best ratio between the optimal value of the linear relaxation with and without a class of inequalities. In the same chapter, we also discuss the relative value of such a concept.

Chapter 5 is on the cycle polytope of a directed graph. Lifting procedures and projection enables us to construct facet defining inequalities for the cycle polytope of a directed graph from each of the facet defining inequalities of the well-known asymmetric traveling salesman polytope of the same graph. It is also demonstrated that not all classes of facet defining inequalities of the cycle polytope of a directed graph can be derived this way, even for small graphs. Lifting procedures are added to guarantee a complete description for graphs on 4 and 5 nodes.

Then, in Chapter 6 formulations for the cell formation problem are compared, and a new integer programming model is proposed. In the chapter thereafter, the polyhedral properties of this formulation are investigated. In the final chapter a specific property of the cell formation polytope is studied: the possibility to patch inequalities defining facets together to a new inequality defining a different facet.

Chapter 2

On the Dimension of Projected Polyhedra

This chapter is joint work with Egon Balas.

An article based on this chapter has been submitted for publication.

In this chapter, we address several basic questions that arise in the use of projection in combinatorial optimization. Central to these is the connection between the dimension of a polyhedron Q and the dimension of its projection on a subspace. We give the exact relationship between the two dimensions. As a byproduct we characterize the relationship between the equality subsystem of a polyhedron and that of its projection. We also derive a necessary and sufficient condition for a face (in particular, a facet) of a polyhedron Q to project into a face (a facet) of the projection of Q , and give a necessary and sufficient condition for the existence of a one to one correspondence between the faces of Q and those of its projection. More generally, we characterize the dimensional relationship between the projection of Q and that of an arbitrary proper face of Q .

2.1 Introduction

Projection has emerged during the last 15 years as one of the most important items in the toolkit of polyhedral combinatorics. In using this tool, one often faces questions concerning the dimension of projected polyhedra. What is the connection between the dimension of the original polyhedron and that of its projections? What about the projections of facets and other faces? It is known that the projection of a polyhedron may have fewer or more facets than the polyhedron itself. But when do facets of the polyhedron project into facets of the projection?

It is this type of questions that have motivated this paper. Our main result in Section 2.2 gives the exact relationship between the dimension of a general polyhedron and that of its projection on a specified subspace. From this, we establish (Section 2.3) the fact that the affine hull of the projection of a polyhedron Q is the projection of the affine hull of Q , and we derive a necessary and sufficient condition for a facet of Q to project into a facet of the projection of Q . More generally, given an arbitrary proper face F of a polyhedron Q , we give a relationship between the dimension of the projection of Q on a subspace, and the dimension of the projection of F on the same subspace. Finally, we show that the operations of projection and monotonicization of polyhedra are commutative.

Consider the polyhedron

$$Q := \{(u, x) \in \mathbb{R}^p \times \mathbb{R}^q : Au + Bx \leq b\},$$

where A , B and b have m rows. The projection of Q onto the subspace

defined by $u = 0$, called the x -space, is defined as

$$Proj_x(Q) := \{x \in \mathbb{R}^q : \exists u \in \mathbb{R}^p \text{ with } (u, x) \in Q\}.$$

It is well known that

$$Proj_x(Q) = \{x \in \mathbb{R}^q : (vB)x \leq vb, \forall v \in W\},$$

where W , called the projection cone, is

$$W := \{v \in \mathbb{R}^m : vA = 0, v \geq 0\}.$$

From the definition, if $Q = \emptyset$, then $Proj_x(Q) = \emptyset$ (and vice versa). So for the rest of this chapter we assume $Q \neq \emptyset$. For background material on projection see [Schrijver, 1986, Nemhauser and Wolsey, 1988, Pulleyblank, 1989, Balas and Fischetti, 1993].

2.2 The Main Result

Let us partition the rows of (A, B, b) into $(A^=, B^=, b^=)$ and $(A^{\leq}, B^{\leq}, b^{\leq})$, where $A^=u + B^=x = b^=$ is the *equality subsystem* of Q , i.e. the set of inequalities satisfied at equality by every $(u, x) \in Q$. Without loss of generality, we may assume that the equality system is of full row rank (otherwise the redundant rows can be removed), and let $r := \text{rank}(A^=, B^=) = \text{rank}(A^=, B^=, b^=)$, where the last equality follows from $Q \neq \emptyset$.

It is well known that $\dim(Q) = p + q - r$, and that Q is full-dimensional, i.e. $\dim(Q) = p + q$, if and only if the equality subsystem is vacuous. We first address this situation.

Proposition 2.2.1 *If $\dim(Q) = p + q$, then $\dim(Proj_x(Q)) = q$.*

Proof: The dimension of $Proj_x(Q)$ is $q - \bar{r}$, where $\bar{r}$ is the rank of the equality subsystem of $Proj(Q)$. We claim that $\bar{r} = 0$. To prove this, we point out that none of the inequalities defining Q is satisfied at equality by every $(u, x) \in Q$. Now suppose there exists $\bar{v} \in W \setminus \{0\}$ such that the inequality $(\bar{v}B)x \leq \bar{v}b$ of the system defining $Proj_x(Q)$ is satisfied at equality by every $x \in Proj_x(Q)$. Then

$$\bar{v}(Au + Bx) = (\bar{v}A)u + (\bar{v}B)x = \bar{v}b$$

for every $(u, x) \in Q$, since $\bar{v}A = 0$. But this implies that every row i of $Au + Bx \leq b$ such that $\bar{v}_i > 0$ is satisfied at equality by every $(u, x) \in Q$, a contradiction. This proves that $\bar{r} = 0$. $\square$

Now, we turn to the general situation, when Q is not necessarily full dimensional. We recall that the matrix $(A^=, B^=)$ is of rank r , and define $r^* := \text{rank}(A^=)$. Let A_0 be a nonsingular $r^* \times r^*$ submatrix of $A^=$. Then, without loss of generality, Q can be written as

$$\left\{ (u_0, u_1, x) : \begin{array}{l} u_0 \in \mathbb{R}^{r^*} \\ u_1 \in \mathbb{R}^{p-r^*} \\ x \in \mathbb{R}^q \end{array} \left| \begin{pmatrix} A_0 & A_1 \\ A_2 & A_3 \\ A_0^\leq & A_1^\leq \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} + \begin{pmatrix} B_0 \\ B_1 \\ B^\leq \end{pmatrix} x \leq \begin{pmatrix} b_0 \\ b_1 \\ b^\leq \end{pmatrix} \right. \right\},$$

where

$$A^= = \begin{pmatrix} A_0 & A_1 \\ A_2 & A_3 \end{pmatrix}, \quad A^\leq = (A_0^\leq, A_1^\leq), \quad B^= = \begin{pmatrix} B_0 \\ B_1 \end{pmatrix}, \quad b^= = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \quad u = \begin{pmatrix} u_0 \\ u_1 \end{pmatrix},$$

and u_0 has r^* components corresponding to the columns of A_0 .

Let us examine the values that r^* can take. Clearly, r^* is nonnegative and cannot exceed the number p of columns of $A^=$. Next, notice that $\text{rank}(B_1) + r^* = r$, and $\text{rank}(B_1) \leq q$. Thus we have

$$\max\{0, r - q\} \leq r^* \leq \min\{p, r\}.$$

If $r^* = p$, then $A_1, A_3, A_1^\leq$ and u_1 are vacuous. If $r^* = r$, then A_2, A_3, B_1 and b_1 are vacuous. If $r^* = 0$, then A_0, A_1, B_0 and b_0 , as well as $A_2, A_0^\leq$ and u_0 are vacuous. In all other cases, all matrices in the above formulation are nonvacuous. Further, if $r^* = r - q$, then B_1 is square, nonsingular.

Next, we project Q onto the subspace of (u_1, x) , by using the equation

$$u_0 = A_0^{-1}b_0 - A_0^{-1}A_1u_1 - A_0^{-1}B_0x$$

to eliminate u_0 . The entries of the matrix $(0, A_3 - A_2A_0^{-1}A_1)$ obtained from (A_2, A_3) through this transformation are all 0. To see this, notice that since $\text{rank}(A^=) = r^* = \text{rank}(A_0)$, the rows of (A_2, A_3) are linear combinations of those of (A_0, A_1) . So, there is a $((r - r^*) \times r^*)$ -matrix R such that $R(A_0, A_1) = (A_2, A_3)$. Since $RA_0 = A_2$ and A_0 is nonsingular, we have $R = A_2A_0^{-1}$, implying $RA_1 = A_2A_0^{-1}A_1 = A_3$. Consequently

$$\text{Proj}_{(u_1, x)}(Q) = \left\{ (u_1, x) \in \mathbb{R}^{p-r^*} \times \mathbb{R}^q \left| \begin{array}{l} \tilde{B}_1x = \tilde{b}_1 \\ \tilde{A}u_1 + \tilde{B}x \leq \tilde{b} \end{array} \right. \right\},$$

where $\tilde{B}_1 := B_1 - A_2 A_0^{-1} B_0$, $\tilde{b}_1 := b_1 - A_2 A_0^{-1} b_0$, $\tilde{A} := A_1^< - A_0^< A_0^{-1} A_1$, $\tilde{B} := B^< - A_0^< A_0^{-1} B_0$, and $\tilde{b} := b^< - A_0^< A_0^{-1} b_0$. We will denote $Q' := \text{Proj}_{(u_1, x)}(Q)$.

Lemma 2.2.1 $\dim(Q') = \dim(Q)$.

Proof: The dimension of a polyhedron can not be less than the dimension of a projection of that polyhedron. So, to prove the lemma we take a set of $z := \dim(Q) + 1$ affinely independent points $(u_0^k, u_1^k, x^k) \in Q$, $k = 1, \dots, z$, and we show that the corresponding set of points $(u_1^k, x^k) \in Q'$ is also affinely independent. Let these latter points be $y^k := (u_1^k, x^k)$, $k = 1, \dots, z$. Showing that they are affinely independent amounts to showing that the points $\begin{pmatrix} y^k \\ -1 \end{pmatrix}$, $k = 1, \dots, z$ are linearly independent. Now suppose they are not; then there exist values λ_k , $k = 1, \dots, z$, not all zero, such that $\sum_{k=1}^z \lambda_k \begin{pmatrix} y^k \\ -1 \end{pmatrix} = 0$. We claim that this implies $\sum_{k=1}^z \lambda_k u_0^k = 0$, which in turn means that the z points $(u_0^k, u_1^k, x^k) \in Q$ are also affinely dependent, a contradiction.

To prove the claim, we notice that

$$\begin{aligned} \sum_{k=1}^z \lambda_k u_0^k &= \sum_{k=1}^z \lambda_k (A_0^{-1} b_0 - A_0^{-1} A_1 u_1^k - A_0^{-1} B_0 x^k) \\ &= \sum_{k=1}^z \lambda_k A_0^{-1} (b_0 - (A_1, B_0) y^k) \\ &= -A_0^{-1} (A_1, B_0, b_0) \sum_{k=1}^z \lambda_k \begin{pmatrix} y^k \\ -1 \end{pmatrix} = 0. \end{aligned}$$

Hence the $p + q - r + 1$ points $y^k \in Q'$, $k = 1, \dots, z$, are affinely independent. $\square$

Since Q' is the projection of Q on the subspace of (u_1, x) , it follows from the definitions that

$$\text{Proj}_x(Q) = \text{Proj}_x(Q').$$

Now, the dimension of Q' is the same as that of Q , but the equality subsystem $\tilde{B}_1 x = \tilde{b}_1$ of Q' has rank $r - r^*$. Notice that Q' can be written as

$$Q' = Q'' \cap \{(u_1, x) \in \mathbb{R}^{p-r^*} \times \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\},$$

where

$$Q'' = \{(u_1, x) \in \mathbb{R}^{p-r^*} \times \mathbb{R}^q : \tilde{A} u_1 + \tilde{B} x \leq \tilde{b}\}$$

is a full dimensional polyhedron, namely one of dimension $p + q - r^*$.

Next, we want to project Q' onto the x -space.

Lemma 2.2.2 $Proj_x(Q') = Proj_x(Q'') \cap \{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}$.

Proof: By definition, $x \in Proj_x(Q')$ if and only if there exists some $u_1 \in \mathbb{R}^{p-r^*}$ such that $(u_1, x) \in Q'$; i.e. such that (i) $(u_1, x) \in Q''$ and (ii) $\tilde{B}_1 x = \tilde{b}_1$. But (i) and (ii) imply that $x \in Proj_x(Q'') \cap \{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}$. $\square$

We are now ready to prove our central result.

Theorem 2.2.1

$$\dim(Proj_x(Q)) = \dim(Q) - p + r^*.$$

Proof: Since the projections of Q and Q' are the same, from Lemma 2.2.2 we have

$$Proj_x(Q) = Proj_x(Q'') \cap \{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}.$$

Further, from this inequality,

$$\dim(Proj_x(Q)) \leq \min\{\dim(Proj_x(Q'')), \dim\{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}\}.$$

The dimension of $\{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}$ is $q - r + r^* = \dim(Q) - p + r^*$. This proves that

$$\dim(Proj_x(Q)) \leq \dim(Q) - p + r^*.$$

To complete the proof, we need to show that the inequality also holds in the other direction. This is equivalent to showing that

$$\dim(Q') \leq \dim(Proj_x(Q')) + p - r^*,$$

since the projection of Q' is the same as that of Q , and from Lemma 2.2.1, $\dim(Q') = \dim(Q)$.

Consider a matrix (D, E) whose rows are vectors $(u_1, x) \in Q'$ that form a maximal independent set. Here D and E correspond to the components u_1 and x , respectively. Then

$$\dim(Q') = \text{rank}(D, E) \leq \text{rank}(D) + \text{rank}(E).$$

Since the rank of D cannot exceed the number of its columns (which is $p - r^*$) and the rank of E cannot exceed $\dim(Proj_x(Q'))$, we have

$$\dim(Q') \leq \dim(Proj_x(Q')) + p - r^*,$$

which is the needed inequality. $\square$

2.3 Consequences

One byproduct of the proof by Theorem 2.2.1 is the result that the solution set of the equality subsystem of $Proj_x(Q)$ is the projection onto the x -space of the solution set of the equality subsystem of Q . Indeed, it is not hard to see that the solution set of the equality subsystem of an arbitrary polyhedron $P \subseteq \mathbb{R}^n$ is the same as the affine hull of P , denoted $aff(P)$, and defined as the set of all $x \in \mathbb{R}^n$ of the form $x = \sum_{i=1}^t x_i \lambda_i$ for some $x_i \in P$ and $\lambda_i \in \mathbb{R}$, $i = 1, \dots, t \in \mathbb{Z}_+$ such that $\sum_{i=1}^t \lambda_i = 1$. (If the equality system is vacuous, i.e. P is full dimensional, then the affine hull of P is the whole space: $aff(P) = \mathbb{R}^n$). With this in mind, the result enunciated above can be stated as

Corollary 2.3.1 $aff(Proj_x(Q)) = Proj_x(aff(Q))$.

Proof: Using the notation introduced in Section 2.2, we have

$$\begin{aligned} aff(Q) &= \{(u, x) \in \mathbb{R}^p \times \mathbb{R}^q : A^-u + B^-x = b^-\} \\ &= \left\{ (u_0, u_1, x) : \begin{array}{l} A_0 u_0 + A_1 u_1 + B_0 x = b_0 \\ \tilde{B}_1 x = \tilde{b}_1 \end{array} \right\} \end{aligned}$$

and

$$aff(Proj_x(Q)) = \{x \in \mathbb{R}^q : \tilde{B}_1 x = \tilde{b}_1\}.$$

The projection cone needed to project $aff(Q)$ onto the x -space has rays (t, w) , where t and w have the same number of components as b_0 and $\tilde{b}_1$, respectively, satisfying $t(A_0, A_1) = 0$, and w unconstrained.

Since A_0 is of full row rank, all the rays of the projection cone are of the form $(0, w)$, which means that $Proj_x(aff(Q))$ is defined by all possible linear combinations of the equation of $\tilde{B}_1 x = \tilde{b}_1$, i.e. is the same as $aff(Proj_x(Q))$.

□

At the beginning of Section 2.2 we discussed the range of $r^* : \max\{0, r - q\} \leq r^* \leq \min\{p, r\}$. In combination with Theorem 2.2.1, this yields the range of $\dim(Proj_x(Q))$:

$$\max\{0, \dim(Q) - p\} \leq \dim(Proj_x(Q)) \leq \min\{q, \dim(Q)\}.$$

We will not discuss all the extreme cases again, except for the one which leads to the next result, which states that the projection operation is dimension preserving if and only if A^- is of full column rank.

Corollary 2.3.2 $\dim(\text{Proj}_x(Q)) = \dim(Q)$ if and only if $r^* = p$. □

Proof: Follows directly from Theorem 2.2.1. □

Next we turn our attention to the dimension of projected faces of Q . Since a face of Q is itself a polyhedron, we should be able to deduce the dimension of its projection from Theorem 2.2.1.

Let $\alpha u + \beta x \leq \beta_0$ be a valid inequality for Q , and suppose

$$F := \{(u, x) \in Q : \alpha u + \beta x = \beta_0\}$$

is a nonempty facet of Q . Let $(\alpha)_A^\top u + (\beta)_B^\top x = (\beta_0)_b^\top$ be the equality subsystem of F , and let $r_F := \text{rank} \left((\alpha)_A^\top, (\beta)_B^\top \right)$. Notice that $r_F - r = 1$, since $\dim(F) = \dim(Q) - 1$. Further, denote $r_F^* := \text{rank} \left((\alpha)_A^\top \right)$. Thus r_F^* is properly defined; the choice of α may not be not unique, but regardless of the choice of α , provided that F is a facet, r_F^* is unambiguous. Then $r_F^* - r^*$ can be interpreted as the difference between the number of dimensions “lost” in projecting Q (which is $p - r^*$) and in projecting F (which is $p - r_F^*$).

Corollary 2.3.3 $\dim(\text{Proj}_x(F)) = \dim(\text{Proj}_x(Q)) - 1 + (r_F^* - r^*)$.

Proof: From Theorem 2.2.1,

$$\begin{aligned} \dim(\text{Proj}_x(F)) &= \dim(F) - p + r_F^* \\ &= \dim(Q) - 1 - p + r_F^* \end{aligned}$$

and

$$\dim(\text{Proj}_x(Q)) = \dim(Q) - p + r^*.$$

Hence

$$\dim(\text{Proj}_x(F)) = \dim(\text{Proj}_x(Q)) - 1 + (r_F^* - r^*).$$

□

Corollary 2.3.3 says that the dimension of $\text{Proj}_x(F)$ is at least equal to the dimension of a facet of $\text{Proj}_x(Q)$; this seems to, but does not in fact, imply that $\text{Proj}_x(F)$ is a facet or an improper face of $\text{Proj}_x(Q)$. The reason for this is that the projection of a face of Q is not always a face of the projection of Q . As an example, think of a pyramid projected onto its base: the vertex of the pyramid, which is one of its faces, does not become a face of the projection, when projected onto the base of the pyramid.

Proposition 2.3.1 *The projection $Proj_x(F)$ of a face F of Q is a face of $Proj_x(Q)$ if and only if Q has a face F_β defined by an inequality of the form $\beta x \leq \beta_0$ (with $\alpha = 0$), such that $Proj_x(F_\beta) = Proj_x(F)$. When this is the case, $F \subseteq F_\beta$.*

Proof: Necessity. Suppose $Proj_x(F)$ is a face of $Proj_x(Q)$. Then there exists an inequality $\beta x \leq \beta_0$ such that $Proj_x(F) = \{x \in Proj_x(Q) : \beta x = \beta_0\}$. But $\beta x \leq \beta_0$ also defines a face F_β of Q , namely $F_\beta = \{(u, x) \in Q : \beta x = \beta_0\}$. Clearly, $Proj_x(F_\beta) = Proj_x(F)$, since the expressions on both sides of the equality are equal to $Proj_x(Q) \cap \{x : \beta x = \beta_0\}$.

Sufficiency. Suppose F_β is the face of Q defined by $\beta x \leq \beta_0$, such that $Proj_x(F_\beta) = Proj_x(F)$. Since $\beta x \leq \beta_0$ is valid for $Proj_x(Q)$ and $Proj_x(F_\beta) = Proj_x(Q) \cap \{\beta x = \beta_0\}$, clearly $Proj_x(F_\beta)$ is a face of $Proj_x(Q)$; hence so is $Proj_x(F)$.

Furthermore, we claim that $F \subseteq F_\beta$. For if not, let $(u, x) \in F \setminus F_\beta$; then $(u, x) \in F$ implies $x \in Proj_x(F)$ and hence $\beta x = \beta_0$, whereas $(u, x) \notin F_\beta$ implies $\beta x < \beta_0$, a contradiction. $\square$

As an immediate consequence of Proposition 2.3.1 we have

Corollary 2.3.4 *There is a one to one correspondence between the faces of Q and the faces of $Proj_x(Q)$ if and only if every face of Q can be defined by an inequality of the form $\beta x \leq \beta_0$ (with $\alpha = 0$).*

Now consider again the important case when F is a facet of Q . Then we have

Corollary 2.3.5 *Let F be a facet of Q . Then $Proj_x(F)$ is a facet of $Proj_x(Q)$ if and only if $r_F^* = r^*$.*

Proof:

From Corollary 2.3.3 $Proj_x(F)$ has the right dimension for being a facet of $Proj_x(Q)$ if and only if $r_F^* = r^*$. It then suffices to show that $Proj_x(F)$ is a face of $Proj_x(Q)$ if $r_F^* = r^*$.

Since F is a facet of Q , it has a defining inequality, say $\alpha u + \beta x \leq \beta_0$. If $r_F^* = r^*$, then there exists a vector λ such that $\alpha = \lambda A^=$; hence the above inequality is equivalent to $(\beta - \lambda B^=)x \leq \beta_0 - \lambda B^=$, which can be written as $\beta'x \leq \beta'_0$. It then follows from Proposition 2.3.1 that $Proj_x(F)$ is a face of $Proj_x(Q)$. $\square$

An immediate consequence of Corollary 2.3.5 is

Corollary 2.3.6 *If F is a facet of Q and $r^* = p$, then $Proj_x(F)$ is a facet of $Proj_x(Q)$.*

Proof: If the matrix $A^=$ is of full column rank, then so is the matrix $\begin{pmatrix} \alpha \\ A \end{pmatrix}^=$, and thus $r^* = r_F^*$. $\square$

An interesting example where this situation arises is that of the cycle polytope of a directed graph. Let Q be the Prize Collecting Traveling Salesman (PCTS) polytope defined on a digraph G with loops, i.e. the convex hull of points (x, y) , where x is the incidence vector of a directed cycle on some node set S , and y is the incidence vector of the set of loops indexed by $N \setminus S$. Then $Proj_x(Q)$ is the cycle polytope of G , i.e. the convex hull of incidence vectors x of directed cycles of G .

Now let F be any facet of the PCTS polytope Q . It is shown in Chapter 5 that $Proj_x(F)$, the projection of F onto the x -space, is a facet of the cycle polytope defined on G . The proof is based on the fact that the matrix $A^=$ in the polyhedral description of Q consists of the unit columns associated with the components of y , hence is of full column rank and so Corollary 2.3.6 applies.

On the other hand, if F is a facet of Q but $Proj_x(F)$ is not a facet of $Proj_x(Q)$, i.e. $r_F^* \neq r^*$, then the only possibility is that $k^* := r_F^* - r^* = 1$. Since this implies that $Proj_x(F)$ is of the same dimension as $Proj_x(Q)$, it follows that either $Proj_x(F) = Proj_x(Q)$, or else $Proj_x(F)$ is a subset, but not a face, of Q .

Corollary 2.3.6 can be generalized to faces other than facets. Indeed, we have the following.

Proposition 2.3.2 *If F is a face of Q of dimension d and $r^* = p$, then $Proj_x(F)$ is a face of $Proj_x(Q)$ of dimension d .*

Chapter 3

Sequential Lifting Procedures

An article based on this chapter has been submitted for publication.

In integer programming we often find some benefit in having disposal of strong valid inequalities for the convex hull of the set of all feasible solutions. To determine such inequalities, so-called lifting procedures can be applied. The main idea of these procedures is to use the description of a face P' of P , defined by fixing some variables, to derive a partial description of the original polytope P . In this chapter sequential lifting is extended in such a way that lifting an inequality defining a facet of P' guarantees to yield an inequality defining a facet of the original polytope P . Furthermore, it is shown that, by generalizing this procedure in an appropriate way, all inequalities defining nontrivial facets of P can be obtained by lifting. For convenience, in this paper we restrict ourselves to zero-one programming, but the results can be extrapolated.

3.1 Introduction and Terminology

Consider the zero-one program $\max\{c^T x \mid x \in S \subseteq \mathbb{B}^n\}$ (where $\mathbb{B}^n$ denotes the set of n -dimensional binary vectors). Define P to be the convex hull of S , where by convention the convex hull of the empty set is the empty set. Unlike integer programming in general, each feasible solution $x \in S$ is an extreme point of P .

Theoretically, given the set S , a facial description of P can be determined by examining the polar cone $\Pi^0(P)$ of P , representing the coefficients $(a, a_0) \in \mathbb{R}^{n+1}$ for which the inequality $a^T x \leq a_0$ is valid for P (see Section 1.2 for a brief explanation). The polar cone of P is described as follows:

$$\Pi^0(P) = \left\{ (a, a_0) \in \mathbb{R}^{n+1} \mid a^T x - a_0 \leq 0 \quad \forall x \in S \right\} \quad (3.1)$$

If P is of full dimension and bounded, and $(a^T, a_0) \neq \mathbf{0}$, then $a^T x \leq a_0$ defines a facet of P if and only if (a^T, a_0) is on an extreme ray of $\Pi^0(P)$. This is a well-known result, see for example [Rockafellar, 1970], [Schrijver, 1986], and [Nemhauser and Wolsey, 1988]. However, determining the extreme rays of $\Pi^0(P)$ is computationally complex. At the moment, programs designed for computing a minimal facial description of the convex hull of a set of points (see e.g. [Fukuda, 1994]) can only handle relatively small sets S of feasible points.

Sequential lifting procedures are based on the idea to organize the quest for equalities of P and facet-defining inequalities of P by making use of the (partial) description of the faces of P induced by the trivial inequalities: $x_j \geq 0$, and $x_j \leq 1$ for any $j \in \{1, \dots, n\}$. To achieve this, we assume

first that $j = n$ and we partition the set S into two subsets, say S^0 and S^1 , as follows: $S^0 = \{x \in S : x_n = 0\}$, and $S^1 = \{x \in S : x_n = 1\}$. P^0 and P^1 are defined as the convex hull of S^0 and S^1 respectively. This implies that P^0 and P^1 are (possibly improper) faces of P . Let p^0 , p^1 , and p denote the dimensions of P^0 , P^1 , and P respectively. This implies $-1 \leq p^0, p^1 \leq p \leq p^0 + p^1 + 1$. The difference of the dimensions of a polyhedron P and a face F of P is important for this chapter.

Definition 3.1 *The dimension gap of a polyhedron P and a face F of P is the difference between the dimensions of P and F .*

Obviously, the dimension gap between P and P^0 can range from 0 to $p^1 + 1$.

Definition 3.2 *An inequality of the form:*

$$a^T x + \beta x_n \leq a_0 \quad \beta \in \mathbb{R} \quad (3.2)$$

is called a lifting from P^0 to P of the inequality $a^T x \leq a_0$ if the inequality (3.2) is valid for P .

In case S^1 is empty, the validity of the inequality (3.2) is independent of the value of β . Otherwise, it is easy to see that (3.2) is a valid inequality for P if and only if the coefficient β satisfies:

$$\beta \leq a_0 - \max\{a^T x \mid x \in S^1\} \quad (3.3)$$

The closer β is to its upper bound, the stronger the lifting. However, optimizing the maximization problem in the upper bound can be very hard. Therefore, in some cases we have to settle for approximations (see e.g. [Gu et al, 1995]).

Assume that the inequality $a^T x \leq a_0$ defines a face of P^0 of dimension p_a^0 , so $p_a^0 \in \{-1, \dots, p^0\}$. Then, inequality (3.2) defines a face of P of dimension at least p_a^0 . When (3.3) is satisfied with equality, the dimension of the face of P defined by (3.2) is at least $p_a^0 + 1$ (the union of a maximal affinely independent set of solutions in S^0 satisfying (3.2) with equality, with at least one solution in S^1 satisfying (3.2) with equality is still affinely independent). Sometimes, when (3.2) defines a facet of P^0 , it also defines a facet of P , as is formulated more precisely in the next proposition.

Proposition 3.1.1 [Wolsey,1975] *Assume $a^T x \leq a_0$ defines a facet of P^0 and the dimension gap between P and P^0 is 1. Then, $a^T x + \beta x_n \leq a_0$ defines a facet of P if and only if $\beta = a_0 - \max\{a^T x \mid x \in S^1\}$.*

All valid inequalities of P^1 can be lifted in a similar way. Let P' be a nonempty intersection of P with a set of hyperplanes defined by the trivial inequalities $0 \leq x_i \leq 1$; in other words: P' is a lower dimensional face of P defined on a subset of the variables. Now, we can derive a strong valid inequality for P by lifting a strong valid inequality for the lower dimensional face P' of P one variable at a time. We can consider different orderings of the variables to be lifted, so from one valid inequality for P' we can derive a family of valid inequalities for P . However, even lifting all facet defining inequalities in all possible orderings of the variables to be lifted, not all inequalities valid for P can be derived (see e.g. [Nemhauser and Wolsey, 1988]).

The sequential lifting procedure can be generalized in several ways. In [Wolsey, 1976] a result is presented, showing how a valid inequality for the convex hull of a set of integer points (not necessarily zero-one) can be derived from a valid inequality in a subset of the variables. In [Zemel, 1978] a parallel, or simultaneous, lifting procedure is proposed. For this procedure, the coefficients of a subset of the variables are determined simultaneously, and if the number of variables equals the dimension gap, an inequality defining a facet of the lower dimensional polytope is lifted to an inequality defining a facet of the higher dimensional polytope. Traditional lifting with a dimension gap of 1 can be interpreted as a special case of this procedure. In general, it is not possible to partition a lifting sequence of variables into reasonable small subsequences of a cardinality that corresponds to the respective dimension gaps. [Balas and Zemel, 1984] showed that with this procedure every non-trivial facet of the polytope corresponding to a multidimensional knapsack problem can be obtained.

However, for all sequential lifting procedures discussed so far, the lifting of an inequality defining a facet of P^0 is not guaranteed to define a facet of P unless the dimension gap of P and P^0 happens to equal 1. So, if an inequality defines a facet F^0 of P^0 , then a procedure is required to transform the corresponding lifting to an inequality defining the same facet F^0 of P^0 but also defining a facet of P . [Boyd and Pulleyblank, 1991] presented an iterative procedure transforming an inequality which is valid for a polyhedron P and facet defining for a non-empty face Q of P , into an inequality that is equivalent to the original inequality with respect to Q , and that is facet defining for P .

Our purpose here is to propose a sequential lifting procedure such that the lifting of an inequality defining a facet of P^0 always defines a facet of P . First, in the next section, we will compare the traditional lifting, simultaneous lifting, and the lifting suggested in this paper, by means of an example.

Then, in Section 3.3 we will investigate the relation between the dimension gap, and the equality systems of P^0 and P . It turns out that facet defining inequalities of P can be generated by combining a specific set of equalities, satisfied by all elements in S^0 , but not by all elements in S . In Section 3.4 we discuss the possibilities to use this specific set of equalities in order to formulate a sequential lifting procedure such that the lifting of an inequality defining a facet of P^0 always defines a facet of P . In Section 3.5 the results of the former two sections are combined to generalize the procedure in such a way that all inequalities defining nontrivial facets of P can be obtained by lifting the facial description of P^0 . Procedures similar to those presented in Sections 3.3, 3.4, and 3.5, can generate facet defining inequalities of P , lifting inequalities of P^1 . In that case, the analysis is analogue. In Section 3.6 we present an approach called strict sequential lifting. Basically, an inequality for P^0 is lifted to P via intermediate polytopes $P^0 = P_0, P_1, \dots, P_{p-p^0} = P$ such that P_{i-1} is a facet of P_i , and therefore the dimension gap between P_i and P_{i-1} is one for all $i \in \{1, \dots, p - p^0\}$. In Section 3.7 we discuss an example on the cell formation polytope, and in the final section we briefly discuss the results presented in this chapter.

3.2 Comparison of the Procedures

In this section, an example is discussed, illustrating the nature and the limitations of the different procedures, as mentioned in the former section.

Consider the set S of feasible solutions consisting of $(0, 0, 0)$, $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$. Then, the polar cone (see Section 1.2) of the convex hull P of S is described as follows:

$$\Pi^0(P) = \left\{ a_x, a_y, a_z, a_0 \in \mathbb{R} \left| \begin{array}{lcl} 0 & \leq & a_0 \\ a_x + a_y & \leq & a_0 \\ a_x + a_z & \leq & a_0 \\ a_y + a_z & \leq & a_0 \end{array} \right. \right\}.$$

This cone has four extreme rays: $(1, 1, 1, 2)$, $(1, -1, -1, 0)$, $(-1, 1, -1, 0)$, and $(-1, -1, 1, 0)$, from which we abstract the following minimal description of P (in general the number of inequalities describing P and $\Pi^0(P)$ are not the same):

$$P = \left\{ x, y, z \in \mathbb{R} \left| \begin{array}{lcl} x + y + z & \leq & 2 \\ x - y - z & \leq & 0 \\ -x + y - z & \leq & 0 \\ -x - y + z & \leq & 0 \end{array} \right. \right\}.$$

A geometrical representation of P is given in Figure 3.1 below.



Figure 3.1: The polytope P , the convex hull of S

Let $S^0 := S \cap \{x, y, z \in \mathbb{B} | z = 0\}$, and $S^{00} := S^0 \cap \{x, y, z \in \mathbb{B} | y = 0\}$. Let P^0 , and P^{00} be the respective convex hulls. It is trivial to derive a minimal description of P^{00} , and to verify that this polytope has only one facet, namely the empty set, which is defined by the inequality $0 \leq 1$ (and many other inequalities). A lifting from P^{00} to P^0 of this inequality, as defined in Definition 3.2, is a valid inequality of the form $\beta y \leq 1$. Given that the inequality is valid for P^0 , the upper bound for β is 1. Since the dimension gap between P^0 and P^{00} is 1, the inequality $y \leq 1$ defines a facet of P^0 . Notice that P^{00} is the face of P^0 defined by the inequality $y \geq 0$, and since the dimension gap is 1, we find that the inequality $y \geq 0$ defines a facet of P^0 too.

A lifting of the inequality $y \leq 1$ from P^0 to P is a valid inequality of the form $y + \gamma z \leq 1$. The maximum value γ can obtain, is $1 - \max\{y | (x, y, z) \in S : z = 1\} = 0$. Since the dimension gap between P and P^0 is 2, the inequality $y \leq 1$ defines a face of P of dimension 1 or 2. Notice that we also find that the inequality $z \geq 0$ defines a face of P of dimension precisely 1.

Could we improve our results by choosing another ordering of lifting the variables? No, since S is symmetric in the variables x , y , and z . We can swap any pair of variables, but the description of S would not change. Actually, we have the following equations:

$$S = \left\{ x, y, z \in \mathbb{B} \left| \begin{array}{ll} x = y & \text{if } z = 0 \\ x = 1 - y & \text{if } z = 1 \end{array} \right. \right\}$$

$$\begin{aligned}
 &= \left\{ x, y, z \in \mathbb{B} \left| \begin{array}{ll} x = z & \text{if } y = 0 \\ x = 1 - z & \text{if } y = 1 \end{array} \right. \right\} \\
 &= \left\{ x, y, z \in \mathbb{B} \left| \begin{array}{ll} y = z & \text{if } x = 0 \\ y = 1 - z & \text{if } x = 1 \end{array} \right. \right\}.
 \end{aligned}$$

In this example, applying lifting of facet defining inequalities, starting from any of the faces defined by trivial inequalities, yields nothing but trivial inequalities, though none of these inequalities defines a facet of P .

Now, let us try to improve these results, by applying **simultaneous lifting** as proposed by Zemel. Two convex hulls whose dimension gap equals the number of variables to be lifted, are P^0 and P^{00} , but for this lifting, the method boils down to ordinary lifting, since the dimension gap is 1. The only other possibility is to lift from a single point to P . Since for this example P is of full dimension, and for the solution $(0, 0, 0)$ the convex hull of that single solution has dimension 0, the dimension gap between these polytopes is n . So, the inequality $0 \leq 1$ can be lifted to a facet defining inequality of P by lifting all variables simultaneously. The simultaneous lifting of the inequality $0 \leq 1$ with respect to all variables, is a valid inequality for P of the form:

$$\beta_x x + \beta_y y + \beta_z z \leq 1 \quad \beta_x, \beta_y, \beta_z \in \mathbb{R}$$

The inequality above is valid for P if and only if the coefficients satisfy the following inequalities, one for each solution of S different from $(0, 0, 0)$:

$$\left\{ \begin{array}{ll} \beta_x + \beta_y \leq 1 & \text{for } (x, y, z) = (1, 1, 0) \\ \beta_x + \beta_z \leq 1 & \text{for } (x, y, z) = (1, 0, 1) \\ \beta_y + \beta_z \leq 1 & \text{for } (x, y, z) = (0, 1, 1) \end{array} \right.$$

Zemel showed that each extreme point of the polyhedron, defined by the three inequalities above, corresponds to a simultaneous lifting with the property that it defines a facet for P . Here, the inequalities induce only one extreme point: $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. The lifting corresponding to this extreme point indeed defines a facet of P :

$$x + y + z \leq 2$$

Now, we will give a rough sketch of how the procedure suggested in this chapter performs on this example. As long as the dimension gap equals 1, the method is identical to lifting based on Definition 3.2. As described

earlier, this is going to yield the inequalities $y \leq 1$ and $y \geq 0$ defining facets of P^0 . For the next lifting step the dimension gap between P^0 and P is 2 and here the methods diverge.

Besides $z = 0$, there are many other equalities that hold for P^0 but not for P . We need a maximal linearly independent set of equalities with this property. All the other equalities with this property are linear combinations of the equalities in the set. Such a set could be $z = 0$ and $x - y = 0$. The cardinality of such a maximal set equals the dimension gap.

The **extended lifting** of the inequality $y \leq 1$ with respect to variable z , is a valid inequality for P of the form:

$$y + \alpha z + \beta(x - y) \leq 1 \quad \alpha, \beta \in \mathbb{R}$$

The inequality above is valid for P if and only if the coefficients α and β satisfy the following inequalities, one for each solution in $S \setminus S^0$:

$$\begin{cases} 1 + \alpha - \beta \leq 1 & \text{for } (x, y, z) = (0, 1, 1) \\ \alpha + \beta \leq 1 & \text{for } (x, y, z) = (1, 0, 1) \end{cases}$$

In Section 3.4 it will be shown that each extreme point of the polyhedron that is defined by the two inequalities above, corresponds to an extended lifting defining a facet of P . In this case, the inequalities induce only one extreme point: $(\frac{1}{2}, \frac{1}{2})$. The extended lifting corresponding to this extreme point can be formulated as:

$$x + y + z \leq 2$$

which indeed is facet defining for P .

In a similar way, the extended lifting of the inequality $y \geq 0$ with respect to variable z , is a valid inequality for P of the form:

$$-y + \alpha z + \beta(x - y) \leq 0 \quad \alpha, \beta \in \mathbb{R}$$

The inequality above is valid for P if and only if the coefficients α and β satisfy the following inequalities, one for each solution in $S \setminus S^0$:

$$\begin{cases} -1 + \alpha - \beta \leq 0 & \text{for } (x, y, z) = (0, 1, 1) \\ \alpha + \beta \leq 0 & \text{for } (x, y, z) = (1, 0, 1) \end{cases}$$

Again, the inequalities induce only one extreme point: $(\frac{1}{2}, -\frac{1}{2})$. The extended lifting corresponding to this extreme point can be formulated as:

$$-x - y + z \leq 0$$

which indeed defines a facet of P .

As in the case of traditional lifting, there is also a byproduct relating to the equalities. Assume we are looking for valid inequalities for P of the following form:

$$\alpha z + \beta(x - y) \leq 0 \quad \alpha, \beta \in \mathbb{R}$$

Such an inequality defines a face of P containing P^0 . The inequality is valid for P if and only if the coefficients α and β satisfy the following inequalities, one for each solution in $S \setminus S^0$:

$$\begin{cases} \alpha - \beta \leq 0 & \text{for } (x, y, z) = (0, 1, 1) \\ \alpha + \beta \leq 0 & \text{for } (x, y, z) = (1, 0, 1) \end{cases}$$

In Section 3.3 it will be shown that each extreme ray of the cone defined by the two inequalities above, corresponds to an extended lifting that is facet defining for P . The inequalities induce only two extreme rays: $(-1, 1)$, and $(-1, -1)$. The extended liftings corresponding to these extreme rays can be formulated as:

$$\begin{cases} -x + y - z \leq 0 \\ x - y - z \leq 0 \end{cases}$$

which define facets of P .

In this example, the procedure of extended lifting does not only yield facet defining inequalities of P , but even gives us a minimal description of the polytope P . In general, to derive a minimal description takes an extra effort, as will be explained in Section 3.5.

3.3 The Dimension Gap and Equality Lifting

In this section we study the relation between the dimension gap of P and P^0 as defined in section 3.1, and the equalities that hold for P^0 but not for P . Our interest in these kind of inequalities is motivated by the following observation:

Observation 3.3.1 *Let Q be a face of P such that $Q \neq P$. Then, there is at least one facet F of P such that $Q \subseteq F$.*

This observation implies that there is at least one inequality $b^T x \leq b_0$ which defines a facet of P and which is satisfied at equality by all $x \in S^0$. We will

see a procedure to abstract the inequalities defining these facets of P , from equalities holding for P^0 .

We first study relations between the dimensions of P and P^0 . Let $L_Q := \{(e^T, e_0) \in \mathbb{R}^{n+1} | e^T x = e_0 \ \forall \ x \in Q\}$. So L_Q represents all equalities satisfied by Q . Then, L_Q is a linear space of dimension $n - q$. Let $\{(e^{iT}, e_0^i), i = 1, \dots, q\}$ form a basis of L_Q and let (E, e) be the $(n - q) \times (n + 1)$ -matrix whose i -th row equals (e^{iT}, e_0^i) . The set of basic vectors will be called an *equality set* of Q , and (E, e) an *equality matrix* of Q .

Let (E, e) be an equality matrix of P^0 , and let $(E^=, e^=)$ be an equality matrix of P . It can occur that for every row (e^{iT}, e_0^i) of (E, e) there is an $x \in S^1$ such that $e^{iT}x \neq e_0^i$, though there might exist linear combinations $\lambda^T(E, e)$ of the rows of (E, e) inducing equalities that are satisfied by all $x \in S$. We need to pay attention to this phenomenon, which is illustrated in the following example.

Example 3.3.1 Consider the instance discussed in the former section. An equality set of P^{00} consists of the rows $(1, 0, -1, 0)$, $(0, 1, 0, 0)$, and $(0, 1, -1, 0)$, corresponding to respectively the equalities $x - z = 0$, $y = 0$, and $y - z = 0$. None of these equalities holds for P^0 . In fact, our choice for the equality set (basic vectors) was an unlucky one. Below, we will show that it is possible to choose the equality set for P^{00} in such a way that the number of basic vectors that do not correspond to equalities for P^0 exactly equals $p^0 - p^{00}$. Notice that for this example $p^0 - p^{00} = 1$.

In general, it is not trivial to determine an equality set of P^0 , not even when an equality set of P^{00} is given. Consider the set of linear combinations of the above equality set of P^{00} that hold for P^0 :

$$\begin{aligned} \Lambda &= \{\lambda \in \mathbb{R}^3 | \lambda_1(x - z) + \lambda_2 y + \lambda_3(z - y) = 0 \quad \forall (x, y, z) \in S^0 \setminus S^{00}\} \\ &= \{\lambda \in \mathbb{R}^3 | \lambda_1 + \lambda_2 - \lambda_3 = 0\} \\ &= \{\lambda \in \mathbb{R}^3 | (\lambda_1, \lambda_2, \lambda_3) = c_1(1, 0, 1) + c_2(0, 1, 1), \quad c_1, c_2 \in \mathbb{R}\}. \end{aligned}$$

The set Λ is a subspace. Any minimal set of elements spanning this subspace, corresponds to an equality set of P^0 . For example, the set of elements $\{(1, 0, 1), (0, 1, 1)\}$ correspond to the equalities $(x - z) + (z - y) = x - y = 0$, and $y + (z - y) = z = 0$, inducing an equality set of P^0 .

Consider the linear spaces L_P and L_{P^0} as defined above. Obviously, $L_P \subset L_{P^0}$. Hence, there exists a basis $\{(e^{iT}, e_0^i), i = 1, \dots, n - p^0\}$ of L_{P^0} such that the vectors of $\{(e^{iT}, e_0^i), i = 1, \dots, n - p\}$ form a basis of L_P . The space defined by the difference set $\{(e^{iT}, e_0^i), i = n - p + 1, \dots, n - p^0\}$ is

called the *dimension gap space* between P and P^0 . A basis (equality set) for this space will be denoted as $\{(d^i)^T, d_0^i), i = 1, \dots, p - p^0\}$. Let (D, d) be the corresponding equality matrix. We will call (D, d) the *equality system* of the dimension gap between P and P^0 .

Henceforth, we make the following assumptions: the equality set of P is a subset of the equality set of P^0 , (D, d) is an equality system of the dimension gap of P and P^0 , and the dimension gap is at least one, that is P^1 is nonempty. Since $p > p^0$, we also assume in the remainder of this chapter that the first row in (D, d) represents $x_n = 0$: $d_1^T x = x_n = 0 = d_0^1$.

Now, let us reconsider Observation 3.3.1. That suggested the existence of at least one inequality $b^T x \leq b_0$ valid for P , and defining a facet F_b of P containing P^0 . The latter property of $b^T x \leq b_0$ implies that there is a unique $(\delta, \lambda) \in \mathbb{R}^{(p-p^0)+(n-p)}$ such that $b^T = \delta^T D + \lambda^T E^=$, and $b_0 = \delta^T d + \lambda^T e^=$. From the validity of $b^T x \leq b_0$ it follows that δ satisfies $\delta^T D x \leq \delta^T d$ for all x in P . So, as a first step to abstracting (b^T, b_0) from (D, d) and $(E^=, e^=)$, we define a *lifting cone* $\Delta^=$ for the equality system (D, d) as follows:

$$\Delta^= := \{\delta \in \mathbb{R}^{p-p^0} \mid \delta^T D x \leq \delta^T d, \quad \forall x \in S^1\}.$$

For this cone we can prove the following proposition:

Proposition 3.3.1 *If S^1 is nonempty, the cone $\Delta^=$ has precisely one extreme point, i.e. the zero vector, and at least one extreme ray.*

Proof: Since $\Delta^=$ is a cone, the zero vector is the only extreme point, if there is any. Suppose the zero vector is not an extreme point of $\Delta^=$. Then, there is a nonzero $\delta \in \Delta^=$ such that also $-\delta \in \Delta^=$. But that implies $\delta^T (D, d)x = \delta^T d$ for all $x \in S^1$, contradicting the definition of (D, d) .

To prove that $\Delta^=$ has an extreme ray, it is sufficient to show that $\Delta^=$ has at least dimension one, since we have already concluded that $\Delta^=$ has a face of dimension zero. So, it suffices to exhibit a nonzero element $\delta \in \Delta^=$: $\delta_1 = -1$, and $\delta_i = 0$ for all $i \in \{2, \dots, p - p^0\}$. $\square$

Definition 3.3 *An inequality of the form*

$$\delta^T D x \leq \delta^T d \quad \delta \in \mathbb{R}^{p-p^0} \quad (3.4)$$

is an extended lifting from P^0 to P of the corresponding equality system (D, d) , if it is a valid inequality for P .

Clearly, inequality 3.4 is valid for P if and only if the vector δ is in $\Delta^=$. The next proposition generalizes this observation.

Proposition 3.3.2 *The inequality $b^T x \leq b_0$ is valid for P and defines a face F_b of P containing P^0 if and only if there exists a (unique) (δ, λ) , with $\delta \in \Delta^\circ \setminus \{0\}$ and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \delta^T D + \lambda^T E^\circ$, and $b_0 = \delta^T d + \lambda^T e^\circ$.*

Proof:

$\Rightarrow$) Since $P^0 \subseteq F_b$, and the rows of (D, d) and (E°, e°) form an equality set of P^0 , there exists a unique (δ, λ) , with $\delta \in \mathbb{R}^{p-p^0}$ and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \delta^T D + \lambda^T E^\circ$, and $b_0 = \delta^T d + \lambda^T e^\circ$. Suppose $\delta \notin \Delta^\circ$. Then, there is at least one $x^* \in S^1$ such that $\delta^T D x^* > \delta^T d$. But that implies $\delta^T D x^* + \lambda^T E^\circ x^* > \delta^T d + \lambda^T e^\circ$, which is equivalent to $b^T x^* > b_0$. This contradicts the assumption that $b^T x \leq b_0$ is valid for P .

$\Leftarrow$) Given is: $\delta^T D x \leq \delta^T d$ for all $x \in S^1$, $\delta^T D x = \delta^T d$ for all $x \in S^0$, and $\lambda^T E^\circ x = \lambda^T e^\circ$ for all $x \in S$. Combining these three expressions, it follows that $\delta^T D x + \lambda^T E^\circ x \leq \delta^T d + \lambda^T e^\circ$, for all $x \in S$, implying the validity of $b^T x \leq b_0$ for P . Combining the latter two of the given expressions, we get $\delta^T D x + \lambda^T E^\circ x = \delta^T d + \lambda^T e^\circ$ for all $x \in S^0$, implying $b^T x = b_0$ for all $x \in P^0$, therefore $P^0 \subseteq F_b$. $\square$

So, with Δ° we characterize all inequalities valid for P and defining a face of P containing P^0 . But which of these inequalities are facet defining? In the simple case where $p - p^0 = 1$, the lifting set boils down to $\Delta^\circ = \{\delta \in \mathbb{R} \mid \delta \leq 0\}$. The only ray of this cone corresponds to the inequality $x_n \geq 0$, which defines a facet of P , since $p^0 = p - 1$. This result can be generalized as follows.

Proposition 3.3.3 *The inequality $b^T x \leq b_0$ is valid for P and defines a facet F_b of P containing P^0 if and only if there exists a (unique) (δ, λ) , with δ on an extreme ray of $\Delta^\circ \setminus \{0\}$ and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \delta^T D + \lambda^T E^\circ$, and $b_0 = \delta^T d + \lambda^T e^\circ$.*

Proof:

$\Rightarrow$) Using the results of Proposition 3.3.2, we only have to show that δ is on an extreme ray of Δ° . Assume there is a nonempty index set I , such that $\delta = \sum_{i \in I} \delta_i$, where $\delta_i \neq 0$ for all $i \in I$. We show that for every $i \in I$ there is a scalar α such that $\delta_i = \alpha \delta$.

Let $b_i^T = \delta_i^T D$, and $b_i^0 = \delta_i^T d$ for all $i \in I$. Then, from Proposition 3.3.2 it follows that $b_i^T x \leq b_i^0$ is valid for P and defines a face F_i of P containing P^0 , for all $i \in I$. Since $b^T x \leq b_0$ is implied by the set of valid inequalities $b_i^T x \leq b_i^0$ for $i \in I$, F_b is a subset of each of the faces F_i of P . Because F_b is a facet of P and $\delta_i \neq 0$, F_i is a facet of P too, and $F_b = F_i$ for each $i \in I$.

Choose any $i \in I$. The row (b^T, b_0) together with the rows of $(E^=, e^=)$, form an equality set of F_b . So, there is a unique $(\alpha, \mu) \in \mathbb{R}^{1+(n-p)}$, such that $(b_i^T, b_{i0}) = \alpha(b^T, b_0) + \mu(E^=, e^=)$. Plugging in the expressions for (b_i^T, b_{i0}) and (b^T, b_0) in (D, d) and $(E^=, e^=)$ we get:

$$\delta_i^T(D, d) = \alpha\delta^T(D, d) + (\alpha\lambda + \mu)^T(E^=, e^=).$$

By the linear independence of the rows of (D, d) and $(E^=, e^=)$, we have $(\delta_i - \alpha\delta) = 0$.

$\Leftarrow$) Using the results of Proposition 3.3.2, we only have to show that F_b is a facet P . Assume there is a nonempty index set I , such that for each $i \in I$ there is a facet F_i of P , and such that $\bigcap_{i \in I} F_i = F_b$. We will show that for each $i \in I$, $F_i = F_b$.

Assume that for each $i \in I$ the inequality $c_i^T x \leq c_{i0}$ defines the facet F_i of P . So, $(b^T, b_0) = \sum_{i \in I} w_i(c_i^T, c_{i0})$ for some nonnegative weights w_i . Also, by Proposition 3.3.2, for each $i \in I$ there is a unique (γ_i, μ_i) such that $\gamma_i \in \Delta^=$, $\mu_i \in \mathbb{R}^{n-p}$, $c_i^T = \gamma_i^T D + \mu_i^T E^=$, and $c_{i0} = \gamma_i^T d + \mu_i^T e^=$. Combining this, we get:

$$(b^T, b_0) = \delta^T(D, d) + \lambda^T(E^=, e^=) = \sum_{i \in I} w_i(\gamma_i^T(D, d) + \mu_i^T(E^=, e^=)).$$

Because the rows of (D, d) and $(E^=, e^=)$ are linearly independent, we conclude: $\delta = \sum_{i \in I} w_i \gamma_i$. Since δ is an extreme ray, all the γ_i 's are on the same ray. So, for every $i \in I$, there is a positive scalar α such that $\alpha\delta = \gamma_i$. Then, every $x \in F_i$ satisfies: $\alpha\delta^T D x = \alpha\delta^T d$. This implies $F_b = F_i$. $\square$

This result implies that we can abstract facet defining inequalities from the equality system of a dimension gap, by generating extreme rays of $\Delta^=$. To get better acquainted with this approach, let us consider an example.

Example 3.3.2 Let $R := \{x \in \mathbb{R}^4 \mid 3x_1 + 2x_2 + 2x_3 + 2x_4 \leq 6\}$, and define the set S of feasible solutions as follows: $S := \mathbb{B}^4 \cap R$. Given a linear objective function, we can optimize over S by using LP techniques to optimize the same function over the convex hull P of S . However, we don't have a closed form description of P yet. Approximating P with $[0, 1]^4 \cap R$ could result in a fractional solution, like $y = (\frac{2}{3}, 0, 1, 1)$.

Apparently, the approximation could be improved, even if we restrict ourselves to the intersections of P with both the subspaces defined by $x_3 = 1$ and $x_4 = 1$ respectively. Let $P^1 := \{x \in P \mid x_4 = 1\}$, and $P^{11} := \{x \in P^1 \mid x_3 = 1\}$. It is easy to deduce a minimal facial description of P^{11} :

$$P^{11} = \{x \in \mathbb{R}^4 \mid x_1 = 0, 0 \leq x_2 \leq 1, x_3 = 1, x_4 = 1\}.$$

The only constraint that is not satisfied by our fractional solution $y = (\frac{2}{3}, 0, 1, 1)$ is the equality constraint $x_1 = 0$. So, let us try to generate an inequality defining a facet of P^1 that is violated by y , lifting an equality system of P^{11} to facet defining inequalities of P^1 .

The equality system we use is induced by the equalities: $x_1 = 0$ and $x_3 = 1$. Consider the lifting set of this equality system:

$$\begin{aligned}\Delta^= &= \left\{ \delta \in \mathbb{R}^2 \mid \begin{array}{l} \delta_1 x_1 + \delta_2 (x_3 - 1) \leq 0 \\ \forall x \in S \text{ s.t. } x_3 = 0, x_4 = 1 \end{array} \right\} \\ &= \{ \delta \in \mathbb{R}^2 \mid \delta_2 \geq 0, \delta_2 \geq \delta_1 \}.\end{aligned}$$

$\Delta^=$ has two extreme rays, which we can represent by $(-1, 0)$ and $(1, 1)$ respectively. These rays correspond to the inequalities $x_1 \geq 0$ and $x_1 + x_3 \leq 1$. Apparently, these inequalities define facets of P^1 . The latter inequality is violated by our fractional solution. Applying traditional lifting we obtain a violated inequality defining a facet of P : $x_1 + x_3 + x_4 \leq 2$.

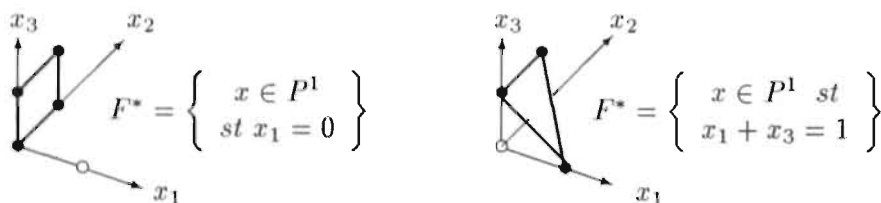


Figure 3.2: Liftings from the equality system

3.4 Extended Lifting

In this section we describe a generalization of the sequential lifting procedure, from now on denoted as the *extended lifting procedure*. With this procedure it is always possible to generate an inequality defining a facet of P from an inequality defining a facet of P^0 . The existence of a procedure with such a feature is indicated by the next proposition.

Proposition 3.4.1 *Let Q be a face of P , and let F_Q be a facet of Q . Then, there is at least one facet F_P of P such that $F_P \cap Q = F_Q$.*

Proof: Notice that Q is nonempty, since there exists a facet of Q . If $Q = P$ then the proposition is trivial, so assume $Q \neq P$. Let (D, d) denote an

equality system for the dimension gap between P and Q . The number of rows is at least one, because Q is a proper face of P . Since Q is a face of P , adding $Dx = d$ to a minimal facial description of P yields a facial (though in general not minimal) description of Q . The description is complete, so at least one of its inequalities defines F_Q , say the inequality $b^T x \leq b_0$. This is true, even if F_Q is empty. To see the latter, if F_Q is empty, Q is a point, and there is at least one facet of P not containing that point. The inequality $b^T x \leq b_0$ also defines a facet F_P of P . Therefore, $F_Q = F_P \cap \{x : Dx = d\} = F_P \cap Q$. $\square$

An important implication of this proposition is that given an inequality $a^T x \leq a_0$ defining a facet F_a of P^0 , there is an inequality $b^T x \leq b_0$ defining both a facet F_b of P and the facet F_a of P^0 . How to go from the coefficients (a^T, a_0) to the coefficients (b^T, b_0) is the topic of this section.

To identify all interesting equivalents of the inequality $a^T x \leq a_0$ we need to know the equality system (D, d) of the dimension gap between P and P^0 . Then, we are able to define for each inequality $a^T x \leq a_0$ a *lifting polyhedron*:

$$\Delta(a^T, a_0) := \{\delta \in \mathbb{R}^{p-p^0} \mid a^T x + \delta^T Dx \leq a_0 + \delta^T d \quad \text{for all } x \in S^1\}.$$

The lifting polyhedron can be interpreted as the projection of a truncation of the cone Δ^+ :

$$\Delta^+ := \{(\alpha, \delta) \in \mathbb{R}^{1+(p-p^0)} \mid \alpha a^T x + \delta^T Dx \leq \alpha a_0 + \delta^T d, \alpha \geq 0 \quad \forall x \in S^1\}.$$

Intersecting Δ^+ with the subspace induced by the equation $\alpha = 1$, and projecting out the α yields $\Delta(a^T, a_0)$. Since the dimension of the truncated cone and its projection are the same, it follows from the analysis of Chapter 2 that there is a one to one relation between the faces of the truncated cone and its projection. Now, it is easy to see that the next proposition holds:

Proposition 3.4.2 *Assume that $a^T x \neq a_0$ for some $x \in S^0$, and P^1 is nonempty. Then, the lifting polyhedron $\Delta(a^T, a_0)$ is nonempty and has at least one extreme point.*

Proof: First we exhibit an element $\delta \in \Delta(a^T, a_0)$ to show that $\Delta(a^T, a_0)$ is nonempty: $\delta_1 = a_0 - \max\{a^T x \mid x \in S^1\}$, and $\delta_i = 0$ for all $i \in \{2, \dots, p-p^0\}$ (recall that the first row of (D, d) represents $x_n = 0$; this solution corresponds to traditional lifting).

It can be shown that the cone Δ^+ has precisely one extreme point, i.e. the zero vector, and at least one extreme ray, with a reasoning analogue to the proof of Proposition 3.3.1. Since the intersection of the cone Δ^+ with

the subspace induced by the equation $\alpha = 1$ has been shown nonempty, this intersection has at least one extreme point. And so does $\Delta(a^T, a_0)$, as follows from the projective relation. $\square$

Definition 3.4 *An inequality of the form*

$$a^T x + \delta^T D x \leq a_0 + \delta^T d, \quad \delta \in \mathbb{R}^{p-p^0} \quad (3.5)$$

is called an extended lifting from P^0 to P of the inequality $a^T x \leq a_0$ if the inequality (3.5) is valid for P .

Clearly, (3.5) is a valid inequality for P if and only if the vector δ is in $\Delta(a^T, a_0)$. Moreover, if $a^T x \leq a_0$ defines a facet F_a of P^0 , then it follows from Proposition 3.4.1 that there is an inequality $b^T x \leq b_0$ defining F_a of P^0 as well as a facet F of P . Because $b^T x \leq b_0$ defines F_a , there is a strictly positive scalar α , a vector $\delta \in \Delta(a^T, a_0)$, and a vector $\lambda \in \mathbb{R}^{n-p}$, such that $(b^T, b_0) = \alpha(a^T, a_0) + \delta(D, d) + \lambda^T(E^=, e^=)$. This implies that there is an extended lifting of $a^T x \leq a_0$ defining the facet F of P .

Since an inequality defining a facet F cannot be a linear combination of inequalities that are valid but do not define F , the only vectors $\delta \in \Delta(a^T, a_0)$ that we have to consider for this particular extended lifting are the extreme points of $\Delta(a^T, a_0)$.

Proposition 3.4.3 *Assume the inequality $a^T x \leq a_0$ is valid for P^0 and defines a facet F_a of P^0 . Then, the following holds:*

The inequality $b^T x \leq b_0$ is valid for P and defines a facet F_b of P containing F_a but not P^0 if and only if there exists a unique $(\alpha, \delta, \lambda)$, with α a positive scalar, δ an extreme point of $\Delta(a^T, a_0)$, and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \alpha a^T + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha a_0 + \delta^T d + \lambda^T e^=$.

Proof:

Since the inequality $a^T x \leq a_0$ is valid for P^0 and defines a facet F_a of P^0 , (a^T, a_0) together with the rows of (D, d) form an equality system for the dimension gap between F_a and P . Applying Proposition 3.3.3 we have: the inequality $b^T x \leq b_0$ is valid for P and defines a facet F_b of P containing F_a if and only if there exists a unique $(\alpha, \delta, \lambda)$, (α, δ) on an extreme nonzero ray of Δ^+ , and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \alpha a^T + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha a_0 + \delta^T d + \lambda^T e^=$. It also follows from Proposition 3.3.3 that the inequality $b^T x \leq b_0$ is valid for P and defines a facet F_b of P containing P^0 if and

only if there exists a unique (δ, λ) , δ on an extreme nonzero ray of Δ^π , and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \delta^T D + \lambda^T E^\pi$, and $b_0 = \delta^T d + \lambda^T e^\pi$. Comparing these results, by the uniqueness of the coefficients we have to conclude that $\alpha \neq 0$ if and only if F_b does not contain P^0 . $\square$

All facet defining inequalities of P^1 can be lifted in a similar way. So, we can derive a facet defining inequality of P by lifting a facet defining inequality of P' one variable at a time. To get better acquainted with the extended lifting procedure, let us revisit the instance of the knapsack problem discussed in Example 3.3.2.

Example 3.4.1 *We have to optimize a linear objective function over $S := \mathbb{B}^4 \cap R$, where $R := \{x \in \mathbb{R}^4 \mid 3x_1 + 2x_2 + 2x_3 + 2x_4 \leq 6\}$. Again, let $P^1 := \{x \in P \mid x_4 = 1\}$, and $P^{11} := \{x \in P^1 \mid x_3 = 1\}$. The minimal facial description of P^{11} is:*

$$P^{11} = \{x \in \mathbb{R}^4 \mid x_1 = 0, 0 \leq x_2 \leq 1, x_3 = 1, x_4 = 1\}.$$

Now, let us try to use this information by lifting the facet defining inequalities of P^{11} to facet defining inequalities of P^1 . As before, the equality system we use, is induced by the equalities $x_1 = 0$ and $x_3 = 1$. First, consider the inequality $x_2 \leq 1$. We represent a lifting polyhedron of this inequality as follows:

$$\begin{aligned} \Delta(a^T, a_0) &= \Delta(0, 1, 0, 0; 1) \\ &= \left\{ \delta \in \mathbb{R}^2 \mid \begin{array}{l} x_2 + \delta_1 x_1 + \delta_2 (x_3 - 1) \leq 1 \\ \forall x \in S \text{ s.t. } x_3 = 0, x_4 = 1 \end{array} \right\} \\ &= \{\delta \in \mathbb{R}^2 \mid \delta_2 \geq 0, \delta_2 \geq \delta_1 - 1\}. \end{aligned}$$

This lifting polyhedron has only one extreme point: $\delta = (1, 0)$. This implies that the extended lifting $x_1 + x_2 \leq 1$ defines a facet of P^1 . An affinely independent set of solutions in $P^1 \cap S$ satisfying this inequality at equality is: $(0, 1, 1, 1)$, $(0, 1, 0, 1)$, and $(1, 0, 0, 1)$. The latter two correspond to the inequalities $\delta_2 \geq 0$ and $\delta_2 \geq \delta_1 - 1$ of the lifting polyhedron. The first one is in P^{11} . Notice that the inequality $x_1 + x_2 \leq 1$ cannot be derived by applying traditional sequential lifting, since that would have yielded $x_2 \leq 1$, which only defines a face of P^1 .

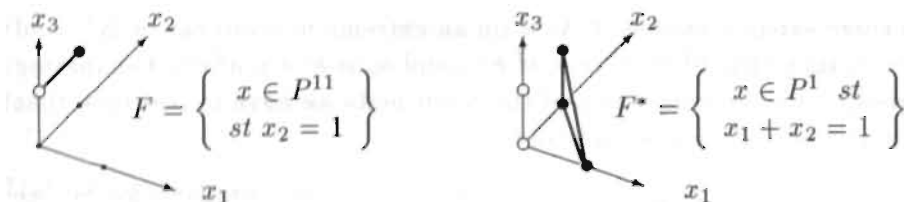


Figure 3.3: Lifting from the facet F of P^{11} defined by $x_2 \leq 1$

Secondly, consider the inequality $x_2 \leq 0$. The lifting polyhedron of this inequality is represented as follows:

$$\begin{aligned} \Delta(a^T, a_0) &= \Delta(0, -1, 0, 0; 0) \\ &= \left\{ \delta \in \mathbb{R}^2 \mid \begin{array}{l} -x_2 + \delta_1 x_1 + \delta_2 (x_3 - 1) \leq 0 \\ \forall x \in S \text{ s.t. } x_3 = 0, x_4 = 1 \end{array} \right\} \\ &= \{ \delta \in \mathbb{R}^2 \mid \delta_2 \geq 0, \delta_2 \geq \delta_1 \}. \end{aligned}$$

This lifting polyhedron has also one extreme point: $\delta = (0, 0)$. This implies that the extended lifting $-x_2 \leq 0$ defines a facet of P^1 . This may seem less spectacular than lifting the other inequality, since the resulting inequality is the same for traditional lifting. But traditional lifting would not have guaranteed that the inequality is facet defining.

The extended lifting of the facet defining inequalities of P^1 is equivalent to traditional lifting, since the equality set contains only one equality, equivalent to $x_4 = 1$. The reader can verify easily that the following inequalities result: $x_2 \geq 0$, and $x_1 + x_2 + x_4 \leq 2$, defining facets of P . Since the variables x_2 , x_3 , and x_4 are similar to a certain extent, we deduce that the inequalities $x_3 \geq 0$, $x_4 \geq 0$, $x_1 + x_2 + x_3 \leq 2$, and $x_1 + x_3 + x_4 \leq 2$, also define facets of P .

3.5 Complete lifting

In this section an algorithm is suggested generating a set of inequalities representing all facets of P , from a minimal facial description of P^0 . This algorithm is a generalization of the algorithms suggested in the two former sections. The next observation indicates that such a procedure might be possible.

Observation 3.5.1 *Let F be a facet of P . Then F contains a face of P^0 . Let F^0 be the face of P^0 that is the intersection of F and P^0 . If F^0 is empty, then F is defined by the trivial inequality $x_n \leq 1$.*

The latter part of this observation is typical for the convex hull of binary vectors. Observation 3.5.1 is of a slightly different nature than Observation 3.3.1 and Proposition 3.4.1. Observation 3.3.1 claims that there are facets of P corresponding to the equality system of P^0 . Proposition 3.4.1 indicates that for each facet of P^0 there is at least one facet of P . Observation 3.5.1 however does not claim that there is a facet of P for each face of P^0 , but points the other way around: there is some face of P^0 for each facet of P . Nevertheless, the construction of the procedure is analogue to the construction of the procedures presented in Sections 3.3 and 3.4.

As in the former sections, our first step is to characterize the set of inequalities, valid for P . Let (A, a) be a $m \times (n + 1)$ -dimensional matrix, such that every inequality that is valid for P^0 is implied by $Ax \leq a$ and the equality set of P^0 . We say that the matrix (A, a) with the equality set forms a *facial description* of P^0 . Furthermore, we assume that m is minimal. This implies that every row of (A, a) corresponds to an inequality defining a facet of P^0 . Now, we can define a *complete lifting cone* of the polytope P^0 :

$$\Delta^0 := \left\{ \alpha \in \mathbb{R}_+^m, \delta \in \mathbb{R}^{p-p^0} \mid \alpha^T Ax + \delta^T Dx \leq \alpha^T a + \delta^T d \quad \forall x \in P^1 \right\}.$$

The complete lifting cone of the polytope P^0 is a cone, for which we can prove the following proposition:

Proposition 3.5.1 *The cone Δ^0 has precisely one extreme point, i.e. the zero vector, and at least one extreme ray.*

Proof: Since Δ^0 is a cone, the zero vector is the only extreme point, if there is any. Suppose the zero vector $\underline{0}$ is not an extreme point of Δ^0 . Then, there is a nonzero $(\alpha, \delta) \in \Delta^0$ such that also $(-\alpha, -\delta) \in \Delta^0$. Since $\alpha \in \mathbb{R}_+^m$, we know $\alpha = \underline{0}$. But that implies $\delta^T Dx = \delta^T d$ for all $x \in S^1$, contradicting the definition of (D, d) . To prove that Δ^0 has an extreme ray, it is sufficient to show that Δ^0 has at least dimension one, since we have already concluded that Δ^0 has a face of dimension zero. So, it suffices to exhibit a nonzero element $(\alpha, \delta) \in \Delta^0$: $\alpha = \underline{0}$, $\delta_1 = -1$, and $\delta_i = 0$ for all $i \in \{2, \dots, p - p^0\}$. $\square$

Definition 3.5 *An inequality of the form*

$$\begin{aligned} (\alpha^T A + \delta^T D + \lambda^T E^-)x &\leq \alpha^T a + \delta^T d + \lambda^T e^- \\ \alpha \in \mathbb{R}^m, \delta \in \mathbb{R}^{p-p^0}, \lambda \in \mathbb{R}^{n-p}. \end{aligned} \quad (3.6)$$

is called a complete lifting from P^0 to P of the minimal facial description of P^0 if the inequality (3.5) is valid for P .

Clearly, inequality (3.5) is valid for P if and only if $(\alpha, \delta) \in \Delta^0$. Furthermore, if an inequality $b^T x \leq b_0$ defines a facet of P , then there is an $(\alpha, \delta) \in \Delta^0$ such that $(b^T, b_0) = (\alpha^T A + \delta^T D + \lambda^T E^=, \alpha^T a + \delta^T d + \lambda^T e^=)$. This observation can be strengthened, as reflected in the next proposition.

Proposition 3.5.2 *The inequality $b^T x \leq b_0$ is valid for P and defines a facet F_b of P if and only if there exists a $(\alpha, \delta, \lambda)$ with (α, δ) on an extreme ray of Δ^0 and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \alpha^T A + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha^T a + \delta^T d + \lambda^T e^=$.*

Proof:

$\Rightarrow$) The inequality $b^T x \leq b_0$ is valid for P , defines a facet F_b of P , and a face F^0 of P^0 . Then there is a $(p - p^0) \times (n + 1)$ -dimensional matrix (A', a') , whose rows are a linearly independent subset of the rows of (A, a) , such that $A'x = a'$ for all $x \in F^0$. Together with the rows of (D, d) , the rows of (A', a') form an equality system of the dimension gap between P and F^0 . It follows from Proposition 3.3.3 that there exists a unique $(\alpha', \delta, \lambda)$ with (α', δ) on an extreme ray of the lifting cone of the equality system of the dimension gap between P and F^0 , and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \alpha'^T A' + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha'^T a' + \delta^T d + \lambda^T e^=$.

Construct α such that $\alpha_i = 0$ if row i of (A, a) is not a row of (A', a') , and α_i equals the value of corresponding element in α' otherwise. Clearly, (α, δ) is in Δ^0 , λ is in $\mathbb{R}^{n-p}$, $b^T = \alpha^T A + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha^T a + \delta^T d + \lambda^T e^=$. For each set of vectors in Δ^0 summing up to (α, δ) , the elements of the vectors corresponding to the rows in (A, a) that are not in (A', a') must have value zero. This, combined with the fact that (α', δ) is on an extreme ray of the lifting cone of the equality system of the dimension gap between P and F^0 , implies that (α, δ) is on an extreme ray of Δ^0 , which completes this part of the proof.

$\Leftarrow$) Assume $(\alpha, \delta, \lambda)$ with (α, δ) is on an extreme ray of Δ^0 and $\lambda \in \mathbb{R}^{n-p}$, such that $b^T = \alpha^T A + \delta^T D + \lambda^T E^=$, and $b_0 = \alpha^T a + \delta^T d + \lambda^T e^=$. Then $b^T x \leq b_0$ is valid for P and defines a face F_b of P . Because (α, δ) is nonzero, F_b does not contain P . Since (α, δ) is on an extreme ray of Δ^0 , F_b is not in the intersection of two different facets of P . Therefore, F_b is a facet of P .

□

A consequence of this proposition is that we can find all inequalities defining a facet of P by checking all extreme rays of Δ^0 . Some rays may refer to the same facet, since a face of P^0 can be the intersection of various different sets of facets of P^0 . However, the extreme rays of Δ^0 all correspond

to some facet of P .

Again, as an illustration, we revisit Example 3.3.2, presented in Section 3.3.

Example 3.5.1 *We lift the polytope P^{11} completely. The lifting cone of P^{11} is:*

$$\begin{aligned} \Delta^{11} &= \left\{ \begin{array}{l} \alpha \in \mathbb{R}^2 \\ \delta \in \mathbb{R}^2 \end{array} \left| \begin{array}{l} \alpha_1 \geq 0, \alpha_2 \geq 0 \\ -\alpha_1 x_2 + \alpha_2(x_2 - 1) + \delta_1 x_1 + \delta_2(x_3 - 1) \leq 0 \\ \forall X \in S \text{ s.t. } x_3 = 0, x_4 = 1 \end{array} \right. \right\} \\ &= \left\{ \begin{array}{l} \alpha \in \mathbb{R}^2 \\ \delta \in \mathbb{R}^2 \end{array} \left| \begin{array}{l} \alpha_1 \geq 0 \\ \alpha_2 \geq 0 \\ \alpha_1 + \delta_2 \geq 0 \\ \alpha_2 + \delta_2 \geq 0 \\ \alpha_2 + \delta_2 \geq \delta_1 \end{array} \right. \right\}. \end{aligned}$$

So, there are five inequalities characterizing Δ^{11} . An extreme ray satisfies at least three linearly independent inequalities at equality. That is, there are at most ten different possibilities. We checked them all, and the result can be summarized as follows:

α_1	α_2	δ_1	δ_2	inequality	index
0	0	-1	0	$x_1 \geq 0$	(a)
0	0	1	1	$x_1 + x_3 \leq 1$	(b)
1	0	0	0	$x_2 \geq 0$	(c)
0	1	1	0	$x_1 + x_2 \leq 1$	(d)
1	1	0	-1	$x_3 \geq 0$	(e)

The inequalities (a) and (b) correspond to extreme rays of the lifting cone of the equality system of P^0 . The inequalities (c) and (d) are extended liftings of the inequalities $0 \leq x_2 \leq 1$. Inequality (e) is a trivial inequality, the lifting of the empty face. We could also have checked whether this defines a facet by checking the dimension of P^{01} . In itself, the resulting inequalities may not be very spectacular in this case. However, the complete lifting guarantees that together, these inequalities form a facial description of P^1 .

Finally, since we have a facial description of P^1 , we lift P^1 completely.

$$\begin{aligned} \Delta^1 &= \left\{ \begin{array}{l} \alpha \in \mathbb{R}_+^5 \\ \delta \in \mathbb{R} \end{array} \left| \begin{array}{l} -\alpha_1 x_1 - \alpha_2 x_2 - \alpha_3 x_3 + \delta(x_4 - 1) \\ + \alpha_4(x_1 + x_2 - 1) + \alpha_5(x_1 + x_3 - 1) \leq 0 \\ \forall X \in S \text{ s.t. } x_4 = 0 \end{array} \right. \right\} \\ &= \left\{ \begin{array}{l} \alpha \in \mathbb{R}_+^5 \\ \delta \in \mathbb{R} \end{array} \left| \begin{array}{ll} \delta \geq -\alpha_4 - \alpha_5 & \delta \geq -\alpha_2 - \alpha_5 \\ \delta \geq -\alpha_3 - \alpha_4 & \delta \geq \alpha_5 - \alpha_1 - \alpha_3 \\ \delta \geq -\alpha_1 & \delta \geq \alpha_4 - \alpha_1 - \alpha_2 \\ \delta \geq -\alpha_2 - \alpha_3 \end{array} \right. \right\}. \end{aligned}$$

This cone has 14 extreme rays:

α_1	α_2	α_3	α_4	α_5	δ	inequality	index
0	0	0	0	0	1	$x_4 \leq 1$	(f)
1	0	0	0	0	0	$x_1 \geq 0$	(g)
0	1	0	0	0	0	$x_2 \geq 0$	(h)
0	0	1	0	0	0	$x_3 \geq 0$	(i)
0	0	0	1	0	1	$x_1 + x_2 + x_4 \leq 2$	(j)
0	0	0	0	1	1	$x_1 + x_3 + x_4 \leq 2$	(k)
0	0	1	0	1	0	$x_1 \leq 1$	(l)
1	0	0	0	1	0	$x_3 \leq 1$	(m)
0	1	0	1	0	0	$x_1 \leq 1$	(n)
1	0	0	1	0	0	$x_2 \leq 1$	(o)
1	0	0	1	1	0	$x_1 + x_2 + x_3 \leq 2$	(p)
1	1	0	1	0	-1	$x_4 \geq 0$	(q)
1	0	1	0	1	-1	$x_4 \geq 0$	(r)
0	0	0	1	1	1	$2x_1 + x_2 + x_3 + x_4 \leq 3$	(s)

The inequalities (l)-(s) do not correspond to extreme rays of the extended lifting of inequalities or the equality system of P^1 . Inequality (s) does not occur in an extended lifting of inequalities or an equality system of P^0 either! Notice that the inequalities (l) and (n), as well as (q) and (r), are pairwise identical.

3.6 Strict Sequential Lifting

The sequential lifting procedures discussed so far, do guarantee a facet defining inequality for P , but require to check the extreme rays or extreme points of a $(p - p^0)$ -dimensional polytope. However, it is possible to lift the equalities of the equality system one by one. Thus, lifting an inequality from P^0 to P can also be done by checking the extreme points of $(p - p^0)$ one dimensional polytopes. This possibility will be briefly discussed in this section. The theoretical results supporting this approach, are of the same nature as the results in the former sections, so we illustrate the principles by means of an example.

In strict sequential lifting, we lift an inequality $p - p^0$ times with dimension gap one. So, we need to construct a sequence of polytopes $P_0, \dots, P_{p-p^0}$, starting with $P_0 = P^0$ and ending with $P_{p-p^0} = P$, such that P_{i-1} is a facet of P_i for all $i \in \{1, \dots, p - p^0\}$. One way to generate a sequence like that is

intersecting P_i by a hyperplane inducing a facet of P_i containing P^0 . The intersection will be identical to the facet, and form P_{i-1} . Such a facet corresponds to an extreme ray of the lifting cone of the equality system of the dimension gap between P_i and $P_0 = P^0$, as discussed in Section 3.3.

Again, as an illustration, we revisit Example 3.3.2 presented in Section 3.3.

Example 3.6.1 *We will lift the inequalities $x_2 \leq 1$ and $x_2 \geq 0$ from P^{11} to P^1 in a strict sequential way.*

To generate a sequence of polytopes with dimension gap one from P^{11} to P^1 we intersect P^1 with a hyperplane inducing a facet of P^1 containing P^{11} . There are two such facets corresponding to the extreme rays of the lifting cone of the equality system of the dimension gap between P^1 and P^{11} , as discussed in Example 3.3.1. Inequalities inducing these facets are $x_1 + x_3 \leq 1$ and $x_1 \geq 0$. We choose one arbitrarily, say $x_1 + x_3 \leq 1$. So, the sequence of polytopes is: $P_0 = P^{11}$, $P_1 = \{x \in P_2 | x_1 + x_3 = 1\}$, and $P_2 = P^1$. An equality system for the dimension gap between P_2 and P_1 is $x_1 + x_3 = 1$. An equality system for the dimension gap between P_1 and P_0 is $x_1 = 0$.

Now, liftings of the inequalities $x_2 \leq 1$ and $x_2 \geq 0$ from P_0 to P_1 are inequalities, valid for P_1 , of the forms $x_2 + \delta x_1 \leq 1$, respectively $-x_2 + \delta x_1 \leq 0$. These inequalities should be satisfied by all $x \in S$ such that $x_4 = 1$, $x_1 + x_3 = 1$. All solutions for which $x_1 = 1$ yield an upperbound for δ . There is only one such a solution: $x = (1, 0, 0, 1)$. For the first inequality we derive 1 as an upper bound for δ , for the second one we find 0. So the inequalities $x_1 + x_2 \leq 1$ and $x_2 \geq 0$ define facets of P_1 .

Liftings of the facet defining inequalities $x_1 + x_2 \leq 1$ and $x_2 \geq 0$ from P_1 to P_2 are inequalities, valid for P_2 , of the form $x_1 + x_2 + \delta(x_1 + x_3) \leq 1 + \delta$, respectively $-x_2 + \delta(x_1 + x_3) \leq 0 + \delta$. These inequalities should be satisfied by all $x \in S$ such that $x_4 = 1$. All solutions for which $x_1 + x_3 \neq 1$ yield an upperbound for δ . There are two solutions satisfying these conditions: $x = (0, 0, 0, 1)$ and $x = (0, 1, 0, 1)$. In both cases we find 0 as an upper bound for δ . So the inequalities $x_1 + x_2 \leq 1$ and $x_2 \geq 0$ define facets of $P_2 = P^1$.

We could have chosen our sequence of polytopes differently, say: $P_0 = P^{11}$, $P_1 = \{x \in P_2 | x_1 = 0\}$, and $P_2 = P^1$. In this example that does not influence the result. In general, it could yield a different result, just as traditional lifting could yield a different result for different orderings of the variables to be lifted.

3.7 An Example on the Cell Formation Polytope

In Chapter 6 a model is suggested for the cell formation problem. In this problem we have a set of M machines and a set of N parts. Each machine and each part has to be assigned to at most one cell, such that some objective function is optimized. The number of cells to be formed is part of the problem. The model is based on binary variables $x_{ij} \in \{0, 1\}$. The interpretation of the variables is that x_{ij} has value one if and only if machine i and part j are in the same cell, and x_{ij} has value zero in all other cases. It is shown that a solution $x \in \mathbb{B}^{M \times N}$ is feasible if and only if it satisfies all of the following so-called *square inequalities*:

$$x_{ij} + x_{hj} + x_{ik} - x_{hk} \leq 2 \quad \begin{array}{l} \forall \quad h, i \in \{1, \dots, M\} \\ \forall \quad j, k \in \{1, \dots, N\} \end{array} \quad (3.7)$$

The inequalities 3.7 have an obvious interpretation: if machine i and part j are in the same cell, and if machine i and part k are in the same cell, then if machine h and part j are also in the same cell, machine h and part k are bound to be in the same cell too.

In Chapter 7 and Chapter 8 the facial structure of the convex hull of the feasible solutions is discussed. It turns out that this polytope has all kind of nice properties. For example, it is of full dimension. As for the square inequalities, they are facet defining, and so are the the trivial inequalities. Actually, together these inequalities form a minimal facial description of the polytope if the number of machines equals two.

Suppose now that we have to adapt our model, because we are given an extra opportunity: for some part there is an alternative production plan, requiring a different set of machines. This is not an unusual scenario. The problem of having to choose between different production plans for a part is known in literature as *concurrent engineering*, see for a more thorough discussion of this topic [Ham et al, 1985].

To start with, let us keep things simple. Assume there are only two machines and two parts. Let part 1 have a set of production plans, and let $S_1 = \{1, \dots, s_1\}$ be the index set of these plans. Now, for each production plan there is a binary variable $p^l \in \{0, 1\}$ that has value one if and only if production plan l is chosen. Assuming we can choose at most one production plan, we have: $\sum_{l \in S_1} p^l \leq 1$. We also need more x -variables for part 1, with an extra index per process plan: x_{ij}^l has value one if and only if machine i and part j are in the same cell, and production plan l is in use.

First, we take only one process plan into consideration: $s_1 = 1$. Let P

be the convex hull of all feasible solutions. Then the equality $p^1 = 1$ defines a face F of P , and for this face we know a minimal facial description:

$$F = \left\{ \begin{array}{l} p^l \in \mathbb{R} \quad \forall l \in S_1 \\ x_{i1}^l \in \mathbb{R} \quad \forall i \in \{1, 2\}, l \in S_1 \\ \hline p^l = 0 \quad \forall l \in S_1 \setminus \{1\} \\ p^1 = 1 \\ x_{i1}^l = p^l \quad \forall l \in S_1 \setminus \{1\} \\ x_{11}^1 + x_{12} + x_{21}^1 - x_{22} \leq 2 \\ x_{11}^1 + x_{12} - x_{21}^1 + x_{22} \leq 2 \\ x_{11}^1 - x_{12} + x_{21}^1 + x_{22} \leq 2 \\ -x_{11}^1 + x_{12} + x_{21}^1 + x_{22} \leq 2 \\ 0 \leq x_{11}^1 \leq 1 \\ 0 \leq x_{12} \leq 1 \\ 0 \leq x_{21}^1 \leq 1 \\ 0 \leq x_{22} \leq 1 \end{array} \right\}$$

We can lift each of these inequalities with respect to p^1 . Since the dimension gap between P and F is one, ordinary sequential lifting suffices to derive facet defining inequalities for P . We add to this the inequality derived by lifting the equality system. Actually, we find the following approximation $\hat{P}$ of P :

$$\hat{P} = \left\{ \begin{array}{l} p^l \in \mathbb{R} \quad \forall l \in S_1 \\ x_{i1}^l \in \mathbb{R} \quad \forall i \in \{1, 2\}, l \in S_1 \\ \hline p^l = 0 \quad \forall l \in S_1 \setminus \{1\} \\ p^1 \leq 1 \\ x_{i1}^l = p^l \quad \forall l \in S_1 \setminus \{1\} \\ -p^1 + x_{11}^1 + x_{12} + x_{21}^1 - x_{22} \leq 1 \\ x_{11}^1 + x_{12} - x_{21}^1 + x_{22} \leq 2 \\ -p^1 + x_{11}^1 - x_{12} + x_{21}^1 + x_{22} \leq 1 \\ -x_{11}^1 + x_{12} + x_{21}^1 + x_{22} \leq 2 \\ 0 \leq x_{11}^1 \leq p^1 \\ 0 \leq x_{12} \leq p^1 \\ 0 \leq x_{21}^1 \leq p^1 \\ 0 \leq x_{22} \leq p^1 \end{array} \right\}$$

and in this special case, one can prove that the approximation is exact and the description minimal.

Now let us make the case more interesting, and bring into consideration the second process plan: $s_1 = 2$. We interpret P as the face of a larger polytope Q , the convex hull of all feasible solutions for two process plans, that is induced by intersecting Q with the subspace defined by the equality $p^2 = 0$. This equality does not induce an equality system, since the dimension gap between Q and P is three. An inequality system is: $p^2 = 0$, $x_{11}^2 = p^2$, and $x_{21}^2 = p^2$. The lifting cone of the inequality $-p^1 + x_{11}^1 + x_{12} + x_{21}^1 - x_{22} \leq 1$ is:

$$L = \left\{ \begin{array}{l} \alpha \in \mathbb{R} \\ \beta \in \mathbb{R} \\ \gamma \in \mathbb{R} \end{array} \left| \begin{array}{l} \gamma - \alpha - \beta \leq 0 \\ \gamma - \beta \leq 0 \\ \gamma - \alpha \leq 0 \\ \gamma \leq 1 \end{array} \right. \right\}$$

This polyhedron has two extreme points: $(0, 0, 0)$ and $(1, 1, 1)$. These correspond to respectively the following inequalities:

$$-p^1 + x_{11}^1 + x_{12} + x_{21}^1 - x_{22} \leq 1$$

and

$$-p^1 - p^2 + x_{11}^1 + x_{11}^2 + x_{12} + x_{21}^1 + x_{21}^2 - x_{22} \leq 1$$

The first one could have been derived by applying ordinary sequential lifting, but without the guarantee that it defines a facet of Q . The second one is due to extended lifting, though it could also have been derived by strict sequential lifting. In the latter case we would have to study the lifting cone of a three dimensional equality system, and still it is not a priori clear which sequence yields interesting results. The second type of inequalities give some extra insight that leads to a straightforward generalization.

Let S_l be the index set of concurrent project plans for part l . Then, the following inequality is facet defining for the polytope that is the convex hull of the feasible solutions of the concurrent engineering cell formation problem:

$$-\sum_{l \in T_j} p_j^l + \sum_{l \in T_j} x_{ij}^l + \sum_{l \in T_j} x_{hj}^l + x_{ik}^n - x_{jk}^n \leq 1$$

for all $n \in S_k$, for all $T_j \subseteq S_j$ such that $T_j \neq \emptyset$, for all $h, i \in \{1 \dots M\}$, for all $j, k \in \{1 \dots N\}$. We refer to this class of inequalities as *concurrent square inequalities*. Although this is an exponential class of inequalities, it is easy to see that the separation problem is polynomial solvable. Other inequalities that are facet defining for this polytope are $\sum_{l \in S_l} p^l \leq 1$, $x_{ij}^l \geq 0$, and $x_{ij}^l \leq p^l$. The latter two types of inequalities are the result of lifting the equality system.

3.8 Polarity, Duality, and Lifting

Beginning with a minimal facial description of P' on a subset of the variables, sequentially complete lifting yields a facial description of the original polytope P . This could be interpreted as a directed recursive procedure to find the extreme rays of the polar cone $\Pi^0(P)$ of P . To see this, notice that we could start lifting from the empty set. For simplicity, assume P is full dimensional. Take the equalities $x_i = 0$ for $i \in \{1, \dots, n\}$ together with the equality $0 = 1$ as a facial description of the empty set. These equalities also form the equality system (notice that there are $p - p^0 = n + 1$ equalities) of the lifting gap between P and the empty set. Then the lifting cone of this equality system is equivalent to the polar cone of P :

$$\Delta^0 = \{(\delta, \delta_{n+1}) \in \mathbb{R}^{n+1} \mid \delta^T x \leq \delta_{n+1} \quad \forall x \in S\}.$$

Another way of interpreting complete lifting is via duality theory. Suppose we maximize $c^T x$ over S^0 and we have a facial description of P^0 . That is equivalent to minimizing $\alpha^T a + \delta^T d + \lambda^T e =$ over $\alpha, \delta, \lambda \in \mathbb{R}^{m+(p-p^0)+(n-p)}$ such that $\alpha \geq 0$ and $c^T = \alpha^T A + \delta^T D + \lambda^T E =$. Extending S^0 with S^1 implies extra conditions on the dual, namely $(\alpha, \delta) \in \Delta^0$.

To get affiliated with this interpretation via duality theory, consider the extreme case where S^0 is empty. Like before, we chose the following equalities as a facial description for the empty set: $x_i = 0$ for all $i \in \{1, \dots, n\}$, and $0 = 1$. Although the primal problem does not have a solution, we still can formulate a dual. The dual would be to minimize δ_{n+1} such that $c_i = \delta_i$ for all $i \in \{1, \dots, n\}$. Extending the empty set S^0 with a nonempty set S^1 , the following inequalities are added as restrictions to the dual: $\delta^T x (= c^T x) \leq \delta_{n+1}$ for all $x \in S^1$.

In general, we cannot expect that sequential complete lifting is very efficient. If the structure of a problem is exploited in order to keep the method as simple as possible, the method might in some cases be suitable to derive a facial description of P , or to be used as a tool to prove the completeness of a description.

We also introduced and discussed extended lifting of inequalities and the equality system. An equality system of P' can be lifted to facet defining inequalities of the original polytope P . Also, together with an equality system of P' an inequality defining a facet of P' can be lifted to facet defining inequalities of the original polytope P , either using all equalities of the equality system at once, or one by one (under certain conditions on the equality system). The sum of the dimensions of the polytopes of which we have to

determine extreme rays, respectively extreme points, is $\dim(P) - \dim(P')$. Again, dependent on the possibilities to use the structure of the underlying problem, these procedures can be appropriate in some cases.

There are several differences between the extended lifting procedure and the simultaneous lifting procedure suggested by [Zemel,1978]. With the extended lifting procedure it is always possible to lift an inequality one variable at a time. It is even possible to lift equality by equality, with strictly sequential lifting. On the obverse, more knowledge is required about the equality systems.

Chapter 4

On the Strength of Inequalities

Recently, in [Goemans, 1995] an indicator has been suggested to compare classes of linear inequalities with respect to their potential effectiveness in a cutting plane algorithm, geared on optimizing linear objective functions over a polyhedron. He showed that this indicator has a nice interpretation in case the polyhedron is of blocking type and the objective is to minimize a linear function with nonnegative coefficients.

In this Chapter we show that the indicator has a similar interpretation for maximizing any linear objective function in case the description in the origin is identical for a relaxation polytope, and its strengthenings, the polytopes resulting from intersecting the relaxation polytope with the halfspaces induced by the respective families of valid inequalities. Some examples are added to illustrate the result.

4.1 Introduction

For a wide variety of problems cutting plane algorithms have been implemented successfully. The process of designing such an algorithm usually starts with a thorough investigation of the facial structure of polyhedra corresponding to the problem under study, yielding classes of valid inequalities. For some classes separating violated inequalities is harder than for others. In fact, for many classes the separation problem is *NP-hard*, indicating we might have to settle for some separation heuristic.

Ideally, one would balance the effort of separation with the improvement of the relaxation of the polytope under research. It is not clear how to measure this improvement. Generally, there is a preference for classes of inequalities that define facets of the convex hull of feasible solutions. But then again we could use some measure for the improvement of the relaxation resulting from adding each of the classes of facet defining inequalities we derived.

Recently, in [Goemans, 1995] a measure is introduced, let us denote it as $\Gamma_{a,b}^{\geq}(P)$, for the strength of an inequality $a^T x \geq b$ with respect to a polytope P :

$$\Gamma_{a,b}^{\geq}(P) = \frac{b}{\min\{a^T x | x \in P\}}.$$

Of course, the strength of an inequality $a^T x \leq b$, with respect to a polytope P , denoted as $\Gamma_{a,b}^{\leq}(P)$, can be written as:

$$\Gamma_{a,b}^{\leq}(P) = \frac{b}{\max\{a^T x | x \in P\}}.$$

Goemans shows that in specific cases this strength has a nice interpretation. The result is based on the following observation.

Observation 4.1.1 *Let P and Q be convex sets in the positive quadrant such that $Q \subseteq P \subseteq \mathbb{R}_+^n$ and $0 \notin P$. So Q is a strengthening of P . Assume both sets are of blocking type; that is, a set S is of blocking type if $x \in S$ and $y \geq x$ imply $y \in S$. Denote $\{\alpha x | x \in P\}$ where α is a scalar as αP . Then, there exists a scalar $\alpha \geq 1$ such that $\alpha P \subseteq Q \subseteq P$.*

Now add the assumption that P and Q are polyhedra. We can write $Q = \{x \in P | a_i^T x \geq b_i, \forall i \in N\}$ for some nonempty index set N representing a family of inequalities. For each inequality $a_i^T x \geq b_i, i \in N$, we compute the indicator $\Gamma_{a_i b_i}^{\geq}(P)$. The maximum of these indicators is denoted as $\Gamma^{\geq}(P, Q)$. Define a strengthening Q_i of P as follows: $Q_i = \{x \in P | a_i^T x \geq b_i\}$. Clearly, $Q = \bigcap_{i \in N} Q_i$ and $\min_x \{c^T x | x \in Q\} \geq \max_{\{i \in N\}} \min_x \{c^T x | x \in Q_i\}$. It is shown in [Goemans, 1995] that now can derive:

$$\begin{aligned} \min \{c^T x | x \in Q\} &\leq \alpha \min \{c^T x | x \in P\} \quad \forall c \in \mathbb{R}_+^n \\ &\iff \alpha \geq \Gamma^{\geq}(P, Q). \end{aligned} \quad (4.1)$$

In other words, adding the inequalities corresponding with N to the polytope P the objective function value cannot increase more than α times the optimal value over P . One of the interesting features of this result is that to compute the indicator $\Gamma^{\geq}(P, Q)$ only the cases in which a single inequality is added to P have to be considered.

Of course, for each polyhedron P we can define a polytope $\text{dom}(P)$ which is the minimal polyhedron of blocking type containing P . But there is not always a one to one relation between families of inequalities valid for P and families of valid inequalities for $\text{dom}(P)$. For example, consider a polyhedron P containing the origin and all unit vectors. Clearly, $\text{dom}(P)$ is in the positive quadrant and contains P , but the analysis described above is not of much use. This raises the question whether there are other polyhedra for which the Goemans measure of strength has a similar and meaningful interpretation.

In the next section we present such a class of polyhedra. The result is indicated by the next observation, which is analogue to Observation 4.1.1.

Observation 4.1.2 *Let P and Q be convex sets containing the origin as an interior point (so P and Q are of full dimension). Assume $Q \subseteq P$. Then, there exists an $\alpha \in (0, 1)$ such that: $\alpha P = \{\alpha x | x \in P\} \subseteq Q \subseteq P$.*

It is always possible to shift a nonempty polytope in such a way that it contains the origin. However, if a polytope is not of full dimension, it does not have an interior. To cope with this, we introduce the concept of zero-identical polyhedra, and claim that it might be a usefull concept to a wide variety of problems.

Definition 4.1 *Two polyhedra, say P and Q , are called **zero-identical** if and only if for all $a \in \mathbb{R}^n$:*

$$a^T x \leq 0 \quad \forall x \in P \iff a^T x \leq 0 \quad \forall x \in Q.$$

In the next section we will study the interpretation of the indicator $\Gamma^{\leq}(P, Q)$ as a worst case ratio comparing the linear relaxation P and its strengthening Q under the assumption that P and Q are zero-identical. In Section 4.3 we will illustrate the results with some examples.

4.2 Worst Case Analysis for Zero-Identical Relaxation Polyhedra

If a polyhedron P is strengthened to a polyhedron Q by adding a set of additional inequalities, but none of these inequalities is violated by the origin, or satisfied at equality by the origin, then P and Q are zero-identical. In other words, we can add additional inequalities of the form $a^T x \leq b$ as many as we judge necessary, as long as b is strictly positive, the relaxation polytope and its strengthening will be zero-identical.

The concept of zero-identical polyhedra is getting more interesting if the origin is a feasible solution. We can prove the following:

Proposition 4.2.1 *Let P and Q be polytopes, and $\underline{0} \in Q \subseteq P$. Then, the following statements are equivalent:*

- a) P and Q are zero-identical;
- b) there exists a $\gamma \in (0, 1]$ such that:

$$\alpha P = \{\alpha x | x \in P\} \subseteq Q \quad \forall \alpha \in (0, \gamma];$$

- c) there exists a $\gamma \in (0, 1]$ such that for any $c \in \mathbb{R}^n$:

$$\min\{c^T x | x \in Q\} \leq \alpha \min\{c^T x | x \in P\} \quad \forall \alpha \in (0, \gamma].$$

Proof: a) $\Rightarrow$ b): Let $X = \{x^1, x^2, \dots, x^k\}$ be the set of all extreme points of P . It is sufficient to show that for all $x \in X$ there is a strictly positive scalar $\beta \in \mathbb{R}$ such that $\beta x \in Q$.

Let $x^j \in X$. If $x^j \in Q$ then we set $\beta_j = 1$. So assume $x^j \notin Q$. That implies an inequality $a^T x \leq b$ that is valid for Q but violated by x^j : $a^T x^j > b > 0$. Without loss of generality, let $a^T x \leq b$ be the inequality for which $b/(a^T x^j)$ is minimal. Take $\beta_j = b/(a^T x^j)$. Clearly, $\beta_j \in (0, 1)$ and $\beta_j x^j \in Q$. Finally, take $\gamma = \min_j \beta_j$.

b) $\Rightarrow$ c): Since $\alpha P \subseteq Q$, it follows (via the separating hyperplane theorem) that we have for all $c \in \mathbb{R}^n$: $\min\{c^T x | x \in Q\} \leq \min\{c^T x | x \in \alpha P\}$. The term on the right hand side can be reformulated to: $\min\{c^T(\alpha x) | x \in P\} = \alpha \min\{c^T x | x \in P\}$.

c) $\Rightarrow$ a): By contradiction, suppose that Q and P are not zero-identical. Then there should exist an inequality of the form $d^T x \geq 0$ that is valid for all elements in Q , but not for all elements in P (not the other way around, since $Q \subseteq P$). Let $y \in P$ such that $d^T y < 0$. Now, take the vector d as the cost vector c . Since α is strictly positive, we get the following contradiction:

$$0 \leq \min\{d^T x | x \in Q\} \leq \alpha \min\{d^T x | x \in P\} \leq \alpha d^T y < 0$$

□

Notice that P and Q are assumed to be polytopes. The results also hold in case they are polyhedra, but then we have to require that r is a ray of P if and only if r is a ray of Q .

Furthermore, statement c) can be reformulated to:

c') there exists a $\gamma \in (0, 1]$ such that for any $c \in \mathbb{R}^n$:

$$\max\{c^T x | x \in Q\} \geq \alpha \max\{c^T x | x \in P\} \quad \forall \alpha \in (0, \gamma].$$

In statements b) and c) of Proposition 4.2.1 the existence of some scalar γ is mentioned. Of course we are interested in the minimum value, say γ^* , that this scalar could have. That is:

$$\gamma^* = \inf_{c \in \mathbb{R}^n} \left\{ \frac{\max\{c^T x | x \in Q\}}{\max\{c^T x | x \in P\}} \right\} = \sup_{c \in \mathbb{R}^n} \left\{ \frac{\min\{c^T x | x \in Q\}}{\min\{c^T x | x \in P\}} \right\},$$

where by convention $0/0$ equals 1.

Under the assumptions of Proposition 4.2.1, it follows from statement c') and the fact that the origin is in P and Q , that $\max\{c^T x | x \in P\} \geq \max\{c^T x | x \in Q\} \geq 0$, and so $\gamma^* \in (0, 1]$.

We can assume that $Q = \{x \in P \mid a_i^T x \leq b_i \ \forall i \in N\}$ where N is a finite nonempty index set representing a family of inequalities. For each inequality $a_i^T x \leq b_i$, $i \in N$, we compute the strength indicator $\Gamma_{a_i, b_i}^{\leq}(P)$, defined in Section 4.1. Similar to [Goemans, 1995], we define $\Gamma^{\leq}(P, Q)$ to be the minimum of the indicators $\Gamma_{a_i, b_i}^{\leq}(P)$, and show in the next proposition that γ^* is a minimum and equals $\Gamma^{\leq}(P, Q)$. The proof of the proposition is analogue to the proof Goemans presents. It has been added for the sake of completeness.

Proposition 4.2.2 *Let P and Q be zero-identical polytopes, and $0 \in Q \subseteq P$. Let γ^* and $\Gamma^{\leq}(P, Q)$ be as defined above. Then: $\gamma^* = \Gamma^{\leq}(P, Q)$.*

Proof: First we show that $\gamma^* \leq \frac{b_i}{\max\{a_i^T x \mid x \in P\}} = \Gamma_{a_i, b_i}^{\leq}(P)$ for all $i \in N$. This is nearly trivial:

$$\gamma^* = \inf_{c \in \mathbb{R}^n} \left\{ \frac{\max\{c^T x \mid x \in Q\}}{\max\{c^T x \mid x \in P\}} \right\} \leq \frac{\max\{a_i^T x \mid x \in Q\}}{\max\{a_i^T x \mid x \in P\}} \leq \frac{b_i}{\max\{a_i^T x \mid x \in P\}}$$

for all $i \in N$.

So, it remains to be shown that $\gamma^* \geq \min_{i \in N} \frac{b_i}{\max\{a_i^T x \mid x \in P\}} = \Gamma^{\leq}(P, Q)$.

We are going to use duality, therefore we need an explicit description of P . We can assume that $P = \{x \in \mathbb{R}^n \mid a_i^T x \leq b_i \ \forall i \in M\}$, where M is a finite index set representing a set of inequalities.

By strong duality:

$$\begin{aligned} \max \quad & \{c^T x\} \\ \text{s.t.} \quad & \sum_{j=1}^n a_{ij} x_j \leq b_i \\ & \forall i \in N \cup M \end{aligned} \quad = \quad \begin{aligned} \min \quad & \{\sum_{i \in N \cup M} b_i y_i\} \\ \text{s.t.} \quad & \sum_{i \in N \cup M} a_{ij} y_i = c_j \\ & \forall j \in \{1, \dots, n\} \\ & y_i \geq 0 \\ & \forall i \in N \cup M. \end{aligned}$$

Let y^* be an optimal dual solution of the program above. Then we have:

$$\begin{aligned} \max_{x \in P} \{c^T x\} &= \max_{x \in P} \left\{ \sum_{j=1}^n \left(\sum_{i \in N \cup M} a_{ij} y_i^* \right) x_j \right\} \\ &\leq \sum_{i \in N \cup M} (\max\{a_i^T x \mid x \in P\}) y_i^*. \end{aligned}$$

Hence,

$$\begin{aligned} \gamma^* &\geq \frac{\max\{c^T x \mid x \in Q\}}{\max\{c^T x \mid x \in P\}} \\ &\geq \frac{\sum_{i \in N \cup M} b_i y_i^*}{\sum_{i \in N \cup M} (\max\{a_i^T x \mid x \in P\}) y_i^*} \\ &= \sum_{i \in N \cup M} \left(\frac{(\max\{a_i^T x \mid x \in P\}) y_i^*}{\sum_{i \in N \cup M} (\max\{a_i^T x \mid x \in P\}) y_i^*} \right) \frac{b_i}{\max\{a_i^T x \mid x \in P\}}. \end{aligned}$$

The latest expression can be interpreted as a weighted sum of $\frac{b_i}{\max\{a_i^T x | x \in P\}}$ since $(\max\{a_i^T x | x \in P\})y_i^* \geq 0$ for all $i \in N \cup M$. Therefore, we have:

$$\gamma^* \geq \min_{i \in N \cup M} \frac{b_i}{\max\{a_i^T x | x \in P\}} \geq \min_{i \in N} \frac{b_i}{\max\{a_i^T x | x \in P\}}.$$

□

The study of less than full dimensional polyhedra is often approached via the study of its full dimensional relaxations. There are many possibilities for deriving full dimensional relaxations. Standard relaxations are known as the submissive (downward monotization) and the dominant (upward monotization) (see [Pulleyblank, 1989, Cunningham and Green-Krotki, 1986, Balas and Fischetti, 1993]). For a polyhedron P in the positive quadrant, these monotizations are defined as follows:

$$\begin{aligned} \text{dom}(P) &= \{ y \in \mathbb{R}^n \mid y \geq x \text{ for some } x \in P \}, \\ \text{sub}(P) &= \{ y \in \mathbb{R}^n \mid 0 \leq y \leq x \text{ for some } x \in P \}. \end{aligned}$$

Suppose we are investigating a polytope Q that is not full dimensional, and its relaxation P in the same affine hull, both in the positive quadrant. Assume also that we have derived some inequalities to strengthen P with respect to Q . Then we can bring the inequalities in a form such that they strengthen $\text{dom}(P)$ with respect to $\text{dom}(Q)$ and use the analysis in [Goemans, 1995] to analyse the strength of the inequalities. Similarly, we can bring the inequalities in a form, most likely different from the form for the dominants, such that they strengthen $\text{sub}(P)$ with respect to $\text{sub}(Q)$ and use the analysis presented here to analyse the strength of the inequalities.

4.3 Some Examples

In this section the theoretical results of the former section are illustrated. For many combinatorial problems, the zero solution and all unit vectors are feasible solutions. This implies that the convex hull of the feasible solutions is a full dimensional polytope, all nonnegativity constraints are facet defining, and it implies that these constraints are the only inequalities defining facets containing the zero solution. Any linear relaxation of the problem that satisfies the nonnegativity constraints is zero-identical to all its strengthenings. The examples are both of this nature.

In the final section we compare the examples. The inequalities of the first example seem stronger than the inequalities of the second example.

However, the polytope of the second example is a projection of the polytope of the first example. From the projective relation between the problems it follows that relaxation of the first polytope can never be better than the relaxation of the second. In other words, here we discovered a huge class of objective functions for the first polytope on which the inequalities perform much worse than the strength indicator suggests, illustrating the limited value of the strength analysis.

4.3.1 The Clique Partitioning Polytope

The clique partitioning problem can be described as follows. Given is a complete graph $G = (V, E)$, $n = |V| \geq 3$, $m = n(n-1)/2$, together with edge weights $w_{ij} \in \mathbb{R}$ for all $\{i, j\} \in E$. A subset $D \subseteq E$ of the edges is called a *clique partition* if the components of the graph $H = (V, D)$ are cliques, i.e. complete subgraphs. The problem is to find a clique partition of maximal weight.

Mathematically, the clique partition problem can be formulated as follows.

$$\begin{array}{ll} \text{Maximize} & \sum_{\{i,j\} \in E} x_{ij} \\ \text{subject to} & x_{ij} + x_{ik} - x_{kj} \leq 1 \quad \text{for all } i, j, k \in V \\ & x_{ij} \in \{0, 1\} \quad \text{for all } \{i, j\} \in V \end{array}$$

The clique partitioning polytope $P(G)$ is defined as the convex hull of all feasible solutions to the above model. This polytope has been studied extensively, see for example [Grötschel and Wakabayashi, 1989]. The inequalities in the integer program are called *triangle inequalities*, and they define facets of $P(G)$. Also, the nonnegativity inequalities define facets of $P(G)$. This implies that we can apply the theory of this chapter fairly easily. The linear relaxation of the above model is zero-identical to the unit cube. So, the strength of an inequality is simply the right hand side divided by the sum of the positive coefficients in the left hand side of the inequality. Therefore, the strength of the class of the triangle inequalities is $1/2$. In other words, optimizing over the relaxation induced by the triangle inequalities will never yield an objective value lower than half of the sum of the positive coefficients of the objective function.

Now consider the following straightforward generalization of the triangle inequalities. Let $s \in V$, and $T \subset V$ such that $s \notin T$. Then,

$$\sum_{j \in T} x_{sj} - \sum_{\{i,j\} \in T \times T} x_{ij} \leq 1$$

defines a facet of $P(G)$. The strength of these inequalities is $1/|T|$. From this we can conclude that the relaxation of $P(G)$ induced by the triangle inequalities can be arbitrarily bad.

4.3.2 The Cell Formation Polytope

The cell formation problem can be described as follows, as will be explained in Chapter 6. Given is a complete bipartite graph $B = (U, W, E)$, $|U|, |W| \geq 2$, together with edge weights $w_{ij} \in \mathbb{R}$ for all $\{i, j\} \in E$. A subset $D \subseteq E$ of the edges is called a *bipartite clique partition* if the components of the graph $H = (V, D)$ are complete bipartite subgraphs. The problem is to find a bipartite clique partition of maximal weight.

Mathematically, the cell formation problem can be formulated as follows.

$$\begin{array}{ll} \text{Maximize} & \sum_{i \in U} \sum_{j \in W} x_{ij} \\ \text{subject to} & x_{ij} + x_{ik} + x_{hj} - x_{hk} \leq 2 \quad \text{for all } h, i \in U, j, k \in W \\ & x_{ij} \in \{0, 1\} \quad \text{for all } i \in U, j \in W \end{array}$$

The clique partitioning polytope $P(B)$ is defined as the convex hull of all feasible solutions to the above model. This polytope will be discussed extensively in Chapter 7 and Chapter 8. The inequalities in the integer program are called *square inequalities*, and they define facets of $P(B)$. Also, the non-negativity inequalities define facets of $P(B)$. Again, this implies that we can apply the theory of this chapter fairly easily. The linear relaxation of the above model is zero-identical to the unit cube. So, the strength of an inequality is simply the right hand side divided by the sum of the positive coefficients in the left hand side of the inequality. Therefore, the strength of the class of the triangle inequalities is $2/3$. In other words, optimizing over the relaxation induced by the square inequalities will never yield an objective value lower than twothird of the sum of the positive coefficients of the objective function.

For the cell formation polytope it is far from trivial to construct classes of inequalities for which the strength is arbitrarily close to zero. But it can be done, as is shown in the last section of Chapter 8.

4.3.3 Comparing the Examples

The cell formation problem can be formulated and solved as a clique partition problem, by adding the missing edges and giving them weight zero. That might be a good idea, since the strength indicators suggest that the

class of triangle inequalities is stronger than the class of square inequalities. However, this idea is of no help. If $V = U \cup W$, then $P(B)$ is a projection of $P(G)$. To see this, we have to convince ourselves of two facts. Firstly, the fact that a feasible solution for the clique partition problem will always induce a feasible solution for the cell formation problem. That follows from the observation that each square inequality is the sum of two triangle inequalities. The other way around is a little harder: from each feasible solution for the cell formation problem we can construct a feasible equivalent solution for the clique partition problem. Assume $h, i \in U$, and we are looking for a feasible value of x_{hi} . To satisfy all triangle inequalities, we must have:

$$\max_{j \in W} \{x_{hj} + x_{ij}\} - 1 \leq x_{hi} \leq 1 - \max_{k \in W} |x_{hk} - x_{ik}|.$$

Such a value for x_{hi} must exist, since otherwise for some $j, k \in W$:

$$x_{hj} + x_{ij} + x_{hk} - x_{ik} > 2.$$

Repeating this argument appropriately, it can be shown that for all missing variables x_{ij} we can find values such that all triangle inequalities are satisfied.

From the projective relation it follows that there is a huge family of objective functions, for which optimizing over the relaxation induced by the triangle inequalities will never yield an objective value lower than twothird of the sum of the positive coefficients of the objective function, in spite of the strength indicator being $1/2$. This illustrates the relative value of the indicator.

Chapter 5

The Cycle Polytope of a Directed Graph

This chapter is joint work with Egon Balas.

An article based on this chapter has been submitted for publication.

The cycle polytope of a directed graph G is the convex hull of the incidence vectors of simple directed cycles of G . In this chapter we show how to obtain facets of the cycle polytope from facets of the asymmetric traveling salesman polytope. It involves sequential lifting, as discussed in Chapter 3, and facet preserving projecting, see Chapter 2. We also obtain facets that have a different origin.

5.1 Introduction

Given a directed graph on n nodes $G := (N, A)$ and the collection $\mathcal{C}$ of its simple directed cycles, the *cycle polytope* $P_{\mathcal{C}}$ of G is the convex hull of incidence vectors of simple directed cycles of G ; i.e.,

$$P_{\mathcal{C}} := \text{conv} \left\{ x \in \{0, 1\}^A : x_{ij} = \begin{cases} 1 & (i, j) \in C \\ 0 & (i, j) \in A \setminus C \end{cases}, \text{ for some } C \in \mathcal{C} \right\}.$$

Notice that the convex hull of *unions* of simple disjoint dicycles is the assignment polytope, and that optimization of a linear function over the assignment polytope is a task whose complexity is $O(n^3)$. Notice also that minimizing a linear function over the cycle polytope $P_{\mathcal{C}}$, i.e. finding the shortest cycle in a directed graph with arbitrary arc lengths is an NP-complete task, as it subsumes the asymmetric traveling salesman (ATS) problem as a special case; but it can be solved in polynomial time if the cost coefficients are restricted in sign so as to exclude the occurrence of negative-length dicycles. Indeed, for cost functions satisfying the above condition, finding a shortest path from j to i in $G - (i, j)$ and adding to it the length of (i, j) yields the shortest cycle containing (i, j) . Repeating this for every arc in the role of (i, j) and then choosing the shortest among the cycles found solves the problem.

The symmetric counterpart of $P_{\mathcal{C}}$, i.e. the cycle polytope of an undirected graph, has been studied in [Coullard and Pulleyblank, 1989], and quite recently in [Bauer, 1993] (see also [Seymour, 1979]).

The cycle polytope of a digraph is closely related to the Prize Collecting Traveling Salesman (PCTS) polytope, studied in [Balas, 1989a, Balas, 1993] (see also [Fischetti and Toth, 1988]). The PCTS problem goes like this. A salesman traveling between cities i and j at a cost of c_{ij} , gets a prize w_i for every city $i \in N$ that he visits. His goal is to find a minimum-cost tour that enables him to collect at least w_0 units of prize money. It has been formulated on a directed graph with loops, $G_L := (N, A \cup L)$, where

$L := \{(1,1), \dots, (n,n)\}$ is the set of loops ($n = |N|$), as the problem of finding

$$\min \sum_{i \in N} \sum_{j \in N \setminus \{i\}} c_{ij} x_{ij} \quad (5.1)$$

subject to

$$\begin{aligned} x(i, N) + y_i &= 1 & i \in N \\ x(N, j) + y_j &= 1 & j \in N \end{aligned} \quad (5.2)$$

$$\sum_{i \in N} w_i y_i \leq U \quad (5.3)$$

$$\begin{aligned} x(S, S) + \sum_{i \in S \setminus \{k\}} y_i - y_\ell &\leq |S| - 1 \\ &\text{for all } S \subset N, 2 \leq |S| \leq n - 2 \\ &\text{and } k \in S, \ell \in N \setminus S \end{aligned} \quad (5.4)$$

$$x_{ij} \in \{0, 1\}, y_i \in \{0, 1\}, \quad i, j \in N \quad (5.5)$$

where x and y are the incidence vectors of arcs and loops, respectively, $U := \sum_{i \in N} w_i - w_0$, and where for $S, T \subseteq N$ we define $x(S, T) := \sum_{i \in S} \sum_{j \in T \setminus \{i\}} x_{ij}$, and we denote $x(\{i\}, S) = x(i, S)$, $x(S, \{i\}) = x(S, i)$.

Every (x, y) satisfying (5.2)-(5.5) defines a simple cycle of length ≥ 2 and a collection of loops, one for every node not in the cycle, with the number of loops (and hence the length of the cycle) constrained by (5.3). The PCTS polytope is then

$$P^* := \text{conv} \{ (x, y) \in \{0, 1\}^{A \cup L} : (x, y) \text{ satisfies (5.2), (5.3), (5.4), (5.5)} \}.$$

In studying the properties of P^* , it is useful to focus on the closely related polytope P_0 obtained from P^* by deleting inequality (5.3):

$$P_0 := \text{conv} \left\{ (x, y) \in \{0, 1\}^{A \cup L} \mid \begin{array}{l} (x, y) \text{ satisfies (5.2), (5.4), (5.5),} \\ x(N, N) \geq 2 \end{array} \right\}.$$

Note that P_0 is the special case of P^* when U equals $n - 2$ and the weights w_i all have value 1. We will call P_0 the cycle-and-loops-polytope of G_L .

For the case when the digraph G is complete, [Balas, 1993] gives a general procedure for obtaining facet defining inequalities for P_0 and P^* from facet

defining inequalities for the asymmetric traveling salesman (ATS) polytope P , i.e. the convex hull of incidence vectors of tours (Hamiltonian dicycles) in G .

This chapter is organized as follows. Section 5.2 discusses some basic properties of P_C . In Section 5.3 we show that the facets of P_0 are in 1-1 correspondence with those of P_C , and that the latter can be obtained from the former by projecting out the loop variables. We thus obtain, roughly speaking, as good a description of the cycle polytope of a directed graph, as the one we have for the ATS polytope defined on the same graph. In the same section we give an overview of all primitive facets derived from the ATS polytope.

In Section 5.4 we discuss additional methods for generating facets of the cycle polytope. For the ATS polytope [Balas and Fischetti, 1993] have given a cloning procedure. We show what this procedure means for the cycle polytope. We also give a (direct) cloning procedure for the cycle polytope. Cloning is a special form of lifting inequalities. Unlike the ATS polytope, the cycle polytope on a complete graph on n nodes is a face of the cycle polytope on a complete graph on $n + 1$ nodes. Therefore, lifting a facet defining inequality is much easier. In Section 5.5 we give an explicit lifting procedure, different from the cloning procedures mentioned earlier.

Finally, in Section 5.6 we study complete descriptions of the cycle polytope for $n = 3$, $n = 4$, and $n = 5$. For $n = 3$ one can derive the complete description from the complete description of the ATSP polytope. Alas, that doesn't hold for the other two cases. We can find some extra classes by lifting the inequalities of a lower dimensional cycle polytope. In itself, this is in none of the cases sufficient to derive a complete description; but in combination with the description derived from the ATSP polytope, it is powerful: for $n = 5$ we derive a complete description. Remarkably enough, this doesn't hold for $n = 4$: we cannot 'find' the trivial nonnegativity inequalities. These inequalities define facets of the cycle polytope on four cities, but define only faces of both the ATSP polytope on four cities and the cycle polytope on three cities.

5.2 The Cycle Polytope

Theorem 5.2.1 *The cycle polytope P_C of G is the convex hull of points $x \in \{0, 1\}^A$ satisfying the system*

$$x(i, N) \leq 1 \quad i \in N \quad (5.6)$$

$$x(N, i) - x(i, N) = 0 \quad i \in N \quad (5.7)$$

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{ij} \geq 1 \quad (5.8)$$

$$x(k, S) + x(\ell, N) - x(S \setminus \{k\}, N \setminus S) \leq 1 \quad (5.9)$$

for all $S \subset N$,
 $2 \leq |S| \leq \lfloor (n-2)/2 \rfloor$,
 $k \in S, \ell \in N \setminus S$

Proof: Constraints (5.6) and (5.7) are satisfied by the incidence vectors of all unions of simple dicycles and the zero vector. Constraint (5.8) excludes the zero vector by requiring the solution-subgraph to contain at least one arc (i, j) such that $i < j$. Clearly, any dicycle and therefore any union of dicycles, has this property. Conversely, any $x \in \{0, 1\}^A$ that satisfies (5.6), (5.7), (5.8) defines a union of simple dicycles in G .

Constraints (5.9) are satisfied by all simple dicycles, but violated by unions of dicycles with more than one member. Indeed, let C be a simple dicycle and x its incidence vector. If $C \subseteq (S, S)$, then $x(\ell, N) = 0$; and if $C \subseteq (N \setminus S, N \setminus S)$, then $x(k, S) = 0$. In either case (5.9) is satisfied. So let $C \cap (S, S) \neq \emptyset$, $C \cap (N \setminus S, N \setminus S) \neq \emptyset$. If $x(k, S) = 0$, (5.9) is clearly satisfied; otherwise k is the tail of an arc in $C \cap (S, S)$, which implies that $x(S \setminus \{k\}, N \setminus S) = 1$, since the cycle C leaves S through some node other than k . In either case, (5.9) is satisfied. Conversely, suppose x is the incidence vector of a union of several cycles. Let C_1 and C_2 be among these. Now choose S such that $C_1 \subseteq (S, S)$ and $C_2 \subseteq (N \setminus S, N \setminus S)$. Then there exists $k \in S$ and $\ell \in N \setminus S$ such that $x(k, S) = 1$, $x(\ell, N) = 1$ and $x(S \setminus \{k\}, N \setminus S) = 0$, i.e. x violates (5.9). $\square$

From now on we assume that the digraph G is complete. We first establish the dimension of P_C . Recall that the dimension of the ATS polytope P is $n^2 - 3n + 1 = (n-1)^2 - n$. The dimension of the cycle-and-loops-polytope P_0 was shown in [Balas, 1989a] to be $(n-1)^2$.

Proposition 5.2.1 *The dimension of P_C is $(n-1)^2$.*

Proof: The system defining P_C has $n(n-1)$ variables and an equality subsystem of rank $\leq n-1$ (the coefficient matrix of (5.7) is the node-arc incidence matrix of G). Hence $\dim(P_C) \leq n(n-1) - n + 1 = (n-1)^2$.

Next we show that any equation $\pi x = \pi_0$ satisfied by all $x \in P_C$ is a linear combination of the equations (5.7). First, for any triplet $\{h, i, j\} \subseteq N$ of

distinct indices, since the incidence vectors of all 2-cycles and 3-cycles defined on the node set $\{h, i, j\}$ belong to P_C , we have

$$\pi_{hi} + \pi_{ih} = \pi_{hj} + \pi_{jh} = \pi_{ij} + \pi_{ji} = \pi_0$$

and

$$\pi_{hi} + \pi_{ij} + \pi_{jh} = \pi_{hj} + \pi_{ji} + \pi_{ih} = \pi_0$$

which implies $\pi_0 = 0$.

Next, in view of (5.7) we may assume w.l.o.g. that $\pi_{j1} = 0$ for all $j \in \{2, \dots, n\}$. For indeed, should this not be the case, an equivalent inequality with this property could be obtained by adding to $\pi x = \pi_0$ the j -th equation (5.7) multiplied by π_{j1} for $j = 2, \dots, n$. From $\pi_{ij} + \pi_{ji} = 0$ for all $i, j \in N$, $\pi_{j1} = 0$ implies $\pi_{1j} = 0$ for all $j \in \{2, \dots, n\}$.

Now we show that $\pi_{ij} = 0$ for all remaining pairs $(i, j) \in N \times N$. For each such (i, j) , consider any 3-cycle on $\{1, i, j\}$. Since its incidence vector is in P_C , we have $\pi_{ij} + \pi_{j1} + \pi_{1i} = 0$; which in light of $\pi_{j1} = 0 = \pi_{1i}$ implies $\pi_{ij} = 0$. Since $\pi_0 = 0$ and $\pi_{ij} = 0$ for all $(i, j) \in N \times N$, the equation $\pi x = \pi_0$ is a linear combination of the equations (5.7). $\square$

Notice that the polytope P_C is the projection of P_0 onto the subspace of the arc variables; in other words, P_C can be obtained from P_0 by projecting out the loop variables y_i , $i \in N$. Yet, the dimension of P_C is the same as that of P_0 . This property holds the key to the one to one correspondence between the facets of P_0 and those of P_C to be discussed below.

Theorem 5.2.2 *The inequalities $x_{ij} \geq 0$ define facets of P_C for $n \geq 4$.*

Proof: Clearly, $x_{ij} \geq 0$ defines a face F of P_C . Let $ax \leq a_0$ define a facet F_A of P_C , containing F . We will show that $ax \leq a_0$ is equivalent to $-x_{ij} \leq 0$ up to multiplication with a positive scalar.

First, choose a triple of nodes h, k, l such that $i \notin \{h, k, l\}$ (this is possible since $n \geq 4$). Any cycle on two or three nodes of $\{h, k, l\}$ induces a point in P_C that satisfies $x_{ij} = 0$, so we deduce like in the proof of Proposition 5.2.1 that $a_0 = 0$. As in that proof, in view of (5.7) we may assume w.l.o.g. that $a_{hk} = 0$ for all $h \in N \setminus \{k\}$ for some $k \notin \{i, j\}$. Since $k \notin \{i, j\}$, all 2-cycles involving k satisfy $x_{ij} \geq 0$ at equality, and hence $a_{kh} = 0$ for all $h \in N \setminus \{k\}$. Now we will show that $a_{hl} = 0$ for all $(h, l) \neq (i, j)$. For each $(h, l) \neq (i, j)$ any 3-cycle on the node set $\{h, l, k\}$ induces a point in P_C satisfying $x_{ij} = 0$, so $a_{hl} + a_{lk} + a_{kh} = 0$. Since $a_{lk} = 0 = a_{kh}$, we have $a_{hl} = 0$.

Finally, since the cycle on i and j satisfies $x_{ij} > 0$, it does not belong to F and hence its incidence vector satisfies $a_{ij}x_{ij} + a_{ji}x_{ji} < 0$; and since $a_{ji} = 0$, we have $a_{ij} < 0$, which concludes our proof. $\square$

In the following we assume that $n \geq 5$.

Clearly, the inequalities $x_{ij} \leq 1$ are dominated by (5.6), so they cannot be facet defining.

Theorem 5.2.3 *The inequalities (5.6) define facets of P_C .*

Proof: Clearly, $x(i, N) \leq 1$ defines a face F of P_C . Let $ax \leq a_0$ define a facet F_A of P_C , containing F . We will show that $ax \leq a_0$ is equivalent to $x(i, N) \leq 1$ up to multiplication with a positive scalar.

As in the proof of Proposition 5.2.1, in view of (5.7) we may assume w.l.o.g. that for any fixed $i \in N$, $a_{ji} = 0$ for all $j \in N \setminus \{i\}$. Since any 2-cycle involving node i satisfies $x(i, N) = 1$, we have $a_{ij} = a_0$ for all $j \in N \setminus \{i\}$.

Now we will show that $a_{jk} = 0$ for all j, k such that $i \notin \{j, k\}$. For each j, k such that $i \notin \{j, k\}$ we construct a 3-cycle on $\{i, j, k\}$. That induces a point $x \in P_C$ satisfying $x(i, N) = 1$, so $a_{jk} + a_{ki} + a_{ij} = a_0$. Since $a_{ki} = 0$ and $a_{ij} = a_0$, we have $a_{jk} = 0$.

Finally, since the 2-cycle on any h and j such that $i \notin \{h, j\}$ satisfies $x(i, N) < 1$, we have $a_0 > a_{hj} + a_{jh} = 0$. $\square$

Next we turn to the inequality (5.8). Unlike the constraints $x_{ij} \geq 0$ and (5.6), (5.7), (5.9), the inequality (5.8) seems to depend on the numbering of the nodes. Symmetry considerations suggest that this cannot be the case; and indeed, as the next Theorem shows, the inequalities (5.8) obtained for all possible numberings of the nodes are valid and have the same properties.

Theorem 5.2.4 *For any permutation π of the node set N , the inequality*

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{\pi(i), \pi(j)} \geq 1 \quad (5.10)$$

defines a facet of P_C .

Proof: Assume w.l.o.g. that $\{\pi(1), \dots, \pi(n)\} = \{1, \dots, n\}$, i.e. that (5.10) is (5.8). It is easy to see that every 2-cycle satisfies (5.8) at equality; therefore (5.8) defines a nonempty face F of P_C . Let $F_\star$ be a facet of P_C containing F and let $bx \leq b_0$ be an inequality defining $F_\star$. In view of (5.7), we may assume w.l.o.g. that $b_{i1} = 0$ for all $i \in \{2, \dots, n\}$.

First, consider the coefficients b_{1i} for $i \in \{2, \dots, n\}$. Since the incidence vector of the 2-cycle on $\{1, i\}$ is in F_* , we have $b_{i1} + b_{1i} = b_{1i} = b_0$ for all $i \in \{2, \dots, n\}$.

Consider now the coefficients b_{ij} for $i \in \{3, \dots, n\}$, $j \in \{2, \dots, i-1\}$, corresponding to the arcs for which the index of the tail is greater than the index of the head. For each such pair, consider the 3-cycle visiting the nodes 1, i and j in that sequence. Since the incidence vector of the 3-cycle is in F_* , we have $b_{1i} + b_{ij} + b_{j1} = b_0 = b_{1i} + b_{ij}$, implying $b_{ij} = 0$ for all $i \in \{3, \dots, n\}$, $j \in \{2, \dots, i-1\}$.

Next consider the coefficients b_{ij} for $i \in \{2, \dots, n-1\}$, $j \in \{i+1, \dots, n\}$. Since the 2-cycle on $\{i, j\}$ induces a point in F_* , we have $b_{ij} + b_{ji} = b_0$, implying $b_{ij} = b_0$ for all $i \in \{2, \dots, n-1\}$, $j \in \{i+1, \dots, n\}$.

We have shown that the inequality $bx \leq b_0$ defining F_* is equivalent to (5.8) up to multiplication with a positive scalar. This proves that (5.8), and hence (5.10), is facet defining. $\square$

Note the huge size of the class (5.10): it contains $n!$ facets of P_C , one for each linear ordering of N .

As to the inequalities (5.9), they are the projections of the corresponding inequalities (5.4) for P_0 onto the subspace of the x variables, i.e. each inequality (5.9) associated with a given triplet $\{S, k \in S, \ell \in N \setminus S\}$ is obtained from the inequality (5.4) associated with the same triplet by substituting $1 - x(i, N)$ for y_i , $i \in (S \setminus \{k\}) \cup \{\ell\}$. The total number of such inequalities is $O(n^2 2^n)$. Now it is well known that when a polyhedron Π is projected onto a subspace, facets of Π do not necessarily project into facets of the subspace-polyhedron. It is therefore noteworthy that in the particular case of P_0 and P_C , they do, as the main theorem in the next section shows.

Notice also that the inequalities (5.9) can be restated in a way that evidences a certain symmetry, hidden in the formulation (5.9):

$$\begin{aligned} x(k, N) + x(\ell, N) - x(S, N \setminus S) &\leq 1 & (5.9') \\ &\text{for all } S \subset N, \\ &2 \leq |S| \leq \lfloor (n-2)/2 \rfloor, \\ &\text{for all } k \in S, \ell \in N \setminus S \end{aligned}$$

The inequality (5.9') can be obtained by adding the inequalities (5.6) for $i = k, \ell$ to the negative of the cutset inequality $x(S, N \setminus S) \geq 1$, valid for the ATS polytope but not for the cycle polytope.

5.3 The Main Result and Some of Its Consequences

As already mentioned, the cycle polytope P_C is the projection of the cycle-and-loops polytope P_0 onto the subspace of the arc variables; and facets of P_0 can be obtained from facets of the ATS polytope by some specialized lifting procedures [Balas, 1993]. Next we show that, unlike in the case of a general polyhedron projected onto an arbitrary subspace, in this case the facets of P_0 project into facets of P_C (see Chapter 2).

Theorem 5.3.1 *Let $\alpha x + \beta y \leq \alpha_0$ be a facet defining inequality for the cycle-and-loops-polytope P_0 . Then the inequality*

$$\sum_{i \in N} \sum_{j \in N \setminus \{i\}} (\alpha_{ij} - \beta_i) x_{ij} \leq \alpha_0 - \sum_{i \in N} \beta_i, \quad (5.11)$$

which is the projection of $\alpha x + \beta y \leq \alpha_0$ onto the subspace of x , defines a facet of P_C .

Proof: Clearly, (5.11) is the projection of $\alpha x + \beta y \leq \alpha_0$ onto the subspace of x , since it can be obtained from the latter by substituting $1 - x(i, N)$ for each y_i , $i \in N$. We show that (5.11) defines a facet of P_C by exhibiting $\dim(P_C) = (n-1)^2$ affinely independent points $x \in \{0, 1\}^A \cap P_C$ that satisfy (5.11) with equality.

Since $\alpha x + \beta y \leq \alpha_0$ defines a facet of P_0 , there exists a set of $(n-1)^2$ affinely independent points $(x^k, y^k) \in \{0, 1\}^{A \cup L} \cap P_0$ that satisfy $\alpha x^k + \beta y^k = \alpha_0$, and $x^k \neq 0$, $k = 1, \dots, q = (n-1)^2$. We will show that if a set of points $(x^k, y^k) \in \{0, 1\}^{A \cup L} \cap P_0$, $k = 1, \dots, q$, is affinely independent, then so is the set of points x^k , $k = 1, \dots, q$. For suppose not, i.e. suppose the points x^k , $k = 1, \dots, q$ are affinely dependent. Then there exist scalars λ_k , $k = 1, \dots, q$, not all zero, such that $\sum_{k=1}^q \lambda_k = 0$ and $\sum_{k=1}^q x_{ij}^k \lambda_k = 0$ for all $(i, j) \in A$. We claim that this implies $\sum_{k=1}^q y_i^k \lambda_k = 0$ for all $i \in N$, which in turn means that the points (x^k, y^k) are also affinely dependent, a contradiction. To prove the claim, we notice that from the assumptions,

$$\sum_{k=1}^q \left(\sum_{j \in N \setminus \{i\}} x_{ij}^k \right) \lambda_k = 0, \quad i \in N;$$

hence for $i \in N$,

$$\begin{aligned} \sum_{k=1}^q y_i^k \lambda_k &= \sum_{k=1}^q (1 - \sum_{j \in N \setminus \{i\}} x_{ij}^k) \lambda_k \\ &= \sum_{k=1}^q \lambda_k - \sum_{k=1}^q (\sum_{j \in N \setminus \{i\}} x_{ij}^k) \lambda_k = 0. \end{aligned}$$

Thus the points $x^k \in \{0, 1\}^A \cap P_C$, $k = 1, \dots, q$, are affinely independent. Further, they clearly satisfy (5.11) with equality, since the corresponding points (x^k, y^k) satisfy $\alpha x^k + \beta y^k = \alpha_0$, $k = 1, \dots, q$. $\square$

Corollary 5.3.1 *The inequalities (5.9) define facets of P_C .*

Proof: The inequalities (5.9) are projections of the inequalities (5.4), which define facets of P_0 . Therefore the claim follows from Theorem 5.3.1. $\square$

We now derive, based on the above theorem, the facet defining inequalities for P_C corresponding to some of the well known facet defining inequalities for the ATS polytope P . We do this in two steps, as done in [Balas, 1993] for the cycle-and-loops-polytope P_0 . First we consider primitive inequalities, then we address the cloning procedure which yields nonprimitive inequalities.

Two nodes h and k of G are called *clones* with respect to a valid inequality $\alpha x \leq \alpha_0$ for the ATS polytope P , if

- (a) $\alpha_{ih} = \alpha_{ik}$ and $\alpha_{hi} = \alpha_{ki}$ for all $i \in N \setminus \{h, k\}$
- (b) $\alpha_{hk} = \alpha_{kh} = \max \{ \alpha_{ik} + \alpha_{kj} - \alpha_{ij} : i, j \in V \setminus \{h, k\}, i \neq j \}$
- (c) the restriction $\bar{\alpha} x \leq \bar{\alpha}_0$ of $\alpha x \leq \alpha_0$ to $\bar{G} := G \setminus \{n\}$, with $\bar{\alpha}_0 := \alpha_0 - \alpha_{hk}$, is valid for the ATS polytope defined on $\bar{G}$.

A valid inequality $\alpha x \leq \alpha_0$ for P is called *primitive* if G has no clones with respect to $\alpha x \leq \alpha_0$.

These concepts were introduced by [Balas and Fischetti, 1993a], who showed that if the primitive members of a family $\mathcal{F}$ of inequalities define facets of P , then all the inequalities in $\mathcal{F}$ define facets of P .

We first consider the family called *comb* inequalities. Let the *handle* H and the *teeth* T_j , $j = 1, \dots, s$, be node sets satisfying

$$(c.1) \quad |T_j \cap H| = |T_j \setminus H| = 1, j = 1, \dots, s, s \geq 3 \text{ and odd}$$

$$(c.2) \quad T_i \cap T_j = \emptyset, \quad i, j = 1, \dots, s$$

$$(c.3) \quad |H \setminus (\cup_{j=1}^s T_j)| \leq 1$$

$$(c.4) \quad |N \setminus (H \cup (\cup_{j=1}^s T_j))| = 1.$$

Then

$$x(H, H) + \sum_{j=1}^s x(T_j, T_j) \leq |H| + (s-1)/2 \quad (5.12)$$

is a primitive comb inequality. The family of comb inequalities is known to be facet defining for the symmetric (see [Grötschel and Padberg, 1985]) as well as the asymmetric (see [Fischetti, 1991]) traveling salesman polytope. Here is the corresponding result for the cycle polytope P_C .

Corollary 5.3.2 *Let $H, T_j \subseteq N$, $j = 1, \dots, s$, satisfy conditions (c.1), (c.2), (c.3) and (c.4). Then the inequality*

$$\sum_{j=1}^s x(T_j, T_j) - x(H, N \setminus H) \leq (s-1)/2$$

defines a facet of P_C .

Proof: The inequality for P_0 corresponding to (5.11) is (see [Balas, 1993])

$$x(H, H) + \sum_{j=1}^s x(T_j, T_j) + \sum_{i \in H} y_i \leq |H| + \frac{s-1}{2}.$$

Substituting for y_i , $i \in H$, yields (5.11). $\square$

Next we address the class of odd CAT inequalities for P introduced in [Balas, 1989]. Two arcs $(i, j), (k, \ell)$ of G are called *incompatible* if $i = k$, or $j = \ell$, or $i = \ell$ and $j = k$. A *closed alternating trail* (CAT) in G is a sequence of distinct arcs $T := (a_1, \dots, a_t)$ such that for $i, j \in \{1, \dots, t\}$, a_i and a_j are incompatible if and only if $|i - j| = 1$ or $t - 1$. An odd CAT is a CAT of odd length (number of arcs). Let $N(T)$ be the set of endpoints of the arcs in T . A node $i \in N(T)$ is a *source* if $\deg^+(i) \geq 2$, a *destination* (sink) if $\deg^-(i) \geq 2$, and *neutral* if $\deg^+(i) \leq 1$ and $\deg^-(i) \leq 1$. A node can be a source and a destination at the same time. Every odd CAT has at least one 2-cycle with a neutral node. A *chord* of T is an arc in $A \setminus T$ joining two nodes of $N(T)$. A chord of type 1 is one that joins a source i to a sink $j \neq i$.

Let T be an odd CAT and K the set of its chords of type 1. Then apart from two small pathological cases (with $5 \leq n \leq 6$), the odd CAT inequality

$$x(T \cup K) \leq (|T| - 1)/2 \quad (5.13)$$

defines a facet of the ATS polytope P (see [Balas, 1989]).

The odd CAT inequality (5.13) is easily seen to be primitive if and only if every 2-cycle of T has a neutral node. In such a case we call T itself primitive.

Corollary 5.3.3 *Let T be a primitive odd CAT different from the two known pathological cases, K the set of its chords, and let $S \subset N(T)$ and $D \subset N(T)$ denote the set of sources and destinations of T , respectively.*

Then the inequality

$$x(T \cup K) - x(S \cap D, N) \leq (|T| - 1)/2 - |S \cap D| \quad (5.14)$$

defines a facet of P_C .

Proof: The inequality for P_0 corresponding to (5.12) is (see [Balas, 1993])

$$x(T \cup K) + \sum_{i \in S \cap D} y_i \leq (|T| - 1)/2,$$

and substituting for y_i , $i \in S \cap D$, yields (5.13). $\square$

The next family that we examine is that of the SD (for source-destination) inequalities, which generalizes the comb, odd CAT and several other families of inequalities ([Balas and Fischetti, 1993a]). The primitive members of the family are defined as follows. Let (S, D, W, I, E, Q) be a partition of N with the following properties:

S is a (possibly empty) set of sources,

D is the (possibly empty) set of destinations,

$H := W \cup I$ is the (nonempty) handle, with $0 \leq |W| \leq 1$ and $|I| = s \geq 1$,

$I := \{i_1, \dots, i_s\}$,

$E := \{e_1, \dots, e_s\}$, and

$T_j := \{i_j, e_j\}$, $j = 1, \dots, s$ are the teeth

and $0 \leq |Q| \leq 1$, with $|S| + |D| + s$ odd. Then the primitive SD inequality

$$x(S \cup H, D \cup H) + \sum_{j=1}^s x(T_j, T_j) \leq \frac{|S| + |D| + 2|H| + s - 1}{2} \quad (5.15)$$

is valid for P and, with the exception of three small pathological cases (with $4 \leq n \leq 6$ and either $|S| = |D| = 0$ or $|S| = |D| = s = 1$), defines a facet of P if $||S| - |D|| \leq \max \{0, s - 3\}$ (see [Balas and Fischetti, 1993a]). The corresponding result for the cycle polytope is then:

Corollary 5.3.4 *Let (S, D, W, I, E, Q) be as above, and let $|S| + |D| + s \geq 3$, with strict inequality if $n \leq 6$. Then the inequality*

$$x(S, H \cup D) + \sum_{j=1}^s x(T_j, T_j) - x(H, N \setminus (H \cup D)) \leq \frac{|S| + |D| + s - 1}{2} \quad (5.16)$$

defines a facet of P_C .

Proof: The corresponding inequality for P_0 , shown to be facet defining in [Balas, 1993], is

$$x(S \cup H, S \cup D) + \sum_{j=1}^s x(T_j, T_j) + \sum_{i \in H} y_i \leq \frac{|S| + |D| + 2|H| + s - 1}{2}$$

Substituting for y_i , $i \in H$, yields (5.16). $\square$

Next we consider a class of lifted cycle inequalities, namely the D_k^+ and D_k^- inequalities (see [Grötschel and Padberg, 1985]). Although these inequalities are strictly speaking not primitive, their only clones are isolated nodes. Let k be any integer satisfying $3 \leq k \leq n - 2$; then the D_k^+ inequality is

$$\sum_{j=1}^{k-1} x_{i_j i_{j+1}} + x_{i_k i_1} + 2x(\{i_1\}, \{i_3, \dots, i_k\}) + \sum_{j=4}^k x(\{i_j\}, \{i_3, \dots, i_{j-1}\}) \leq k - 1. \quad (5.17)$$

In [Balas and Fischetti, 1993] these inequalities were shown to be facet defining for P . The corresponding inequality for P_C is as follows.

Corollary 5.3.5 *For any integer k , $3 \leq k \leq n - 2$, the inequality*

$$\begin{aligned} & x(\{i_1\}, \{i_3, \dots, i_k\}) + x_{i_2 i_3} + x(\{i_\ell\}, N) \\ & \quad - x(\{i_1\}, N \setminus \{i_1, \dots, i_k\}) \\ & \quad - \sum_{j=3}^{k-1} x(\{i_j\}, N \setminus \{i_3, \dots, i_{j+1}\}) \\ & \quad - x(\{i_k\}, N \setminus (\{i_1\} \cup \{i_3, \dots, i_k\})) \leq 1 \end{aligned} \quad (5.18)$$

defines a facet of P_C .

Proof: The inequality for P_0 corresponding to (5.17), shown in [Balas, 1993] to be facet defining, is

$$\sum_{j=1}^{k-1} x_{i_j i_{j+1}} + x_{i_k i_1} + 2x(\{i_1\}, \{i_3, \dots, i_k\}) + \sum_{j=4}^k x(\{i_j\}, \{i_3, \dots, i_{j-1}\}) \\ + y_{i_1} + \sum_{j=3}^k y_{i_j} - y_{i_\ell} \leq k - 1$$

for some $i_\ell \in N \setminus \{i_1, \dots, i_k\}$. Substituting for y_{i_1}, y_{i_ℓ} and $y_{i_j}, j = 3, \dots, k$ yields (5.18). $\square$

The D_k^- inequalities are analogous to the D_k^+ inequalities, with all the arcs with coefficient equal to 2 directed into i_1 rather than out of i_1 . A result analogous to Corollary 5.3.5 gives a facet defining inequality for P_C corresponding to the D_k^- inequality for P .

Finally, we briefly mention the E_4 -inequalities shown by [Grötschel, 1977] to be facet defining for P . For any $\{i_1, \dots, i_4\} \subset N$, the corresponding E_4 -inequality is

$$2x_{i_1 i_2} + x_{i_1 i_4} + 2x_{i_2 i_1} + x_{i_2 i_3} + x_{i_4 i_2} \leq 3. \quad (5.19)$$

Corollary 5.3.6 *For any $\{i_1, \dots, i_4\} \subset N$, the inequality*

$$x_{i_1 i_2} + x_{i_2 i_1} + x_{i_4 i_2} + x_{i_3 i_1} - x_{i_1 i_3} - x_{i_2 i_4} \leq 1 \quad (5.20)$$

defines a facet of P_C .

Proof: It is not hard to see that the inequality

$$2x_{i_1 i_2} + x_{i_1 i_4} + 2x_{i_2 i_1} + x_{i_2 i_3} + x_{i_3 i_1} + x_{i_4 i_2} + y_1 + y_2 \leq 3$$

derived from (5.19) by sequential lifting is valid and facet defining for P_0 . Substituting for y_1 and y_2 then yields (5.20), which according to Theorem 5.3.1 must be valid and facet defining for P_C . $\square$

5.4 Cloning

Next we turn to general nonprimitive facet defining inequalities. As mentioned earlier, [Balas and Fischetti, 1993a] gave a procedure whereby a facet

defining inequality $\alpha x \leq \alpha_0$ for the ATS polytope $P(G^n)$ defined on the complete digraph G^n on n nodes can be lifted into a facet defining inequality for the polytope $P(G^{n+1})$ defined on the complete digraph on $n+1$ nodes G^{n+1} . This procedure is called *cloning*, and it is done by making the new node of G^{n+1} , say k , a *clone* of some node h of G^n with respect to $\alpha x \leq \alpha_0$, i.e. by choosing the lifting coefficients α_{ki} and α_{ik} , $i = 1, \dots, n$, so as to satisfy the conditions (a), (b), (c) stated above. Applying cloning repeatedly leads to *clique lifting*, whereby nodes can be expanded into cliques (of clones).

[Balas, 1993] has extended this procedure of cloning/clique lifting to facet defining inequalities for the cycle-and-loops-polytope P_0 . In the procedure of [Balas, 1993], when cloning a node of G^n to lift a facet defining inequality $\alpha x + \beta y \leq \alpha_0$ for $P_0(G_L^n)$ into one for $P_0(G_L^{n+1})$, the lifting coefficients α_{ki} , α_{ik} , $i = 1, \dots, n$ of the arcs incident with the new node k (which is to be made a clone of, say, node h) are determined as above, in conformity with conditions (a), (b), (c); but in addition, a lifting coefficient

$$\alpha_{kk} = \max \{ \alpha_{ik} + \alpha_{kj} - \alpha_{ij} : i, j \in V \setminus \{h, k\} \}$$

is determined for the loop variable y_k of node k . Since this lifting coefficient (for y_k) is not necessarily the same as the coefficient α_{hh} of the loop variable y_h , node k is called a *quasi-clone* (rather than a clone) of h . In all other respects, the clique lifting procedure for P_0 works the same way as the one for P ; above all, it guarantees that facet defining inequalities, when lifted, give rise to facet defining inequalities.

In turning now to the cycle polytope P_C , it follows from the above discussion that in order to find the facet defining inequality for P_C that corresponds to a given (arbitrary) facet defining inequality $\alpha x \leq \alpha_0$ for the ATS polytope $P(G)$, one can proceed as follows:

1. Find the primitive inequality $\tilde{\alpha}x \leq \tilde{\alpha}_0$ for $P(\tilde{G})$, the ATS polytope defined on some induced subgraph $\tilde{G}$ of G , from which $\alpha x \leq \alpha_0$ can be determined by cloning (clique lifting).
2. Derive from $\tilde{\alpha}x \leq \tilde{\alpha}_0$ the corresponding primitive facet inducing inequality $\tilde{\alpha}x + \tilde{\beta}y \leq \tilde{\alpha}_0$ for $P_0(\tilde{G}_L)$ defined on the digraph $\tilde{G}_L$ with loops.
3. Apply the cloning procedure to $\tilde{\alpha}x + \tilde{\beta}y \leq \tilde{\alpha}_0$ to obtain a facet inducing inequality $\alpha x + \beta y \leq \alpha_0$ for $P_0(G_L)$.
4. Substitute for y to obtain the corresponding facet inducing inequality $\hat{\alpha}x \leq \hat{\alpha}_0$ for $P_C(G)$, the cycle polytope of G .

We illustrate the procedure by an example.

Example 5.4.1 Consider the inequality

$$(E.1) \quad x_{12} + x_{13} + x_{15} + x_{32} + x_{34} + x_{43} + x_{45} + x_{52} + x_{54} + 2x_{35} + 2x_{53} \leq 4$$

whose support graph is shown in Figure 5.1(a). This graph has an arc (i, j) drawn in a plain line if $\alpha_{ij} = 1$, an arc (i, j) drawn in a double line if $\alpha_{ij} = 2$, and no arc (i, j) if $\alpha_{ij} = 0$. (E.1) is valid for the ATS polytope P on $n \geq 5$ nodes.

It is not hard to see that nodes 3 and 5 are clones, in that $\alpha_{i3} = \alpha_{i5}$ and $\alpha_{3i} = \alpha_{5i}$ for all i , and $\alpha_{35} = \alpha_{53} = \max \{\alpha_{i3} + \alpha_{3j} - \alpha_{ij} : i, j \in N \setminus \{3, 5\}\} = 2$. The primitive inequality from which (E.1) can be obtained by cloning is the odd CAT inequality with $|T| = 5$, also known as T_2 , whose support graph is shown in Figure 5.1(b):

$$(E.2) \quad x_{12} + x_{13} + x_{32} + x_{34} + x_{43} \leq 2$$

This inequality is known to be facet defining for the ATS polytope P defined on the complete digraph with 4 nodes. The corresponding inequality for P_0 defined on the complete digraph with loops on 4 nodes, illustrated in Figure 5.1(c), is

$$(E.3) \quad x_{12} + x_{13} + x_{32} + x_{34} + x_{43} + y_3 \leq 2.$$

Applying the cloning procedure of [Balas, 1993] to node 3 yields the inequality whose support graph is shown in Figure 5.1(d):

$$(E.4) \quad x_{12} + x_{13} + x_{15} + x_{32} + x_{34} + x_{43} + x_{45} + x_{52} + x_{54} + 2x_{35} + 2x_{53} + y_3 + 2y_5 \leq 4$$

Note that node 3 has loop coefficient of 1, whereas its quasi-clone 5 has loop coefficient of 2. Finally, substituting for y_3 and y_5 we obtain the inequality whose support graph is shown in Figure 5.1(e), with negative coefficients drawn in dotted lines:

$$(E.5) \quad x_{12} + x_{13} + x_{15} + x_{35} + x_{43} + x_{45} - x_{31} - x_{52} - x_{54} - 2x_{51} \leq 1$$

5.5 Direct Lifting Procedures

In this section we describe a lifting procedure for the cycle polytope, that subsumes the cloning procedure of the last section as a special case. Two

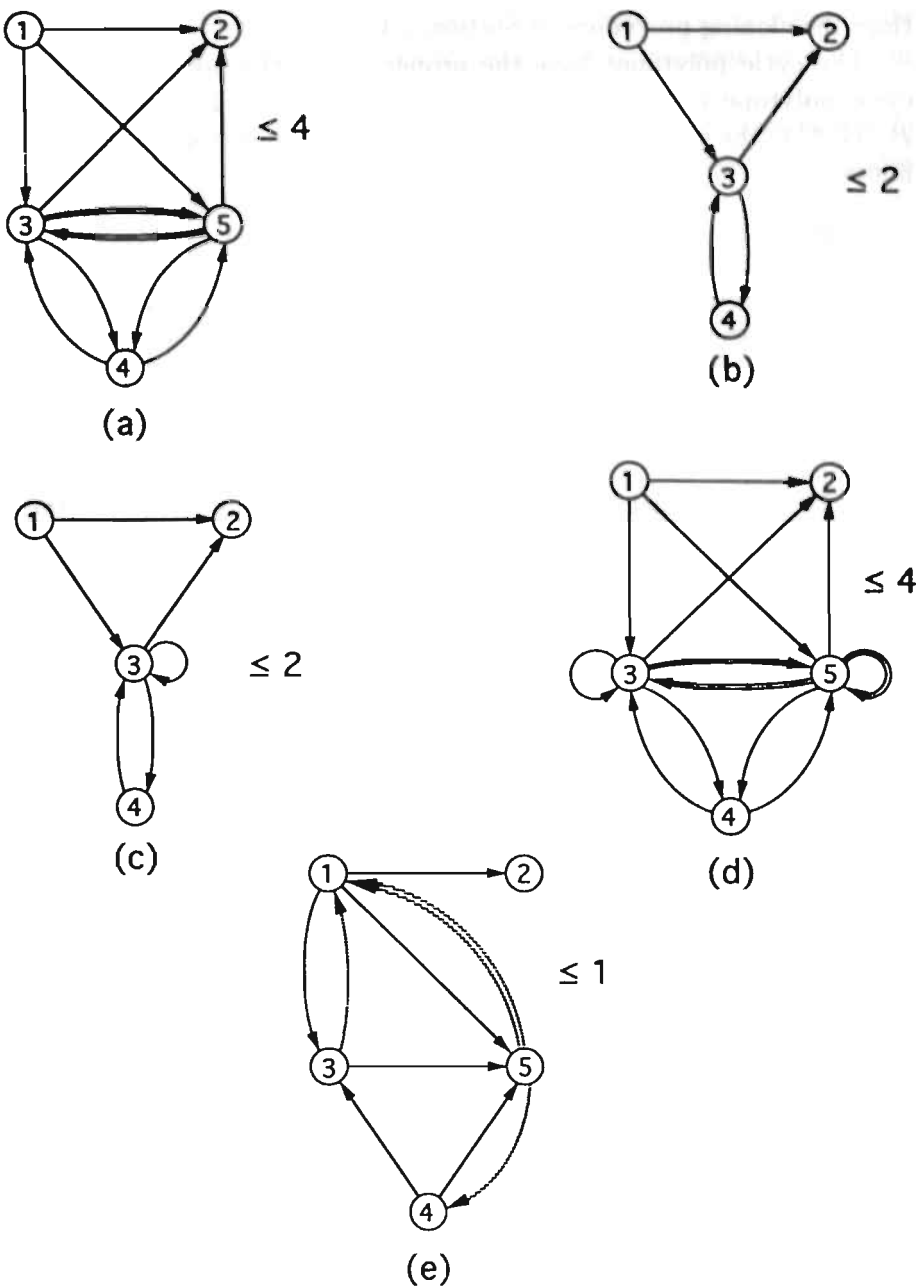


Figure 5.1: Cloning via the ATS polytope

other special cases will be discussed, which may give rise to different facets than the cloning procedure of Section 5.4.

The cycle polytopes have the property that the polytope $P_C(G^n)$, the cycle polytope defined on a complete digraph with n nodes, is a face of $P_C(G^{n+1})$, the cycle polytope defined on a complete digraph with $n + 1$ nodes:

$$P_C(G^n) = \{x \in P_C(G^{n+1}) \mid x_{in+1} = x_{n+1i} = 0 \ \forall i \in \{1, \dots, n\}\}.$$

Let $\alpha x \leq \alpha_0$ be a valid inequality that defines a proper face F_α of $P_C(G^n)$. For our purposes here, a lifting of $\alpha x \leq \alpha_0$ is an inequality

$$\alpha x + \sum_{i=1}^n \beta_i x_{in+1} + \sum_{j=1}^n \gamma_j x_{n+1j} \leq \alpha_0 \quad (5.21)$$

which is valid. Notice that (5.21) defines a face F of $P_C(G^{n+1})$ such that $F_\alpha \subseteq F$. The set of all vectors $(\beta, \gamma) \in \mathbb{R}^{n+n}$ for which (5.21) is valid, will be called the *lifting set* of $\alpha x \leq \alpha_0$, and denoted L_{α, α_0} .

In Chapter 3 we have studied the lifting sets of valid inequalities of general (as opposed to full dimensional) polyhedra, and we have described their basic properties. One of the main results specializes to a useful theorem for our case, which for the sake of completeness we derive here from scratch.

To start with, in the following proposition we introduce a relatively simple explicit representation of L_{α, α_0} for the case that $\alpha x \leq \alpha_0$ defines a nontrivial facet of $P_C(G^n)$.

Proposition 5.5.1 *Let*

$$L = \left\{ \beta, \gamma \in \mathbb{R}^n \mid \begin{array}{ll} \beta_i + \gamma_j \leq \alpha_{ij} & \forall (i, j) \in A \\ \beta_i + \gamma_i \leq \alpha_0 & \forall i \in N \end{array} \right\}.$$

Then we have:

- (i) *If $\alpha x \leq \alpha_0$ is valid for $P_C(G^n)$, then $L \subseteq L_{\alpha, \alpha_0}$.*
- (ii) *If $\alpha x \leq \alpha_0$ is valid for $P_C(G^n)$, and defines a nontrivial facet of $P_C(G^n)$, then $L = L_{\alpha, \alpha_0}$.*

Proof:

(i) Let $(\beta, \gamma) \in L$, and $y \in P_C(G^{n+1})$, y integer. We show that (β, γ) and y satisfy (5.21).

If the cycle $C(y)$ defined by y does not contain node $n+1$, the inequality holds by the validity of $\alpha x \leq \alpha_0$. Therefore, assume $C(y)$ contains node $n+1$. Let h and k be the predecessor, respectively successor, of $n+1$ in $C(y)$.

If $h \neq k$, i.e. $|C(x)| > 2$, then from the fact that removing the arcs $(h, n+1)$ and $(n+1, k)$ from $C(y)$ and adding arc (h, k) yields a cycle in G^n whose incidence vector satisfies $\alpha y \leq \alpha_0$, it follows that

$$\begin{aligned} \alpha y + \sum_{i=1}^n \beta_i y_{in+1} + \sum_{j=1}^n \gamma_j y_{n+1j} &= \alpha y + \beta_h + \gamma_k \leq \alpha_0 \\ (\alpha_0 - \alpha_{hk}) + \beta_h + \gamma_k &= \alpha_0 + (\beta_h + \gamma_k - \alpha_{hk}) \leq \alpha_0. \end{aligned}$$

If $h = k$, i.e. $|C(x)| = 2$, then we have:

$$\alpha y + \sum_{i=1}^n \beta_i y_{in+1} + \sum_{j=1}^n \gamma_j y_{n+1j} = \beta_h + \gamma_k \leq \alpha_0.$$

So, for any $(\beta, \gamma) \in L$, (5.21) is satisfied by all $x \in P_C(G^{n+1})$. therefore, $L \subseteq L_{\alpha, \alpha_0}$.

(ii) Let $(\beta, \gamma) \in L_{\alpha, \alpha_0}$. We show $L_{\alpha, \alpha_0} \subseteq L$. Together with the result of (i) this proves (ii).

From the 2-cycle on nodes i and $n+1$ and $(\beta, \gamma) \in L_{\alpha, \alpha_0}$, it follows that $\beta_i + \gamma_i \leq \alpha_0$, for all $i \in N$, so (β, γ) satisfies the second type of inequalities defining L .

Now we consider $\beta_i + \gamma_j \leq \alpha_{ij}$ for any pair i and j such that $(i, j) \in A$. Because $\alpha x \leq \alpha_0$ defines a facet of $P_C(G^n)$, and $\alpha x \leq \alpha_0$ is not equivalent to a nonnegativity constraint, there exists for every arc (i, j) at least one solution $y \in P_C(G^n)$ such that $y_{ij} = 1$ and $\alpha y = \alpha_0$. From $C(y)$ we can construct a cycle C in G^{n+1} , by replacing the arc (i, j) in $C(y)$ with the pair of arcs $\{(i, n+1), (n+1, j)\}$. Let z be the incidence vector of C . Then, from the validity of z , we have

$$\alpha z + \sum_{h=1}^n \beta_h z_{hn+1} + \sum_{k=1}^n \gamma_k z_{n+1k} = \alpha y - \alpha_{ij} + \beta_i + \gamma_j \leq \alpha_0.$$

Since $\alpha y = \alpha_0$, we derive $\beta_i + \gamma_j \leq \alpha_{ij}$. □

In the remainder of this section, we will assume that $\alpha x \leq \alpha_0$ is valid for $P_C(G^n)$ and defines a nontrivial facet of $P_C(G^n)$, because that is the only case of interest for the purposes of this section. Now we know what the lifting set looks like, we can study its polyhedral properties, some of which are given in the following proposition.

Proposition 5.5.2 *Assume that $\alpha x \leq \alpha_0$ is valid for $P_C(G^n)$ and defines a nontrivial facet of $P_C(G^n)$. Then L_{α, α_0} has the following properties:*

- (i) L_{α, α_0} is full dimensional.
- (ii) If $(\beta, \gamma) \in L_{\alpha, \alpha_0}$, then $(\beta + c\mathbf{1}, \gamma - c\mathbf{1}) \in L_{\alpha, \alpha_0} \quad \forall c \in \mathbb{R}$, where $\mathbf{1}$ denotes an n -dimensional vector with all entries having value 1.
- (iii) The extreme faces of L_{α, α_0} have dimension 1.

Proof:

(i) A polyhedron is of full dimension if and only if it has an interior point, i.e. a vector satisfying all constraints, but satisfying none of the constraints at equality. It is easy to verify that the following vector is an interior point of L_{α, α_0} :

$$\begin{aligned} \beta_i &= 0 & \forall i \in N \\ \gamma_j &= \min\{\min_{i \in N: i \neq j} \{\alpha_{ij}\}, \alpha_0\} - 1 & \forall j \in N \end{aligned}$$

(ii) We show that $(\beta + c\mathbf{1}, \gamma - c\mathbf{1}) \in L_{\alpha, \alpha_0} \quad \forall c \in \mathbb{R}$:

$$\begin{aligned} (\beta_i + c) + (\gamma_i - c) &= \beta_i + \gamma_i \leq \alpha_0 & \forall i \in N \\ (\beta_i + c) + (\gamma_j - c) &= \beta_i + \gamma_j \leq \alpha_{ij} & \forall i, j \in N : (i, j) \in A \end{aligned}$$

(iii) An extreme face of L_{α, α_0} is a nonempty subset of L_{α, α_0} satisfying a maximal affinely independent set of inequalities at equality. The dimension of such a face is the dimension of L_{α, α_0} minus the cardinality of the affinely independent set of inequalities satisfied at equality by all elements of the face. From (i) we know that the dimension of L_{α, α_0} is $2n$. To complete the proof we show that the cardinality of a maximal affinely independent set of inequalities satisfied at equality is $2n - 1$.

The constraint matrix of L_{α, α_0} can be interpreted as the incidence matrix of a connected graph. This graph has a node set consisting of N and a copy $\bar{N}$ of N . There is an arc from every node in N to every node in $\bar{N}$. The rank of an incidence matrix of a connected graph is the number of nodes of the graph minus one, in our case: $2n - 1$. $\square$

The second property is that for every vector $(\beta, \gamma) \in L_{\alpha, \alpha_0}$, we can construct an affinely independent vector $(\bar{\beta}, \bar{\gamma}) \in L_{\alpha, \alpha_0}$. Notice that a solution $x \in P_C(G^{n+1})$ satisfies the lifting induced by $(\bar{\beta}, \bar{\gamma})$ at equality if and only if x satisfies the lifting induced by (β, γ) at equality. In other words, the liftings corresponding to these two vectors define the same face of $P_C(G^{n+1})$!

Now we are ready for the main theorem of this section.

Theorem 5.5.1 *Let $\alpha x \leq \alpha_0$ be valid for $P_C(G^n)$, and define a facet of $P_C(G^n)$.*

Then, a lifting (5.21) of $\alpha x \leq \alpha_0$ defines a facet F of $P_C(G^{n+1})$ if and only if (β, γ) is on a one-dimensional face of L_{α, α_0} .

Proof: $\Rightarrow$) Inequality (5.21) defines a facet F^0 of $P(G^n)$ and a face F of $P(G^{n+1})$, $F^0 \subseteq F$. First we have to show that there is a facet F^* of $P(G^{n+1})$ containing F , such that $F \cap P_C(G^n) = F^0$. The statement is trivial if F is a facet of $P(G^{n+1})$, so assume F is not a facet of $P(G^{n+1})$. Observe that there is a facet $\hat{F}$ of $P_C(G^{n+1})$ containing F . Then either $\hat{F} \cap P_C(G^n) = F^0$, or $\hat{F} \cap P_C(G^n) = P_C(G^n)$. In the latter case, consider a maximal affinely independent set $T \subseteq \hat{F} \cap B^n$. Because $P_C(G^n) \subseteq \hat{F}$, we may assume that there are $\dim(P_C(G^n))$ elements in T belonging to F^0 , and one element, say y , belonging to $P_C(G^n)$ but not to F^0 . Let $T^- := T \setminus \{y\}$. Since $|T \cap (P_C(G^{n+1}) \setminus P_C(G^n))| = |T^- \cap (P_C(G^{n+1}) \setminus P_C(G^n))| = 2n - 2$, there are integer solutions $x \in (P_C(G^{n+1}) \setminus P_C(G^n))$ such that $C(x)$ covers node $n + 1$ and $T^- \cup \{x\}$ is an affinely independent set. Such a set induces a hyperplane. Each of the hyperplanes contains all integer solutions $x \in (P_C(G^{n+1}) \setminus P_C(G^n))$ such that $C(x)$ covers node $n + 1$ and $T^- \cup \{x\}$ is not an affinely independent set, because each of those solutions is the affine combination of solutions satisfying (5.21) at equality. At least one of these hyperplanes supports $P_C(G^{n+1})$, and therefore it contains a facet $\bar{F}$ of $P_C(G^{n+1})$. Now, $\bar{F}$ contains F , and $\bar{F} \cap P_C(G^n) = F^0$.

So, there is a facet F^* of $P(G^{n+1})$ containing F , such that $F \cap P_C(G^n) = F^0$. Therefore, there is a vector $(\lambda, \mu) \in L_{\alpha, \alpha_0}$ inducing a lifting of $\alpha x \leq \alpha_0$ defining F^* . Because F is on a one-dimensional face of L_{α, α_0} , it contains a set of $2n - 1$ integer solutions $x \in P_C(G^{n+1})$ such that $C(x)$ covers node $n + 1$. Because $F \subseteq F^*$, F^* contains the same set. So, (β, γ) , $(\beta + 1, \gamma - 1)$, and (λ, μ) are solutions of a system of $2n - 1$ linearly independent equalities, involving $2n$ variables. There cannot be more than two affinely independent solutions to such a system. Apparently, the set of vectors $\{(\beta, \gamma), (\beta + 1, \gamma - 1), (\lambda, \mu)\}$ is affinely dependent. This implies that (β, γ) and (λ, μ) are on the same one-dimensional face of L_{α, α_0} , defining the same face of $P_C(G^{n+1})$: $F^* = F$.

$\Leftarrow$) Suppose (β, γ) is not on an extreme face of L_{α, α_0} . Then (β, γ) is the linear of a set of affinely independent vectors on extreme faces of L_{α, α_0} , and at least three of the vectors in this set have strictly positive weights. These vectors cannot be on the same one-dimensional face, since that would imply dependency. But if these vectors belong to different faces of L_{α, α_0} , then

they define different faces of $P_C(G^n)$. So, the lifting implied by (β, γ) , is the sum of two inequalities, valid for $P_C(G^{n+1})$ and defining different faces of $P_C(G^{n+1})$, implying that the lifting induced by (β, γ) is not facet defining: a contradiction. $\square$

Preferably, one would like to have disposal of explicit ways of constructing a vector on an extreme face of L_{α, α_0} , ie. deriving facet defining liftings from a nontrivial facet defining inequality. Actually, for some inequalities we already have discussed such a procedure in the former section: cloning. In the remainder of this section we will discuss a general way of generating vectors on different extreme faces of L_{α, α_0} . We will demonstrate that all extreme faces correspond to sequential liftings.

Theorem 5.5.2 *Let $\alpha x \leq \alpha_0$ be valid for $P_C(G^n)$, and define a facet of $P_C(G^n)$. Then, a lifting (5.21) of $\alpha x \leq \alpha_0$ with $\gamma_0 = 0$ defines a facet F of $P_C(G^{n+1})$ if and only if there exists a sequence $\Pi = (\pi_1, \dots, \pi_{2n})$ of β_i and γ_j , $i, j \in \{1, \dots, n\}$, with $\pi_1 = \gamma_0$, $\pi_2 = \beta_i$ for some i , such that:*

$$\begin{aligned}\beta_i &= \min_{j \in \{1 \dots n\}} \{\alpha_{ij} - \gamma_j\} \text{ if } \beta_i = \pi_k, \gamma_j = \pi_l, \text{ and } k > l \\ \gamma_j &= \min_{i \in \{1 \dots n\}} \{\alpha_{ij} - \beta_i\} \text{ if } \beta_i = \pi_k, \gamma_j = \pi_l, \text{ and } k < l\end{aligned}$$

where $\alpha_{ii} = \alpha_0$.

Proof.:

$\Rightarrow$) Since (5.21) is facet defining, (β, γ) is on a face of lowest dimension of L_{α, α_0} . So, there is a set of $2n - 1$ linearly independent constraints satisfied at equality. This set induces a bipartite graph $G = (U, V, F)$ where there is precisely one element u_i in U for every coefficient β_i , where there is precisely one element v_j in V for every coefficient γ_j , and where there is an edge in F for every pair β_i, γ_j occurring in the same constraint of the linearly independent set. Since G is connected and contains $2n$ nodes and $2n - 1$ edges, G is a maximal spanning tree.

Now, we choose $\pi_1 = \gamma_1$, and $\pi_2 = \beta_i$ for some i such that $\beta_i + \gamma_1 = \alpha_{ij}$. Such an i exists, since otherwise G could not be a connected graph. Next, we choose for π_3 a β_h such that u_h neighbors v_1 , or a γ_j such that v_j neighbors u_i , and so on, until all nodes of G are assigned to a sequence number. The reader can verify that this Π , β , and γ satisfy the conditions.

$\Leftarrow$) This follows straightforward from sequential lifting (see Chapter 3). $\square$

As an illustration, we consider the following inequality, that is facet defining for the cycle polytope on four nodes:

$$x_{12} + x_{31} + x_{34} - x_{23} \leq 1 \quad (5.22)$$

Now we study all liftings of this inequality. There are many different sequences Π : 2680. Fortunately, a lot of those sequences yield the same values for (α, β) . For example, let $\pi_1 = \gamma_1$ and $\pi_2 = \beta_1$. Then $\gamma_1 = 0$, $\beta_1 = 1$, $\beta_2, \beta_3, \beta_4 \leq 0$, $\gamma_2 \leq 0$, and $\gamma_3, \gamma_4 \leq -1$. Because of these upper bounds, all other constraints are redundant. Regardless of the sequencing of the remaining coefficients, the result will be the same.

Not for all choices of π_2 the case is so trivial. Checking all possibilities, we end up with the following five different liftings:

$$\begin{array}{rcl}
 x_{12} + x_{31} + x_{34} - x_{23} + x_{15} - x_{53} - x_{54} & \leq & 1 \\
 x_{12} + x_{31} + x_{34} - x_{23} + x_{52} - x_{53} - x_{45} & \leq & 1 \\
 x_{12} + x_{31} + x_{34} - x_{23} - x_{25} & \leq & 1 \\
 x_{12} + x_{31} + x_{34} - x_{23} + x_{54} - x_{15} - x_{25} & \leq & 1 \\
 x_{12} + x_{31} + x_{34} - x_{23} + x_{52} + x_{53} + x_{54} - x_{15} - x_{45} - 2x_{25} & \leq & 1
 \end{array}$$

By renumbering the nodes, we can reduce these liftings to each other. For example, in the last lifting, reassign number 1 to 3, 2 to 4, 3 to 2, 4 to 1, and 5 to 5. using the degree equalities, one can find the first lifting.

5.6 Some Complete Descriptions for Small n .

In this section we present the complete and minimal description of the cycle polytope on 3, 4 and 5 nodes, and for each class of inequalities, the corresponding inequalities of the ATS polytope and the cycle polytope on one node less. We used a program of [Fukuda, 1994] to determine the complete descriptions.

Consider Table 5.1. It consists of three columns. The first one contains symbols referring to the classes of facet defining inequalities of the cycle polytope on three nodes. It also contains numbers, referring to the number of different facets defined by the corresponding class. The total number of facets is five. The second column contains graphical representations of the classes of facet defining inequalities of the cycle polytope on three nodes. The dotted lines refer to edges with coefficient minus one, the full lines to edges with coefficient plus one, and the omitted lines to edges with coefficients zero. In the third column, we have graphical representations of the classes of facet defining inequalities of the ATS polytope on three nodes, generating the corresponding classes of facet defining inequalities of the cycle polytope on three nodes.

For three nodes, there are two classes of facet defining inequalities. The degree inequalities, denoted with "D", and the zero-excluding inequalities,

denoted with "Z". There are three different degree inequalities. They can be abstracted from the degree equalities for the ATS polytope. They also could be lifted from the inequality $x_{12} \leq 1$, which defines a facet of the cycle polytope on two nodes. There are two zero-excluding inequalities defining different facets. Notice that there are two ways of abstracting these inequalities from the ATS description.

Table 5.2 has the same structure as Table 5.1, but now for four nodes. There is an extra column for the class of facet defining inequalities for the cycle polytope on three nodes. There are forty different facets, partitioned into five classes: nonnegativity inequalities (N), (the only class that can not be abstracted by the approaches), degree inequalities (D) (the only class that can be abstracted by both approaches), generalized two-subtour elimination inequalities (G) (valid but not facet defining for the ATS polytope on four nodes), zero-excluding inequalities (Z), and lifted cycle inequalities also known as T_2 inequalities (L, T_2).

Neither of the approaches is sufficient to generate the description of the cycle polytope on four nodes, but combined they are sufficient. This also holds for the cycle polytope on five nodes, as is depicted in Table 5.3.

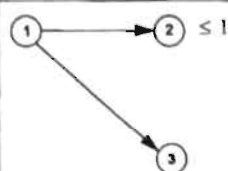
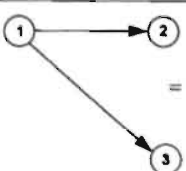
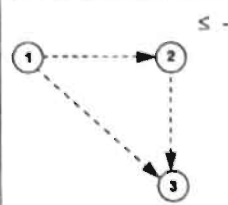
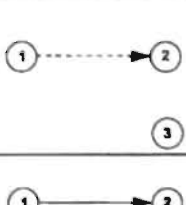
	facets $P_{\square}$	corresp. facets ATS_3
D		
Z		
n 5		

Table 5.1 Complete descriptions for $n = 3$

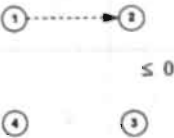
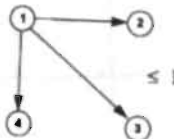
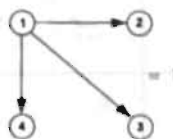
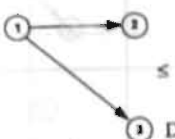
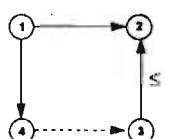
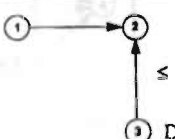
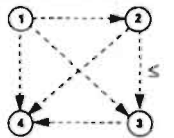
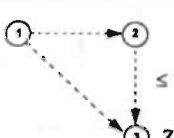
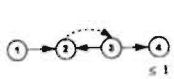
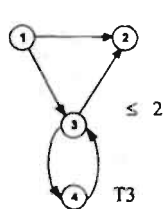
	facets P_{C_i}	Corr. ATS - facet	Corr. facet P_{C_i}
N 12			
D 4			
G 12			
Z 6			
T_2 Odd CP_5 6			
40			

Table 5.2 Complete descriptions for $n = 4$

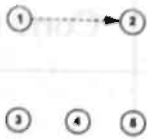
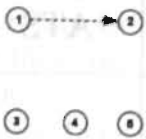
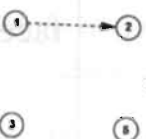
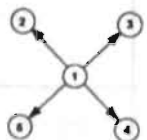
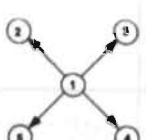
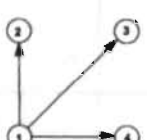
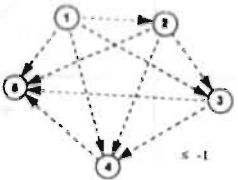
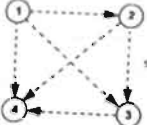
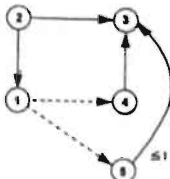
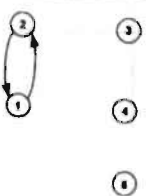
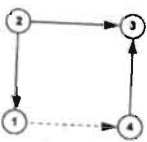
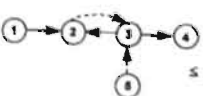
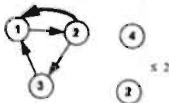
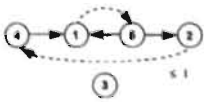
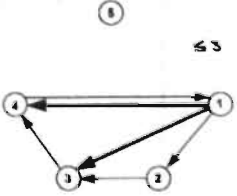
	facets P_{C_5}	facets ATS_5	facets P_{C_4}
N 20	 ≤ 0	 ≤ 0	 ≤ 0
D 5	 ≤ 1	 $= 1$	 ≤ 1
Z 24	 ≤ -1		 ≤ -1
G 60	 ≤ 1	 ≤ 1	 ≤ 1
L/T_3 120	 ≤ 1	 ≤ 2	 ≤ 1
D_4/C_3 120	 ≤ 1	 ≤ 3	

Table 5.3 Complete descriptions for $n = 5$

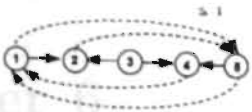
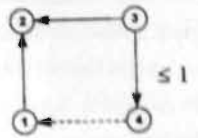
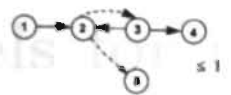
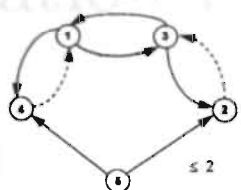
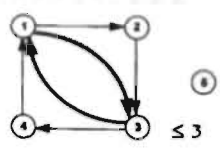
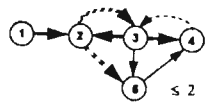
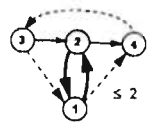
	facets $\mathcal{P}_{C_5}$	facets $\mathcal{ATS}_5$	facets $\mathcal{P}_{C_4}$
G^+ 60			
L^+ 60			
I_4 60			
u 120			
<u>709</u>			

Table 5.3 Complete descriptions for $n = 5$ (continued)

Chapter 6

Models for the Cell Formation Problem

This chapter is joint work with Yves Crama.

It has been published as an article in the International Journal of Production Research.

For cellular manufacturing strategies to succeed, the productive system first has to be divided into highly independent cells. This means that a partition of the machines into machine groups, a partition of the parts into part families, and a matching between the machine groups and the part families have to be simultaneously determined. Mathematically, this question can be expressed as the problem of finding a near block diagonal permutation of the machine-part incidence matrix. Research on such grouping problems has primarily concentrated on the design of heuristics. Different grouping efficiency criteria have been proposed to express the quality of the groupings proposed by these heuristics.

This chapter is concerned with mathematical programming approaches to the formation of production cells. Existing models are reviewed and their features are briefly discussed. An alternative model is proposed, which allows for the formulation of various constraints and grouping efficiency criteria. Finally, some test problems are used to support the claim that this model may be adequate for the solution to optimality of the cell formation problem.

6.1 Introduction

Due to increasing international competition and fastidiousness of customers, manufacturers are forced from mass production to the production of a larger product mix. To meet these new requirements, it is very important to have the ability to produce many small volume batches consisting of complex parts in a short production period. This leads to an increased complexity of the management task, increased investments in inventory and a decreased efficiency of mass production systems. To maintain high efficiency levels, it is an accepted strategy to adopt a group technology philosophy, and to organize a large portion of the manufacturing system into cells ([Schonberger, 1982, Hyer and Wemmerlöv, 1989]). A group technology cell consists of a number of machines located close to each other (a machine group) and geared for the manufacturing of a number of similar parts (a part family). To boost efficiency, tooling, loading and scheduling decisions within each cell should be made (almost) independently of the other cells. This requires that machine groups and part families be identified on basis of their interrelations, so that a minimum of interference occurs between the cells. Intercell relations can be dealt with in various ways. Transporting a part during its production process complicates the scheduling and controlling of the cells to a considerable extent. In-

stead of intercell movements, some other options can be considered like extending the number of machines of a certain type (machine duplication), allowing parts to be rerouted, reviewing the design of the bottleneck parts, or even transferring the production of a part to another production line (part subcontracting). For each of these options, heuristics or exact algorithms have been designed. For example [Kumar and Vanelli, 1987] present a heuristic for part subcontracting. Machine duplication has been considered by [Vanelli and Kumar, 1986]. Algorithms for combinations of these options are presented by [Vanelli and Kumar, 1986], [Kusiak, 1991], and [Kern and Wei, 1991]. The algorithms presented by [Sule, 1991] and [Hertz et al, 1991] also take intercell movements into account. Most of these approaches, however, assume that a decomposition of the production system into cells has already been determined. In this paper, we will exclusively deal with the problem of forming such cells.

The data for the cell formation problem is commonly assumed to be summarized in the *machine-part incidence matrix* $A = [a_{ij}]$, where $a_{ij} = 1$ if part j has to be processed on machine i and $a_{ij} = 0$ otherwise. More generally, the elements of A could also represent the processing time required by each part on each machine, or some other numerical data; the models to be discussed here also apply to this situation, with only slight modifications. A 'natural' machine-part grouping is expected to emerge when the rows and columns of the incidence matrix are reordered in a proper way. In that case the diagonal blocks should be relatively 'filled' compared to the other blocks, as illustrated in table below. In this example, deleting the element (1, 2) and adding the element (5, 6) would result in a perfect block diagonal structure.

part →		1	2	3	4	5	6
mach ↓	1	1	1	0	1	0	0
	2	0	1	1	0	0	1
	3	1	0	0	1	0	0
	4	0	0	0	0	1	0
	5	0	1	1	0	0	0

	1	4	2	3	6	5
3	1	1	0	0	0	0
1	1	1	1	0	0	0
2	0	0	1	1	1	0
5	0	0	1	1	0	0
4	0	0	0	0	0	1

Table 6.1 Example of near block diagonalization by permutation of rows and columns

The ones outside the diagonal blocks ('exceptions') represent intercell relations and should be avoided as much as possible. On the other hand, it is also undesirable to have zeroes in the diagonal blocks ('voids') or in

other words to have a part and a machine in the same cell when they are not directly related. Informally stated, the cell formation problem consists in minimizing some combination of the number of exceptions and of voids. We now proceed to define the problem more formally.

A matrix $X = [x_{ij}]$ is called *block diagonal* if there exists a partition $R_1, \dots, R_k, R_{k+1}$ of its row-set and a partition $C_1, \dots, C_k, C_{k+1}$ of its column-set such that $x_{ij} \neq 0$ if and only if, for some $1 \leq l \leq k$, $i \in R_l$ and $j \in C_l$ (notice that, to be very precise, we should say that X is block diagonal up to permutations of its rows and columns).

When X is a block diagonal machine-part incidence matrix, we interpret the partition of its rows and columns as describing the formation of k cells, where the l -th cell consists of the machines in R_l and the parts in C_l ($l = 1, \dots, k$). The machines in R_{k+1} and the parts in C_{k+1} are not included in any cell; in many applications R_{k+1} and C_{k+1} may be assumed empty.

If A and X are machine-part incidence matrices of the same dimension and X is block diagonal, then we say that element (i, j) is an *exception* of A (with respect to X) if $a_{ij} = 1$ and $x_{ij} = 0$; we call (i, j) a *void* of A (with respect to X) if $a_{ij} = 0$ and $x_{ij} = 1$.

The cell formation problem can be abstracted into the following *block diagonalization problem*:

given an incidence matrix A and a function $f(.,.)$, find a block diagonal incidence matrix X of same dimension as A which minimizes $f(A, X)$.

The function $f(.,.)$ is meant to give a measure of the distance, or dissimilarity, between A and X . It may for instance compute some weighted combination of the number of exceptions and voids. We will say more on this topic in Section 6.3.3. This basic formulation is often extended with additional constraints, e.g. on the minimal and maximal number of machines and/or parts per cell, or the number of cells (see Section 6.3.2). Such problems will be referred to as constrained block diagonalization problems.

It should be noticed that the block diagonalization model has many potential applications outside of the group technology framework discussed above. For instance, [McCormick et al, 1972] originally devised their bond energy algorithm as a clustering method for decomposing or organizing data arrays; this approach has subsequently been applied in many fields (see e.g. [Arabie and Hubert, 1990]. [Kumar et al, 1986] discuss the use of block diagonalization for production planning in flexible manufacturing systems. [King and Nakornchai, 1982] mention an application to the decomposition

of large scale linear programming problems. [Stern, 1991] is interested in sparse matrix computations.

Many heuristic approaches have been proposed for dealing with the block diagonalization problem. We now briefly review some of this work.

Even though the problem which they consider is slightly different from ours (they are more interested in 'seriating' the rows and columns than in 'partitioning' them), the bond energy algorithm developed by [McCormick et al, 1972] has been widely used to produce good solutions for the block diagonalization problem. Their ideas have been extended e.g. by [Askin et al., 1991, Hilger et al, 1991, Ng, 1991].

Another approach interprets the columns and rows as binary words and attempts to simultaneously reorder them in decreasing lexicographic order. This was first suggested by [King, 1980] (for a very efficient implementation, see [Page, Tarjan]); later variants were proposed e.g. by [Chan and Milner, 1982, King and Nakornchai, 1982, Chandrasekharan and Rajagopalan, 1986].

Other authors suggest to introduce similarity coefficients to measure the similarity of parts (respectively machines), and to determine a part (respectively machine) grouping based on a clustering algorithm; this may leave open the question of determining the best corresponding grouping of machines (respectively parts), see for example [Chandrasekharan and Rajagopalan, 1987, Kusiak and Heragu, 1987, Boe and Cheng, 1991, Srinivasan and Narendran, 1991].

[Faber and Carter, 1986], and [Vohra et al, 1990] view the block diagonalization problem as a network partitioning problem and present heuristics based on iteratively solving max-flow min-cut subproblems.

Local improvement algorithms are also suitable for determining groupings. The algorithm of [McCormick et al, 1972] is of this nature. A simulated annealing algorithm to deal with the problem was applied by [Stern, 1991].

Interestingly enough, many of the authors mentioned above do not precisely formulate the objective function which they consider: they are looking for a "nearly block diagonal" approximation of the incidence matrix, without specifying the meaning of "nearly" (see [Kusiak and Cheng, 1990]).

By contrast, a few researchers model the block diagonalization problem as a mixed-integer programming problem, thus stating exactly their objective function (see e.g. [McCormick et al, 1972, Kumar et al, 1986, Kusiak and Heragu, 1987, Boctor, 1989, Boctor, 1991, Vohra et al, 1990].

Moreover, we are only aware of two *exact* algorithms for the block diagonalization problem: [Boctor, 1991] develops an exact algorithm based on his

formulation; [Kusiak and Cheng, 1990] propose an exact algorithm for the constrained block diagonalization problem, based on a heuristic algorithm of [Kusiak and Chow, 1987]. Notice that, by exact algorithm, we mean here an algorithm that computes an *optimal* solution of an *explicitly* stated model, as opposed to a *heuristic* solution of either an explicit or an implicit model.

■ In this chapter a mathematical programming model for the block diagonalization problem is formulated. This model can be used to derive strong lower bounds on the optimal value of the problem. Such lower bounds could prove useful, e.g. to assess the quality of the heuristic approaches listed above, or, eventually, to develop an exact algorithm for the problem. In the next section, the model formulations found in the literature are reviewed, and it is shown that the use of these formulations cannot be expected to yield very strong lower bounds. A new formulation is proposed in Section 6.3. In Section 6.4, some small numerical examples are used to suggest that an algorithm based on the new formulation can be useful in deriving optimal solutions with respect to several objective functions. Finally, in the last section, a short discussion of the results is given.

6.2 Review of Existing Mathematical Programming Formulations

It is by now well-known that the choice of a “good” model may crucially affect our ability to solve large instances of integer programming problems. For a thorough discussion of this point, see [Nemhauser and Wolsey, 1988]. In this section, we survey the main mathematical programming models previously proposed in the literature for the cell formation problem, and we briefly discuss the quality of the lower bounds which they yield. We systematically leave out of the discussion those models which do not address the cell formation problem as formulated in Section 6.1 (e.g., the quadratic assignment model of [McCormick et al, 1972], or the clustering models surveyed in [Kusiak and Heragu, 1987], which only determine a partition of the machine set but not of the part set).

6.2.1 Formulations of Existing Models

All mathematical programming models discussed here have been originally developed with the objective of minimizing the (weighted) sum of exceptions, under some additional constraints restricting the number of machines and

parts in each cell, and/or limiting the number of cells. But they can easily be extended to formulate more general objective functions.

One of the first such models is the quadratic integer programming formulation with linear constraints of [Kumar et al, 1986]. The variables and constraints in their formulation occur in all other formulations to be discussed in this section.

Let M be the number of machines, P the number of parts, and A the given incidence matrix. [Kumar et al, 1986] also assume that an upper bound K is given on the number of cells to be formed; if this upper bound K is not exogenously given, a similar model can be formulated with $K = \min(M, P)$, and the number of cells in the optimal solution of the problem is then completely unconstrained at the outset. The indices i, j and k will be used for machines, parts and cells respectively. Now introduce binary variables m_{ik} and p_{jk} , where $m_{ik} = 1$ if machine i is assigned to cell k , $m_{ik} = 0$ otherwise, and $p_{jk} = 1$ if part j is assigned to cell k , $p_{jk} = 0$ otherwise. Then the constraints

$$\sum_{k=1}^K m_{ik} = 1 \quad i \in \{1 \dots M\} \quad (6.1)$$

$$\sum_{k=1}^K p_{jk} = 1 \quad j \in \{1 \dots P\} \quad (6.2)$$

$$m_{ik}, p_{jk} \in \{0, 1\} \quad i \in \{1 \dots M\}, j \in \{1 \dots P\}, k \in \{1 \dots K\} \quad (6.3)$$

express that each machine and each part must be assigned to exactly one cell. Thus, each feasible solution of (6.1)–(6.3) implicitly defines a block-diagonal matrix $X = [x_{ij}]$, where $x_{ij} = 1$ if $m_{ik} = p_{jk} = 1$ for some k in $\{1, \dots, K\}$. Furthermore, [Kumar et al, 1986] describe cardinality constraints that can be added to restrict the number of parts and machines in each cell, for instance:

$$L \leq \sum_{i=1}^M m_{ik} + \sum_{j=1}^P p_{jk} \leq U \quad k \in \{1 \dots K\} \quad (6.4)$$

where L and U are given bounds. [Kumar et al, 1986] propose to solve the model

$$\max \sum_{i=1}^M \sum_{j=1}^P \sum_{k=1}^K a_{ijk} m_{ik} p_{jk} \quad (6.5)$$

$$\text{subject to (6.1)–(6.4)} \quad (6.6)$$

It is easy to see that maximizing the function (6.5) is tantamount to minimizing the number of exceptions under the constraints imposed on the number and size of the cells. Similarly, any weighted sum of the number of exceptions and voids could be expressed as a quadratic function in the variables m_{ik} and p_{jk} .

The model (6.5)–(6.6) features a quadratic objective function and, as such, does not easily lend itself to the traditional approaches used in integer programming. It would therefore be desirable to replace the quadratic function (6.5) by a linear expression. This goal is achieved by substituting for (6.5)–(6.6) an easy transportation model which, however, only provides an approximation of the original model. More classically, the function (6.5) can also be linearized by replacing each quadratic term $m_{ik}p_{jk}$ by a new binary variable y_{ijk} , with the interpretation that $y_{ijk} = 1$ if machine i and part j are both assigned to cell k , and $y_{ijk} = 0$ otherwise. To enforce the relations $y_{ijk} = m_{ik}p_{jk}$, the following constraints can be added to the model: for each $i \in \{1, \dots, M\}$, $j \in \{1, \dots, P\}$, $k \in \{1, \dots, K\}$,

$$\left. \begin{array}{l} y_{ijk} \leq m_{ik} \\ y_{ijk} \leq p_{jk} \end{array} \right\} \quad (6.7)$$

$$y_{ijk} \geq m_{ik} + p_{jk} - 1 \quad (6.8)$$

(see e.g. [Hansen, 1979]). The inequalities (6.8) are unnecessary if the objective function only involves nonnegative coefficients, as is the case with (6.5). They become necessary when the objective function contains negative coefficients, which is for instance the case when the number of voids is accounted for.

When applied to the model of [Kumar et al, 1986], these linearizing steps lead to the formulation

$$\max \sum_{i=1}^M \sum_{j=1}^P \sum_{k=1}^K a_{ij} y_{ijk} \quad (6.9)$$

$$\text{subject to (6.1)–(6.4), (6.7).} \quad (6.10)$$

[Vohra et al, 1990] base their Lagrangian approach on this formulation, as will be further discussed in the next subsection.

A related linearization is proposed by [Boctor, 1989], citeBo91. His approach can be viewed as starting with the objective function

$$\min 1/2 \sum_{i=1}^M \sum_{j=1}^P \sum_{k=1}^K a_{ij} ((1 - m_{ik})p_{jk} + m_{ik}(1 - p_{jk})) \quad (6.11)$$

which exactly computes the number of exceptions, rather than with the function (6.5) used by [Kumar et al, 1986]. Then, new variables u_{ijk} and v_{ijk} are introduced with the following interpretation: $u_{ijk} = 1$ if part j is assigned to cell k and machine i is not, and $u_{ijk} = 0$ otherwise; similarly $v_{ijk} = 1$ if machine i is assigned to cell k and part j is not, and $v_{ijk} = 0$ otherwise. This leads to the model:

$$\min 1/2 \sum_{i=1}^M \sum_{j=1}^P \sum_{k=1}^K a_{ij}(u_{ijk} + v_{ijk}) \quad (6.12)$$

$$\text{subject to (6.1)–(6.4) and} \quad (6.13)$$

$$m_{ik} - p_{jk} + u_{ijk} - v_{ijk} = 0 \quad \begin{array}{l} i \in \{1 \dots M\}, \\ j \in \{1 \dots P\}, k \in \{1 \dots K\} \end{array} \quad (6.14)$$

$$u_{ijk} \geq 0, v_{ijk} \geq 0 \quad \begin{array}{l} i \in \{1 \dots M\}, \\ j \in \{1 \dots P\}, k \in \{1 \dots K\} \end{array} \quad (6.15)$$

where the constraints (6.14) (resp. (6.15)) enforce the relations $u_{ijk} = (1 - m_{ik})p_{jk}$ (resp. $v_{ijk} = m_{ik}(1 - p_{jk})$). Bector observes that the integrality conditions on the variables p_{jk} can be dropped without altering this model (actually this author works with constraints slightly different from (6.4), but his observations are easily seen to carry over to the above model). At the cost of introducing more variables, it is straightforward to formulate a similar model aimed at the minimization of a weighted sum of exceptions and voids.

6.2.2 Strength of the Formulations

The efficiency of most algorithms solving integer programming problems is dependent on the availability of sharp bounds on the value of the objective function. Therefore an important criterion in judging of the quality of an integer programming model concerns its ability to provide such bounds. We are now going to evaluate the models described above on the basis of this criterion.

A common way to bound the optimal value of an integer programming problem is to solve the continuous relaxation of this problem obtained by dropping the integrality requirements on the variables. Consider for instance the model (6.9)–(6.10) of [Vohra et al, 1990], and replace the constraints (6.3) by: $0 \leq m_{ik} \leq 1$, $0 \leq p_{jk} \leq 1$ for all i, j, k . It is straightforward to check that, if this relaxation is feasible at all, then it admits the feasible solution $m_{ik} = p_{jk} = y_{ijk} = 1/K$ for all i, j, k . This implies that the optimal

value of the relaxed model is at least equal to $\sum_i \sum_j a_{ij}$, which is a very trivial upper bound for the original integer model. Similarly, the continuous relaxation of Boctor's model (6.12)–(6.15) admits the optimal solution $m_{ik} = p_{jk} = 1/K$, $u_{ijk} = v_{ijk} = 0$ for all i, j, k , which yields 0 as a trivial lower bound on the number of exceptions. Thus, we conclude that the continuous relaxations of both models are weak.

[Vohra et al, 1990] and [Boctor, 1991] suggest to strengthen the constraint set (6.1) by requiring the first machine to be in the first cell, the second machine to be in one of the first two cells, and so on. This can be expressed by the additional constraints ([Boctor, 1991]):

$$m_{ik} = 0 \quad i \in \{1, \dots, K-1\}, k \in \{i+1, \dots, K\} \quad (6.16)$$

(the formulation of this requirement proposed by [Vohra et al, 1990] is incorrect, as it allows for instance $m_{11} = m_{12} = 1$). Constraints (6.16) might indeed result in a strengthening of the bounds obtained by continuous relaxation, and [Vohra et al, 1990] report some evidence to this effect. The bounds which they compute remain, however, rather far from the value of the best known heuristic solutions.

Another way to determine bounds on the optimal value is to apply Lagrangian relaxation. For instance, as suggested by [Vohra et al, 1990], an upper bound $Z(\lambda, \mu)$ for the model (6.9)–(6.10) can be obtained by dualizing the linearizing inequalities (6.7) with nonnegative multipliers λ_{ijk} and μ_{ijk} , and by solving the Lagrangian dual problem

$$Z(\lambda, \mu) = \max \left\{ \sum_i \sum_j \sum_k a_{ij} y_{ijk} + \sum_i \sum_j \sum_k a_{ij} y_{ijk} \{ (m_{ik} - y_{ijk}) \lambda_{ijk} + (p_{jk} - y_{ijk}) \mu_{ijk} \} \right\} \quad (6.17)$$

$$\text{subject to (6.1)–(6.4) and} \quad (6.18)$$

$$y_{ijk} \in \{0, 1\} \text{ for all } i, j, k \quad (6.19)$$

Ventura et al. (1990) also add to this model the constraints

$$\sum_{k=1}^K y_{ijk} \leq 1 \quad i \in \{1 \dots M\}, j \in \{1 \dots P\} \quad (6.20)$$

which were implied by (6.1) and (6.7) in the original model, but are no longer redundant in the Lagrangian dual (6.17)–(6.19).

Rearranging the terms, we get

$$\begin{aligned}
 Z(\lambda, \mu) = & \sum_i \sum_j a_{ij} \max \left\{ \begin{array}{l} \sum_{k=1}^K (1 - \lambda_{ijk} - \mu_{ijk}) y_{ijk} \\ \text{subject to (6.19)–(6.20)} \end{array} \right\} \\
 & + \max \left\{ \begin{array}{l} \sum_{i=1}^M \sum_{k=1}^K \sum_{j=1}^P a_{ij} \lambda_{ijk} m_{ik} + \sum_{j=1}^P \sum_{k=1}^K \sum_{i=1}^M a_{ij} \mu_{ijk} p_{jk} \\ \text{subject to (6.1)–(6.4)} \end{array} \right\} \quad (6.21)
 \end{aligned}$$

So computing $Z(\lambda, \mu)$ boils down to solving $(M * P + 1)$ subproblems, namely $M * P$ easy knapsack problems and a network flow problem. A better upper bound, say Z_D , can be obtained by minimizing $Z(\lambda, \mu)$ over all possible choices of the Lagrange multipliers λ_{ijk} and μ_{ijk} . This gives rise to a continuous nonlinear minimization problem, which [Vohra et al, 1990] handle by a subgradient procedure.

Let us observe here that all the subproblems arising in the computation of $Z(\lambda, \mu)$ have the so-called “integrality property”, meaning that the integrality requirements (6.3) and (6.19) can be dropped from the formulation (6.21) without altering its optimal solution. It is a well-known result that, in such a case, the best bound Z_D computable by Lagrangian relaxation is equal to the bound obtained by continuous relaxation. Thus, it follows from our previous discussion that the bound Z_D is again a weak one, certainly not better than the trivial upper bound $\sum_i \sum_j a_{ij}$.

A similar reasoning applies to the model (6.17)–(6.20) strengthened by the constraints (6.16). Here also, the bound obtained for a best possible choice of the Lagrange multipliers is equal to the optimal value of the continuous relaxation of (6.9), (6.10), (6.16). From this viewpoint, the only (though not negligible) advantage of the Lagrangian approach is that it is computationally more manageable than the continuous relaxation approach.

All these remarks are easily transposable to the model (6.12)–(6.15) proposed by [Boctor, 1991].

6.3 A new Mathematical Programming Model

In Section 6.2, we argued that the mathematical programming models found in the literature for the block diagonalization problem are characterized by weak continuous relaxations, and hence, are not well suited for the solution to optimality of large scale instances. In the present section, we propose a

new model for the block diagonalization problem. We show that this model can accommodate many of the restrictions and objective functions considered in the literature.

6.3.1 Model Formulation

Consider again the formulation of the block diagonalization problem stated in Section 6.1: the problem consists in determining a 'close' approximation of the $M \times P$ machine-part incidence matrix A by a 0-1 block diagonal matrix X . In view of this formulation, it is very natural to choose as decision variables the elements of X , with their obvious interpretation:

$$x_{ij} \begin{cases} = 1 & \text{if machine } i \text{ and part } j \text{ are in the same cell,} \\ & \text{for } i \in \{1, \dots, M\} \text{ and } j \in \{1, \dots, P\} \\ = 0 & \text{otherwise.} \end{cases}$$

Block diagonal matrices can be characterized as follows:

Proposition 6.3.1 *The $M \times P$ 0-1 matrix X is block diagonal if and only if its elements satisfy the inequalities*

$$x_{ij} + x_{hj} + x_{ik} - x_{hk} \leq 2 \text{ for } i, h \in \{1, \dots, M\} \text{ and } j, k \in \{1, \dots, P\} \quad (6.22)$$

To present a short proof of this proposition, it is useful to think of an alternative formulation of the block diagonalization in graph theoretic terms (see e.g. [King and Nakornchai, 1982] or [Kumar et al, 1986]; we use the graph theoretic terminology of [Bondy and Murty, 1976]). Every $M \times P$ binary matrix X can be viewed as the adjacency matrix of a bipartite graph $G(X)$ with vertex set $\{1, \dots, M\} \cup \{1, \dots, P\}$; there is an edge $\{i, j\}$ between vertices $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, P\}$ if and only if $x_{ij} = 1$. Recall that a bipartite graph on the vertex set $V_1 \cup V_2$ is called complete bipartite if it contains all possible edges between V_1 and V_2 , i.e. if its edge set is $V_1 \times V_2$. It is easy to see that a matrix X is block diagonal if and only if each connected component of $G(X)$ is a complete bipartite graph. So, the block diagonalization problem can be interpreted as the problem of approximating a given bipartite graph G by a disjoint union of complete bipartite graphs on the same vertex set as G .

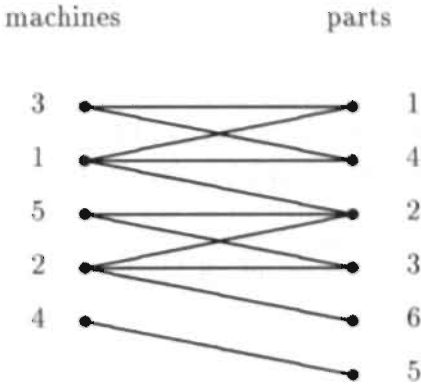


Figure 6.1 Example of a graph representation of machine–part relations

As an illustration, Figure 6.1 shows the graph associated with the incidence matrix displayed in Table 6.1.1. Notice that deleting the edge (1,2) from this graph and adding the edge (5,6) would produce three complete bipartite components, associated with the three diagonal blocks of the corresponding matrix.

The inequalities (6.22) have an obvious interpretation in graph theoretic terms: if all components of $G(X)$ are complete bipartite, and if three edges $\{i, j\}$, $\{i, k\}$ and $\{h, j\}$ are present in $G(X)$, then $\{h, k\}$ must also be an edge. This interpretation suggests to refer to (6.22) as the *square inequalities*.

We are now ready for a proof of Proposition 6.3.1.

Proof: (Only if) Suppose that X is a 0–1 matrix which violates (6.22) for some choice of i, h, j, k . Then it must be the case that $x_{ij} = x_{hj} = x_{ik} = 1$ and $x_{hk} = 0$, which easily implies that X is not block diagonal.

(If) Assume by contradiction that X is a 0–1 matrix which satisfies (6.22), but that X is not block diagonal. If $G = G(X)$ is the bipartite graph associated with X , then G has a component which is not complete bipartite. In other words, there are two vertices of G , say $h \in \{1, \dots, M\}$ and $v \in \{1, \dots, P\}$, such that h and v are in the same component of G but $\{h, v\}$ is not an edge. Consider a shortest path in G from h to v , say $P = (h, j, i, k, \dots, v)$. Notice that $j \neq v$, since $\{h, v\}$ is not an edge. On the other hand, $\{i, j\}$, $\{i, k\}$ and $\{h, j\}$ are edges of G , and hence (6.22) implies that $\{h, k\}$ also is an edge. But then $(h, k, \dots, v)$ is a path from h to v and is shorter than P : contradiction. $\square$

So, the block diagonal incidence matrices correspond exactly to the integer solutions of the system (6.23)–(6.24) :

$$x_{ij} + x_{hj} + x_{ik} - x_{hk} \leq 2$$

$$\text{for } i, h \in \{1, \dots, M\} \text{ and } j, k \in \{1, \dots, P\} \quad (6.23)$$

$$0 \leq x_{ij} \leq 1 \quad i \in \{1, \dots, M\}, j \in \{1, \dots, P\} \quad (6.24)$$

Some limited computational experience seems to indicate that this system provides a useful description of the set of all block diagonal matrices (see Section 6.4). This impression can be confirmed by a couple of theoretical results. To state these results, let us denote by $Q_{M,P}$ the convex hull of all 0–1 solutions of (6.23)–(6.24); that is, $Q_{M,P}$ is the polyhedron whose vertices are exactly the block diagonal matrices of dimension $M \times P$. We prove in Chapter 7 that the inequalities (6.23) define facets of $Q_{M,P}$, and that $Q_{M,P}$ is completely described by (6.23)–(6.24) when $M = 2$. We also produce in that chapter several other classes of inequalities which define additional facets of $Q_{M,P}$.

Our ultimate goal is to develop a cutting-plane approach for the block diagonalization problem, based on the above formulation of the problem in terms of the variables x_{ij} (recall that a cutting-plane procedure essentially relies on the iterative tightening of an initial formulation of the problem by addition of a possibly large number of valid inequalities; see [Nemhauser and Wolsey, 1988]). Such an approach was successfully applied by [Faigle et al, 1986] and [Grötschel and Wakabayashi, 1990] to a problem closely related to the block diagonalization problem, namely the following clique partitioning problem: compute a ‘good’ approximation of a (nonbipartite) graph by a disjoint union of complete (nonbipartite) graphs (recall that a graph is complete if it contains an edge between each pair of vertices). Thus, the clique partitioning problem can clearly be seen as a nonbipartite analogue of the block diagonalization problem. [Faigle et al, 1986] and [Grötschel and Wakabayashi, 1990] relied on a formulation of the clique partitioning problem using the binary variables r_{ij} , where r_{ij} takes the value one if and only if vertex i and vertex j belong to the same complete subgraph. They observed that the following ‘triangle inequalities’ provide a very tight description of the set of feasible solutions to their problem:

$$r_{ij} + r_{ik} - r_{kj} \leq 1 \text{ for all } i, j, k. \quad (6.25)$$

Intuitively, the inequalities (6.25) express that, if two edges $\{i, j\}$ and $\{i, k\}$ are present in a complete subgraph, then the edge $\{k, j\}$ should also be

present. The similarity of these 'triangle inequalities' with our 'square inequalities' should be obvious.

It is possible to compute heuristic solutions as well as nontrivial lower bounds for instances of the block diagonalization problem by transforming these instances into instances of the clique partitioning problem, and applying techniques already developed for the latter problem. In this transformation the bipartite graph has to be extended to a complete graph by adding edges with zero weight. This causes the number of edges, and therefore the number of variables in the LP formulation, to double, without yielding any extra information. Therefore we prefer to handle the formulation (6.23)–(6.24), which involves less variables and remains closer to the original structure of the problem.

Another reason why we prefer the formulation (6.23)–(6.24) is that the complexity of the problem is determined by the minimum of the number of parts and the number of machines. For example, if the number of machines is fixed, the problem is computationally easy to solve. To see this, notice that for each partitioning of the machines into machine groups it is straightforward how to partition the parts into part families such that the weighted sum of exceptions and voids is minimal. Approaching the block diagonalization problem with techniques designed for the clique partitioning problem, we cannot make use of this observation.

6.3.2 Additional Constraints

In the previous subsection, we showed that the basic unconstrained block diagonalization problem can be modelled using the variables x_{ij} .

More specifically, in the context of the cell formation problem, any solution of the system (6.23)–(6.24) implicitly determines a partition of the parts and machines into cells, without a priori restrictions on the number or the composition of the cells.

But, as mentioned in Chapter 1, Section 3.7, and Section 6, it is sometimes necessary to extend this basic model with additional constraints in order to express limitations of a physical, technological or organizational nature. These constraints may express special relations between machines and/or parts; they may also be cardinality constraints on the size or the number of cells. We now briefly discuss how the basic model can be extended to handle such constraints. Notice that we are not concerned here by the origin of these constraints, but simply by the fact that the modeller may sometimes wish to consider them as being exogeneously imposed.

Machine-part relations.

Expressing that machine i and part j must be in the same cell can be done very simply by fixing x_{ij} to one. Similarly, fixing x_{ij} to zero implies that machine i and part j may not be in the same cell.

Machine-machine or part-part relations.

If machines h and i must be in the same cell, then the equations $x_{hj} = x_{ij}$ can be added to the model for all j in $\{1, \dots, P\}$. A similar modification applies when parts j and k must be in the same cell.

It is more delicate to model the relation between two machines (or two parts) which are not allowed to be in the same group. For two machines i and h , this can be modelled by set packing inequalities like:

$$x_{hj} + x_{ij} \leq 1 \quad j \in \{1, \dots, P\},$$

but set packing constraints often have weak continuous relaxations. Another possibility would be to introduce a dummy part p and to impose $x_{hp} = 1$, $x_{ip} = 0$ (machine h is in the same group as p , while machine i is not).

Constraints on cell size.

Assume that a lower bound L and an upper bound U are exogeneously imposed on the number of machines in any cell. This can be expressed by the constraints

$$L \leq \sum_i x_{ij} \leq U \quad j \in \{1, \dots, P\},$$

since, by definition of the variables x_{ij} , $\sum_i x_{ij}$ is the number of machines in the cell to which part j is assigned. Such constraints can also be used to prevent the formation of cells involving either no parts or no machines, by letting $L = 1$ (see Section 6.1). One can similarly restrict the number of parts in any cell.

Alternatively, suppose that L and U are respectively lower and upper bounds on the total number of machines *and* parts in any cell (as in Section 6.2, constraint (6.4)). This can be modelled by the inequalities

$$Lx_{ij} \leq \sum_k x_{ik} + \sum_h x_{hj} \leq 2U - Ux_{ij} \quad i \in \{1, \dots, M\}, j \in \{1, \dots, P\}.$$

Indeed, when $x_{ij} = 1$, this constraint restricts the number of machines and parts in the cell containing i and j to lie between L and U ; if $x_{ij} = 0$, then

the constraint is redundant.

Constraints on the number of cells.

Suppose now that, for any imaginable reason, the machines may be assigned to at most C cells, for some given C . This can be enforced by introducing C dummy parts, numbered from $P + 1$ to $P + C$, and imposing the additional constraints:

$$\sum_{k=1}^C x_{i,P+k} = 1 \quad i \in \{1, \dots, M\} \quad (6.26)$$

(of course, the square inequalities (6.22) must bear now on all variables x_{ij} , for $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, P+C\}$). The constraints (6.26) imply that each machine must be assigned to a cell containing a dummy part. Thus, at most C machine groups may be formed. In fact, the variables $x_{i,P+k}$, for $k \in \{1, \dots, C\}$, can best be identified with the variables m_{ik} introduced in Section 6.2; that is, $x_{i,P+k} = 1$ means that machine i is assigned to cell k . Then, (6.26) can be identified with (6.1). This also suggests that the formulation can be strengthened by adding constraints similar to (6.16), that is:

$$x_{i,P+k} = 0 \quad i \in \{1, \dots, C-1\}, k \in \{i+1, \dots, C\}.$$

6.3.3 Objective Functions

In this subsection, we discuss some of the objective functions which have been proposed in the literature for the block diagonalization problem, and we show how they can be modelled using the x_{ij} variables. We also point to limitations of our model with this regard.

Let us assume that we are approximating an incidence matrix A by the block diagonal matrix X . We denote by $e(X)$ (resp. $v(X)$) the number of exceptions (resp. voids) of A with respect to X . As mentioned earlier, a basic criterion for the goodness of block diagonalization is defined by the sum $e(X) + v(X)$. More generally, let e_{ij} (resp. v_{ij}) be an (arbitrary) penalty to be incurred when element (i, j) is an exception (resp. a void) of A with respect to X (these penalties could reflect the cost of dealing with the exceptions and voids, for example the cost of extra transportation or machine idle time). Then, the total cost of the approximation can be measured by the sum of these penalties, that is:

$$f(A, X) = \sum_{ij} \{e_{ij}a_{ij}(1 - x_{ij}) + v_{ij}(1 - a_{ij})x_{ij}\} \quad (6.27)$$

Minimizing this function is equivalent to minimizing a function of the form

$$\sum_{ij} w_{ij} x_{ij}, \quad (6.28)$$

for an appropriate definition of the weights w_{ij} . When $e_{ij} = v_{ij} = 1$ for all $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, P\}$, then (6.27) boils down to $e(X) + v(X)$. Similarly, each of the objective functions $e(X)$ and $v(X)$ can be obtained as a special case of (6.27). An objective function of type (6.27) also arises if we measure the quality of the approximation by some weighted Hamming distance $d(A, X)$ between A and X . Recall that a weighted Hamming distance is defined as:

$$d(A, X) = \sum_{ij} q_{ij} |a_{ij} - x_{ij}| = \sum_{ij} q_{ij} (a_{ij} - x_{ij})^2,$$

where q_{ij} are arbitrary nonnegative parameters. Equivalently,

$$d(A, X) = \sum_{ij} q_{ij} \{a_{ij}(1 - x_{ij}) + (1 - a_{ij})x_{ij}\},$$

which shows that $d(A, X)$ is a function of type (6.27) with $e_{ij} = v_{ij} = q_{ij}$ for all i, j . In particular, the unweighted Hamming distance assumes $q_{ij} = 1$ for all i, j , and hence coincides again with $e(X) + v(X)$.

Various criteria are also extensively discussed by [Kumar and Chandrasekharan, namely: grouping efficiency, modified grouping efficiency and grouping efficacy. All these criteria take values in $[0, 1]$; larger values correspond to more desirable partitions.

Grouping efficiency is defined as:

$$\eta(A, X) = q\eta_1 + (1 - q)\eta_2 \quad (6.29)$$

where q is a parameter, $0 \leq q \leq 1$, and

$$\begin{aligned} \eta_1 &= \sum_{ij} a_{ij} x_{ij} / \sum_{ij} x_{ij} \\ \eta_2 &= \sum_{ij} (1 - a_{ij})(1 - x_{ij}) / \sum_{ij} (1 - x_{ij}) \end{aligned}$$

(in words, η_1 is the number of ones of A in the blocks divided by the total number of elements in the blocks, and η_2 is the number of zeroes of A outside the blocks divided by the number of elements outside the blocks). Grouping efficiency appears to be a highly nonlinear expression in the variables x_{ij} , and

we do not see how to accomodate it in our model. [Kusiak and Cheng, 1990] also mentioned some drawbacks of this criterion when q is arbitrarily chosen. On the other hand, they noticed that, when $q = \sum_{ij} x_{ij}/MP$, then (6.29) becomes

$$\bar{\eta}(A, X) = 1 - (e(X) + v(X))/MP.$$

They call $\bar{\eta}(A, X)$ the *modified grouping efficiency* (of A with respect to X). Obviously, maximizing $\bar{\eta}$ is equivalent to minimizing our well-known basic criterion $e(X) + v(X)$.

[Kusiak and Cheng, 1990] suggest that the following *grouping efficacy* criterion may be the fittest for the group technology cell formation problem:

$$\Gamma(A, X) = (C - e(X))/(C + v(X)),$$

where C denotes the number of ones in A . The function Γ is again a nonlinear function of X , and, as such, may seem difficult to maximize by classical integer programming techniques. However, its maximization can be reduced to a finite sequence of 0–1 linear optimization problems by relying on a well-known approach for fractional programming problems (this approach goes back to [Dinkelbach, 1967]; see for example [Gondran and Minoux, 1984] for a general presentation). The gist of this approach is to iteratively improve a current solution X_k , in the following way. Let $\Gamma_k = \Gamma(A, X_k)$. There exists a better block diagonal matrix than X_k with respect to grouping efficacy if and only if the inequality

$$(C - e(X))/(C + v(X)) > \Gamma_k$$

has a solution (over the set of block diagonal matrices), i.e., if and only if the minimum of the following expression is negative:

$$\Gamma_k(C + v(X)) + e(X) - C. \quad (6.30)$$

Now, (6.30) can be rewritten as

$$\Gamma_k \sum_{ij} (1 - a_{ij})x_{ij} + \sum_{ij} a_{ij}(1 - x_{ij}) + (\Gamma_k - 1)C. \quad (6.31)$$

Thus, we can find out whether the current matrix X_k is optimal by minimizing (6.31) (that is, a linear expression of type (6.27)) over the set of block diagonal matrices which is equivalent to the set of integer solutions of (6.23)–(6.24). This is a linear programming problem in 0–1 variables. If the minimum of (6.31) is zero, then X_k is optimal with respect to the

grouping efficacy criterion. Otherwise, a better solution X_{k+1} is found and the procedure can be iterated. Notice that, since the numerator of $\Gamma(A, X)$ takes integer values in $[0, C]$ and its denominator takes integer values in $[C, MP]$, the number of possible values of $\Gamma(A, X)$ is finite. Therefore, the improving sequence $X_1, X_2, \dots$ is also finite. In fact, if binary search over the range of values of Γ is used, then an optimal X can be computed by solving $O(\log(MP))$ linear subproblems in $0 - 1$ variables.

6.4 Numerical examples

We now illustrate the use of the models described in the previous sections on three examples from the literature: a (10×15) instance from [Chan and Milner, 1982], a (8×20) instance from [Chandrasekharan and Rajagopalan, 1986], and a (13×25) instance from [Leskowski et al., 1987].

Example 6.1 The machine-part incidence matrix A for this example is displayed in Table 6.2.

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5
1		1				1	1				1	1			
2					1				1	1			1		1
3	1		1			1								1	
4	1		1	1		1								1	
5								1	1	1			1		1
6		1				1	1				1	1			
7		1				1					1	1			
8	1		1	1		1					1			1	
9			1	1		1	1							1	
10								1	1	1			1		1

Table 6.2 Data set 6.1 [Chan and Milner, 1982]

Assume that we want to find a close approximation of this matrix by a block diagonal matrix X , with the objective of maximizing the modified grouping efficiency $\bar{\eta}(A, X)$. As discussed in Section 6.3, this can be equivalently achieved by solving the following linear programming problem in $0 - 1$ variables (see (6.23), (6.24), (6.27)):

$$\min e(X) + v(X) = \sum_{ij} a_{ij}(1 - x_{ij}) + \sum_{ij} (1 - a_{ij})x_{ij} \quad (6.32)$$

subject to

$$x_{ij} + x_{hj} + x_{ik} - x_{hk} \leq 2 \quad i, h \in \{1, \dots, M\}, j, k \in \{1, \dots, P\} \quad (6.33)$$

$$0 \leq x_{ij} \leq 1 \quad i \in \{1, \dots, M\}, j \in \{1, \dots, P\} \quad (6.34)$$

$$x_{ij} \text{ integer} \quad i \in \{1, \dots, M\}, j \in \{1, \dots, P\} \quad (6.35)$$

with $M = 10$ and $P = 15$. Our first goal is to solve the continuous relaxation (6.32)–(6.34) of this model. Observe that the number of square inequalities (6.33) is extremely large. In general, for an $M \times P$ instance matrix, there are $M(M-1)P(P-1)$ square inequalities, i.e. 18,900 in this special case. Few linear programming packages can explicitly handle such a large number of constraints (not to speak of the number of constraints arising for larger, more realistic problem instances).

In view of this difficulty, we adopt a cutting plane approach, as described for instance in [Nemhauser and Wolsey, 1988]. This approach can be briefly described as follows:

Step 1 select a subset, say S , of the inequalities (6.33)–(6.34);

Step 2 solve the linear programming submodel defined by the objective function (6.32) and the inequalities in S (observe that the optimal value of this submodel already yields a lower bound on the optimal value of (6.32)–(6.35));

Step 3 check whether the current solution X satisfies all inequalities (6.33)–(6.34); if yes, then X is optimal for (6.32)–(6.34); otherwise, enlarge the set S with some of the violated inequalities, and return to Step 2.

Of course, many questions must be answered in order to turn this informal procedure into an efficient algorithm: how should the initial set S be chosen in Step 1, how should we check whether a current solution satisfies (6.33)–(6.34), and which violated inequalities should be added to S in Step 2, etc. These questions can only be meaningfully answered after extensive testing on proper data sets. In our preliminary implementation of this algorithm, we choose as initial set S the collection of trivial inequalities (6.34) (hence, the optimal solution found in the first execution of Step 2 is $X = A$). We check by complete enumeration whether the system (6.33)–(6.34) is satisfied, and we only add to S the first 25 (or, in the first execution of Step 2, the first 100) violated inequalities which are encountered. Moreover, in order to keep the model small, we also remove from S in Step 3 all constraints

which are nonbinding at the current solution (notice that a same inequality may be selected and removed several times, and that this may theoretically cause the procedure to cycle).

When applied to our example, this procedure yields the results summarized in Table 6.3.

solu	# of ineq total/ binding	$e(X) + v(X)$	$\bar{\eta}$	# of pivots	# of violated ineq
X_0	0/0	$0 + 0 = 0$	1.000	0	184
X_1	100/65	$8 + 1.5 = 9.5$	0.937	59	87
X_2	90/84	$6 + 4 = 10$	0.933	71	37
X_3	109/109	$6 + 4 = 10$	0.933	68	0

Table 6.3 Performance of the cutting plane algorithm on Example 6.1; objective function (6.32).

A sequence of four solutions $X_0 (= A), X_1, X_2, X_3$ is computed by the algorithm (in four successive runs of Step 2). The second column of Table 6.3 indicates the number of square inequalities in the current model, and how many of these are satisfied at equality by the optimal solution of the model. Column 3 gives the optimal value of the objective function (6.32) for the current model, and column 4 gives the corresponding value of $\bar{\eta} = 1 - (e(X) + v(X))/MP$. Column 5 displays the number of simplex pivot steps needed to solve each of the models. Finally, column 6 indicates the number of inequalities violated by the current solution. In particular, X_3 satisfies all constraints and hence is optimal for (6.32)–(6.34).

Some comments are in order concerning this example. First, observe that the linear programming models to be optimized always remain very small. As a consequence, solving them is a question of seconds (using the package LINDO on a PC 286). Moreover, the lower bound $e(X) + v(X)$ converges very fast to 10 (as a matter of fact, since we know that the optimal value of $e(X) + v(X)$ must be an integer, we could already have rounded up the lower bound 9.5, obtained after one iteration, to the value 10). Even more importantly, the matrix X_3 obtained in the last iteration of the procedure is in fact a 0–1 block diagonal matrix, and hence, is an optimal solution of the original 0–1

programming problem (6.32)–(6.35). It defines a partition of the machines into three groups $\{1, 6, 7\}$, $\{2, 5, 10\}$, $\{3, 4, 8, 9\}$, and of the parts into three corresponding families $\{1, 3, 4, 6, 14\}$, $\{2, 7, 11, 12\}$, $\{8, 9, 10, 13, 15\}$; part $\{5\}$ forms a cell by itself.

Thus, we have been able to solve to optimality the model (6.32)–(6.35), which involves 150 binary variables and about 19,000 constraints, by solving only 4 linear programming problems of very small size. Of course, as demonstrated by our third example below, we cannot expect to solve just as easily every instance of the block diagonalization problem. But the previous example gives at least some hope that model (6.32)–(6.35), possibly strengthened by additional valid inequalities, may be adequate for the solution to optimality of large scale instances. In order to further sustain this claim, we now proceed with our discussion of numerical examples.

The block diagonal approximation of A defined by X_3 features an optimal grouping efficiency of 0.933, and a grouping efficacy $\Gamma(A, X_3)$ equal to $(C - e(X_3))/(C + v(X_3)) = 43/53$. One may wonder whether another solution exists with better grouping efficacy. This question can be answered, as explained in Section 6.3, by solving the subproblem

$$\min (43/53)v(X) + e(X) \quad (6.36)$$

subject to (6.33)–(6.35) (see (6.30)). The continuous relaxation of this model can again be easily optimized by the cutting plane procedure, as shown in Table 6.4. The entries of this table can be interpreted in a similar fashion as those of Table 6.3. The column labelled LB gives here a lower bound on the optimal value of the objective function (6.36).

solu	# of ineq total/ binding	$e(X)$	$v(X)$	LB	# of pivots	# of violated ineq
Y_0	0/0	0	0	0.000	0	184
Y_1	100/71	7.5	2	9.123	73	89
Y_2	96/89	6	4	9.245	70	75
Y_3	114/114	6	4	9.245	74	0

Table 6.3 Performance of the cutting plane algorithm on Example 6.1; objective function (6.36).

Here again, the cutting plane procedure converges in four iterations to an optimal integer solution with 6 exceptions and 4 voids (this is in fact the same matrix X_3 that was already found when optimizing with respect to modified grouping efficiency). Thus, we conclude that 43/53 is the optimal value of the grouping efficacy criterion for Example 6.1, and that this value is attained by X_3 .

Example 6.2 The machine-part incidence matrix for this example is given in Table 6.5.

										1	1	1	1	1	1	1	1	1	1	2
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0
1		1	1					1	1		1		1	1		1	1		1	
2			1	1		1	1							1				1		1
3		1						1	1		1		1	1		1	1		1	
4			1	1		1	1			1								1		1
5	1				1	1				1		1			1		1			
6	1				1				1	1		1			1					1
7			1	1		1	1				1	1						1		1
8			1	1		1	1											1		1

Table 6.5 Data set 6.2 [Chandrasekharan and Rajagopalan, 1986]

As for the first data set, we first attempt to find a best block diagonalization of this matrix with respect to modified grouping efficiency. The cutting plane procedure is used to solve the continuous relaxation of (6.32)–(6.35) (since $M = 8$ and $P = 20$, there are 21,280 square inequalities in this formulation). It yields the results summarized in Table 6.6.

solu	# of ineq total/ binding	$e(X) + v(X)$	$\bar{\eta}$	# of pivots	# of violated ineq
X_0	0/0	$0 + 0 = 0$	1.000	0	401
X_1	100/74	$5 + 0 = 5$	0.969	66	299
X_2	99/88	$6 + 1 = 7$	0.956	73	198
X_3	113/81	$8 + 0 = 8$	0.950	72	155
X_4	106/87	$9 + 0 = 9$	0.944	74	160
X_5	112/109	$8 + 1 = 9$	0.944	77	64
X_6	134/133	$9 + 0 = 9$	0.944	73	0

Table 6.6 Performance of the cutting plane algorithm on Example 6.2;
objective function (6.32)

The last solution found, i.e. X_6 , is a 0 – 1 block diagonal matrix and hence is optimal with respect to modified grouping efficiency. It defines three machine groups, namely $\{1,3\}$, $\{2,4,7,8\}$ and $\{5,6\}$, and three corresponding family parts $\{2,8,9,11,13,14,16,17,19\}$, $\{3,4,6,7,18,20\}$ and $\{1,5,10,12,15\}$.

We ask now whether there is a better solution than X_6 with respect to grouping efficacy. The grouping efficacy of X_6 is equal to 52/61. Thus, we first attempt to solve the problem:

$$\min (52/61)v(X) + e(X) \tag{6.37}$$

subject to (6.33)–(6.35).

The cutting plane procedure applied to the relaxation of this problem yields the results in Table 6.7 (to be interpreted as Table 6.4).

solu	# of ineq total/ binding	$e(X)$	$v(X)$	LB	# of pivots	# of violated ineq
Y_0	0/0	0	0	0.000	0	401
Y_1	100/74	5	0	5.000	76	299
Y_2	99/99	5	2	6.705	88	211
Y_3	124/106	7	1	7.852	97	164
Y_4	131/103	8	1	8.852	90	181
Y_5	128/90	9	0	9.000	87	96
Y_6	115/103	9	0	9.000	90	94
Y_7	128/128	9	0	9.000	88	0

Table 6.7 Performance of the cutting plane algorithm on Example 6.2;
objective function (6.36).

The last solution found here, i.e. Y_7 , is identical to the optimal solution X_6 with respect to modified efficiency. Hence, this solution is also optimal with respect to grouping efficacy.

Example 6.3. The machine-part incidence matrix for this example is given in Table 6.8.

	1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
1	1																									
2		1																								
3			1																							
4				1																						
5					1																					
6						1																				
7							1																			
8								1																		
9									1																	
10										1																
11											1															
12												1														
13													1													

Table 6.8 Data set 6.3 [Leskowski et al., 1987]

Again the cutting plane procedure is used to solve the continuous relaxation of (6.32)–(6.35) in order to find a best block diagonalization of this matrix with respect to modified grouping efficiency. Since $M = 13$ and $P = 25$ we have to cope with 93,600 square inequalities in this formulation. After solving 30 linear programming problems of very small size (the maximum number of square inequalities involved in a model is 201) the algorithm stops. In this case, the final solution X_{30} satisfies all square inequalities indeed, but it is highly fractional, in fact: $X_{30} = 2/3A$. From this we conclude that for any block diagonalization of A the number of exceptions and voids is at least onethird of the number of non zeroes in A viz. 21. However, by making use of the special structure of this example, it can be proven that the minimal number of exceptions and voids is 33. This shows that for “bad” matrices with overlapping clusters as in this example, the gap between the optimal solution and the lower bound generated with our primitive cutting plane algorithm can be considerable, and one may wish to be able to improve the lower bound. This is possible by adding valid inequalities to the model (6.32)–(6.35), which are redundant for the 0–1 programming problem, but cut off fractional solutions as X_{30} from the continuous relaxation. In Chapter 7 we provide several classes of strong valid inequalities, some of which are severely violated by X_{30} .

6.5 Conclusions

In this chapter, we have briefly surveyed some of the previous attempts to attack the group technology cell formation problem, expressed as a block di-

agonalization problem, via mathematical programming techniques. We have argued that these attempts usually give rise to (mixed) integer programming models which provide weak bounds on the optimal value of the problem. We have proposed an alternative model which allows for the formulation of various constraints and objective functions, and provides stronger, nontrivial bounds on the optimal value of the problem. Finally, we have shown on some examples that this model may be adequate to solve instances of the block diagonalization problem *to optimality*.

Of course, the cutting plane algorithm presented in Section 6.4 is rather rudimentary. A more sophisticated cutting plane algorithm for the block diagonalization problem would first require to strengthen the formulation (6.23)–(6.24) by deriving additional classes of valid inequalities for this formulation (i.e., inequalities satisfied by all block diagonal matrices X). More precisely, we would like to give a partial description of the polyhedron whose extreme points are the 0 – 1 solutions of the system (6.23)–(6.24). Some special properties of this polytope together with a partial description of the polytope are discussed in Chapter 7 and Chapter 8.

Chapter 7

The Polytope of the Cell Formation Problem

This chapter is joint work with Yves Crama.
It will be published as an article in *Networks*.

A matrix X is called *block diagonal* if there exists a partition $R_1, \dots, R_k, R_{k+1}$ of its row-set and a partition $C_1, \dots, C_k, C_{k+1}$ of its column-set such that $x_{ij} \neq 0$ if and only if $i \in R_l$ and $j \in C_l$ for some $1 \leq l \leq k$. Now the block diagonalization problem can be formulated as follows: given an $m \times p$ incidence matrix A and a function $f(\cdot)$, find an $m \times p$ block diagonal incidence matrix X which minimizes $f(A, X)$.

This problem can be interpreted as an abstraction of the cell formation problem encountered in cellular manufacturing, as discussed in Chapter 6. Similar problems also arise in the analysis of large data arrays (e.g. for marketing or archeology applications), in production planning for flexible manufacturing systems, in sparse matrix computations, etc.

In this chapter, we discuss the facial structure of the polytope whose extreme points are exactly the $m \times p$ 0-1 block diagonal matrices. More precisely, some general properties of facet-defining inequalities for the polytope are stated, specific families of facet-inducing inequalities are described, and some lifting results are presented. The latter procedures guarantee that for every integer number z we can construct a facet defining inequality whose coefficients are relative primes, such that at least one of the coefficients has value z .

7.1 Introduction

In this chapter, we investigate the facial structure of the polytope whose extreme points are exactly the $m \times p$ 0-1 block diagonal matrices ($m, p \in \mathbb{N}$). More precisely, as in Chapter 6, we define a matrix X to be *block diagonal* if there exists a partition $R_1, \dots, R_k, R_{k+1}$ of its row-set and a partition $C_1, \dots, C_k, C_{k+1}$ of its column-set such that $x_{ij} \neq 0$ if and only if $i \in R_l$ and $j \in C_l$ for some $1 \leq l \leq k$ (notice that what we really mean is that X is block diagonal up to permutations of its rows and columns). We let

$$S_{mp} = \{ X \in \{0, 1\}^{m \times p} \mid X \text{ is block diagonal} \}$$

and we denote by Q_{mp} the convex hull of S_{mp} . The goal of this chapter is to provide a (partial) description of the polytope Q_{mp} by linear inequalities.

As explained in Chapter 6, our interest for the polytope Q_{mp} mainly stems from its relation to the cell formation problem encountered in cellular manufacturing. The data for this problem are generally assumed to be summarized in the *machine-part incidence matrix* A , where $a_{ij} = 1$ if part j needs to be processed on machine i , and $a_{ij} = 0$ otherwise. Recall that a

group technology cell consists of a number of machines (a machine-group) geared on the manufacturing of a number of similar parts (a part-family). The cell formation problem asks for a partition of the machines into machine-groups, a partition of the parts into part-families, and a matching between the machine-groups and the part-families which optimizes some measure of the inter- and intra-cell relationships. It can be abstracted into the following *block diagonalization problem*: given an $m \times p$ incidence matrix A and a function $f(.,.)$, find an $m \times p$ block diagonal incidence matrix X which minimizes $f(A, X)$ (the function $f(.,.)$ gives an estimate of the distance, or dissimilarity, between the original incidence matrix A and the 'ideal' cellularized system represented by X). In Chapter 6, we showed that, for many of the objective functions $f(.,.)$ proposed in the literature, the cell formation problem can be reduced to the problem of minimizing a linear function of the variables x_{ij} ($i = 1, \dots, m; j = 1, \dots, p$) over the polytope Q_{mp} . Similar block diagonalization problems also arise in the analysis of large data arrays (e.g. for marketing or archeology applications), in production planning for flexible manufacturing systems, in sparse matrix computations, etc (see Chapter 6 for references).

In our presentation, we will often rely on a graph-theoretic interpretation of block diagonal matrices and of the polytope Q_{mp} . We follow the graph-theoretic terminology of [Bondy and Murty, 1976]. Moreover, when $B = (U, V, E)$ is a bipartite graph and $G = (U, V, F)$ is a subgraph of B , we say that G is a *complete bipartite partitioning* of B if all connected components of G are complete bipartite (we look at isolated vertices as complete bipartite graphs). In particular, consider the complete bipartite graph $K_{mp} = (U_m, V_p, U_m \times V_p)$, where $U_m = \{u_1, \dots, u_m\}$ and $V_p = \{v_1, \dots, v_p\}$. We regard an arbitrary $m \times p$ 0-1 matrix X as the adjacency matrix of a subgraph G of K_{mp} , say $G = (U_m, V_p, F)$, where $(u_i, v_j) \in F$ if and only if $x_{ij} = 1$. It is easy to see that the matrix X is block diagonal if and only if its associated graph is a complete bipartite partitioning of K_{mp} .

This graph-theoretic interpretation stresses the analogy of the polytope Q_{mp} with the clique partitioning polytope P_n investigated by [Faigle et al, 1986] and [Grötschel and Wakabayashi, 1989], and with the related multiway cut polytope studied by [Chopra and Rao, 1991]. In fact, Q_{mp} can be viewed as the projection of P_{m+p} on some appropriate subspace. But this observation does not seem very useful in deriving a description of Q_{mp} from the results available about P_n ; for example, there is not a one to one relationship between (classes of) facet defining inequalities of these polytopes, (see also Chapter 2).

In Section 7.2, some general properties of facet defining inequalities for the polytope Q_{mp} are stated. In Section 7.3 specific families of facet inducing inequalities are described. Finally, section 7.4 contains some lifting theorems. In Chapter 8 a technique will be presented to patch facet defining inequalities into new valid inequalities which, under certain conditions, also define facets.

7.2 Properties of facet defining inequalities

We describe in this section some general properties of facet defining inequalities for the polytope Q_{mp} : two 'lifting' results, relating facets of lower-dimensional polytopes to facets of higher-dimensional ones, and one proposition describing the 'graphical' structure of facet-inducing inequalities.

In our discussion, it will be often convenient to consider the polytope associated with block diagonal submatrices of a given matrix, or equivalently, with complete bipartite partitionings of a given graph. To define these concepts more accurately, let $B = (U_m, V_p, E(B))$ be an arbitrary bipartite graph, where, as before, $U_m = \{u_1, \dots, u_m\}$ and $V_p = \{v_1, \dots, v_p\}$. The set of incidence matrices of complete bipartite partitionings of B is denoted by S_B , and the convex hull of S_B is denoted by Q_B . Clearly, if $B = K_{mp}$, then $S_B = S_{mp}$ and $Q_B = Q_{mp}$. In fact, the polyhedron Q_B can be viewed in the space $\mathbb{R}^{m \times p}$ as the face of Q_{mp} with the property that, for all $X \in Q_B$, $x_{ij} = 0$ when $(i, j) \notin E(B)$.

The dimension of Q_B is $|E(B)|$, since the subgraph of B containing no edges at all, as well as any subgraph of B containing only one edge of B , are complete bipartite partitionings of B . For the same reason the *trivial inequalities* $x_e \geq 0$ for $e \in E(B)$ are facet inducing for each Q_B . And so are the *trivial inequalities* $x_e \leq 1$, since all subgraphs of B containing exactly two edges of B are complete bipartite partitionings of B , too.

Suppose the following inequality is valid for S_B and defines a facet of Q_B :

$$\sum_{i,j:(i,j) \in B} \pi_{ij} x_{ij} \leq \pi_0 \quad (7.1)$$

The first properties of (7.1) we will discuss concern lifting procedures (see Chapter 3 for a study of lifting procedures). So, let B^+ be a bipartite graph on the same vertex set as B such that $E(B) \subset E(B^+)$. A lifting of (7.1) from Q_B to Q_{B^+} is an inequality valid for S_{B^+} of the form:

$$\sum_{i,j:(i,j) \in B^+} \pi_{ij} x_{ij} \leq \pi_0 \quad (7.2)$$

To start with, assume that B_1 and B_2 differ only in one edge. Our first result follows directly from the traditional sequential lifting procedure described in [Nemhauser and Wolsey, 1988] combined with the observation that Q_B is a facet of Q_{B^+} :

Proposition 7.2.1 *Consider inequality (7.1) and assume that it is valid for S_B and it defines a facet of Q_B . Let $E(B^+) = E(B) \cup \{(u_h, v_k)\}$. Then the inequality (7.2) is valid for S_{B^+} and defines a facet of Q_{B^+} if and only if*

$$\pi_{hk} = \pi_0 - \max \left\{ \sum_{i,j:(i,j) \in E(B)} \pi_{ij} x_{ij} \mid X \in S_{B^+} \text{ and } x_{hk} = 1 \right\}.$$

This proposition guarantees that, when a facet defining inequality is derived for a 'partial' polyhedron Q_B , this inequality can always be lifted to an inequality defining a facet of Q_{mp} .

The following proposition shows that an inequality defining a facet of Q_{mp} is also facet defining for each of the polyhedra corresponding to block diagonal matrices with at least m rows and p columns. It is similar in spirit to Theorem 3.3 in [Grötschel and Wakabayashi, 1990], Theorem 3.2 in [Chopra and Rao, 1991], and Theorem 2.2 in [Deza and Laurent, 1992]. Notice that no conditions are imposed on the specific form of the inequality.

Proposition 7.2.2 *Let $a, b \in \mathbb{N}$, $E(B) = U_m \times V_p$ and $E(B^+) = U_{m+a} \times V_{p+b}$. Consider inequality (7.1) and assume that it is valid for S_B and it defines a facet of Q_B . Then the inequality (7.1) also defines a facet of Q_{B^+} .*

Proof: It is obvious that (7.1) is valid for all $X \in S_{mp}$ if and only if it is valid for all $Y \in S_{m+a, p+b}$. To prove that the inequality defines a facet of Q_{B^+} , it is sufficient to show that the property holds for $a = 1$ and $b = 0$, since then the property holds for all $a, b \in \mathbb{N}$ by induction and by the symmetry of the vertices sets.

Because the inequality (7.1) is facet defining for Q_{mp} , there exists a set of $m \cdot p$ affinely independent solutions, all satisfying the inequality with equality. If $X = \underline{0}$ does not belong to this set, these solutions are also linearly independent. On the other hand, if $X = \underline{0}$ is in the set, then $\pi_0 = 0$, and hence (7.1) is a linear combination of the inequalities $x_e \geq 0$ with $(e \in U_m \times V_p)$, which are facet defining for all Q_B as mentioned earlier. This would imply that (7.1) is identical to an inequality $x_e \geq 0$ for a certain $e \in U_m \times V_p$.

So we may assume to have a set of $m \cdot p$ feasible and linearly independent elements X of S_B , all satisfying (7.1) at equality. This set can easily be

transformed into a set $\mathcal{Y}$ of $m \cdot p$ feasible and linearly independent elements Y of S_{B+} all satisfying (7.1) at equality, by setting $y_{m+1,j} = 0$ for all $j \in \{1, \dots, p\}$.

For the construction of p additional solutions, it is needed to select from $\mathcal{Y}$ a subset $\mathcal{Z}$ of p solutions such that the subsets of variables $\{z_{1j} \mid j \leq p\}$ of each of the p solutions $Z \in \mathcal{Z}$ are linearly independent. Determining such a set is equivalent to selecting a $(p \times p)$ nonsingular submatrix in a $(m \cdot p) \times p$ matrix with rank p . Each of the solutions in $\mathcal{Z}$ can be adapted in the following way. Let $Z = (z_{ij})$ be a solution from $\mathcal{Z}$, resulting in an adapted solution $\bar{Z} = (\bar{z}_{ij})$. Then $\bar{z}_{ij} = z_{ij}$ for all $i \leq m, j \leq p$, and $\bar{z}_{m+1,j} = z_{1j}$ for all $j \leq p$.

The set $\bar{\mathcal{Z}}$ of adapted solutions $\bar{Z}$ clearly is linearly independent, since the subsets of variables $\{\bar{z}_{m+1,j} \mid j \leq p\}$ of each of these p solutions in $\bar{\mathcal{Z}}$ are linearly independent. Now it is easy to see that the solutions in $\bar{\mathcal{Z}} \cup \mathcal{Y}$ form a set of $(m+1) \cdot p$ feasible and linearly independent elements $Y \in S_{m+1,p}$ all satisfying the equality (7.1) at equality. So inequality (7.1) is valid for $Q_{m+1,p}$, and defines a facet of $Q_{m+1,p}$. $\square$

Again consider now inequality (7.1) and assume that it defines a non-trivial facet of Q_B . Without any information about the numerical values of the coefficients (Π, π_0) , some general structural properties of the inequality can be stated. To do this, associate with the inequality two edge sets E and E^+ , defined as follows:

$$\begin{aligned} E &:= \{(u_i, v_j) \mid (u_i, v_j) \in B \text{ and } \pi_{ij} \neq 0\}, \\ E^+ &:= \{(u_i, v_j) \mid (u_i, v_j) \in B \text{ and } \pi_{ij} > 0\}. \end{aligned}$$

We call $H := (V(E), E)$ (respectively $H^+ := (V(E^+), E^+)$) the support (respectively the positive support) of the inequality (7.1).

Proposition 7.2.3 *If B is a nonempty bipartite graph, and (7.1) defines a non-trivial facet of Q_B , then:*

- (1) $\pi_0 > 0$;
- (2) E^+ is nonempty;
- (3) the support H of (7.1) is connected;
- (4) the positive support H^+ of (7.1) is connected;
- (5) $V(E) = V(E^+)$.

If moreover B is a complete bipartite graph, then:

(6) $E \setminus E^+$ is nonempty, i.e. Π has negative elements;

(7) the support H of (7.1) is two-connected.

Proof:

(1) Since the solution $X = \underline{0}$ is feasible, π_0 has to be nonnegative. As discussed in the proof of Proposition 7.2.2, all facet defining inequalities for which $\pi_0 = 0$ are trivial inequalities.

(2) Suppose E^+ is empty. Then inequality (7.1) is implied by the trivial inequalities $x_e \geq 0$, and therefore it cannot be facet defining.

(3) Suppose H is disconnected. Let H_1 be a connected component of H with vertex set W , and define H_2 to be the subgraph of H induced by $V(E) \setminus W$. It is easy to see that there exists a solution $\bar{X} \in S_B$ satisfying (7.1) at equality, such that $\bar{x}_{ij} = 0$ for all edges (u_i, v_j) with exactly one endpoint in W and one endpoint in $V(E) \setminus W$. Then define:

$$\pi_1 := \sum_{e \in E(H_1)} \pi_e \bar{x}_e, \quad \pi_2 := \sum_{e \in E(H_2)} \pi_e \bar{x}_e.$$

Clearly $\pi_0 = \pi_1 + \pi_2$. The inequality $\sum_{e \in E(H_1)} \pi_e x_e \leq \pi_1$ is a valid inequality for Q_B . Indeed, suppose $\sum_{e \in E(H_1)} \pi_e y_e > \pi_1$ for a solution $Y \in S_B$. Then from $\bar{X}$ and Y a solution $Z \in S_B$ could be constructed that violates (7.1). The same argument shows that the inequality $\sum_{e \in E(H_2)} \pi_e x_e \leq \pi_2$ is also valid. Since (7.1) is the sum of these two valid inequalities, it cannot induce a facet.

(4) Suppose H^+ is disconnected. Let H_1 be a connected component of H^+ with vertex set W , and define H_2 to be the subgraph of H^+ induced by $V(E^+) \setminus W$. Since H is connected and H^+ is a subgraph of H , there is an edge, say (u_i, v_j) , covering a vertex of H_1 as well as a vertex of H_2 , such that $\pi_{ij} < 0$.

Because inequality (7.1) is facet defining, there exists a solution $X \in S_B$ satisfying (7.1) at equality such that $x_{ij} = 1$ (otherwise the facet would be contained in the facet defined by the trivial inequality $x_{ij} \geq 0$). Now, the solution Y with $y_e = 0$ if the edge e covers a vertex of H_1 as well as a vertex of H_2 , and $y_e = x_e$ otherwise, is feasible, but violates (7.1). This contradicts the validity of the inequality of (7.1).

(5) Obviously $V(E^+) \subseteq V(E)$. Now suppose there exists a vertex $i \in V(E) \setminus V(E^+)$. Since H is connected, there exists a vertex $j \in V(E)$ such that $\pi_{ij} < 0$. Because inequality (7.1) is facet defining, there exists a solution $X \in S_B$ satisfying (7.1) at equality such that $x_{ij} = 1$. But, by the same reasoning as used to prove the former property, this contradicts the validity

of inequality (7.1).

(6) Suppose Π has no negative elements at all. The solution Y with $y_e = 1$ for all $e \in E(B)$ is feasible. Since (7.1) is facet defining, it easily follows that $\sum_{e \in E(B)} \pi_e = \pi_0$. Hence, (7.1) is a linear combination of the trivial inequalities $x_e \leq 1$, a contradiction.

(7) From Property (3) it follows that H is connected. If H is not two-connected, the vertex-set of H contains at least one cut-vertex, say h . This cut-vertex induces a split of H into two components H_1 and H_2 which have only vertex h in common and have no edges in common. There exists a solution $\bar{X} \in S_B$ satisfying (7.1) at equality. Define:

$$\pi_1 := \sum_{e \in E(H_1)} \pi_e \bar{x}_e, \quad \pi_2 := \sum_{e \in E(H_2)} \pi_e \bar{x}_e.$$

Clearly $\pi_0 = \pi_1 + \pi_2$.

The inequality $\sum_{e \in E(H_1)} \pi_e x_e \leq \pi_1$ is a valid inequality for Q_B . Indeed, suppose $\sum_{e \in E(H_1)} \pi_e y_e > \pi_1$ for a solution $Y \in S_B$. Now, a solution $Z \in S_B$ could be constructed from $\bar{X}$ and Y in the following way. Let

$$E(\bar{X}) = \{e \in E(H_2) \mid \bar{x}_e = 1\} \text{ and } E(Y) = \{e \in E(H_1) \mid y_e = 1\}.$$

Let F be a minimal edge-set such that $E(\bar{X}) \cup E(Y) \cup F$ is the edge-set of a complete bipartite partitioning of B (this is possible since B is complete bipartite). It is easy to see that any edge $e \in F$ has exactly one endpoint in each of H_1 and H_2 . Thus, $\pi_e = 0$ for all $e \in F$. Define now $Z \in S_B$ by: $z_e = 1$ if $e \in E(\bar{X}) \cup E(Y) \cup F$ and $z_e = 0$ otherwise. Then Z is feasible and violates (7.1), contradicting the validity of the inequality (7.1).

The same reasoning shows that the inequality $\sum_{e \in E(H_2)} \pi_e x_e \leq \pi_2$ is also valid for Q_B . Since (7.1) is the sum of two valid inequalities, it cannot induce a facet, a contradiction. $\square$

7.3 Facet Defining Inequalities

We present in this section various classes of inequalities defining facets for Q_{mp} . These inequalities will be obtained by lifting inequalities defining facets for a face Q_B of Q_{mp} (according to Proposition 7.2.1). Some of the subgraphs B which we will consider are 'squares', so-called 'spiked' K_{23} -free connected bipartite graphs, and cycles. For a given $X \in [0, 1]^{m \times p}$ and a given subgraph B of K_{mp} , we use the shorthand $x(B)$ to denote the sum $\sum_{i,j:(u_i,v_j) \in E(B)} x_{ij}$.

7.3.1 Square Inequalities

In Chapter 6 we observed that the *square inequalities*:

$$x_{hj} + x_{hk} + x_{ij} - x_{ik} \leq 2 \quad (h, i \in \{1, \dots, m\}, j, k \in \{1, \dots, p\}) \quad (7.3)$$

are valid for S_{mp} (hence, for its convex hull Q_{mp}), and that they yield, together with the integrality constraints on X , a valid description of S_{mp} ; that is,

$$S_{mp} = \{X \in \{0, 1\}^{mp} \mid x_{hj} + x_{hk} + x_{ij} - x_{ik} \leq 2 \\ \text{for all } h, i \in \{1, \dots, m\}, j, k \in \{1, \dots, p\}\}.$$

[Faigle et al, 1986] as well as [Grötschel and Wakabayashi, 1990] observed experimentally that the closely related 'triangle' inequalities, provide a tight description of the clique partitioning polytope. Here we prove:

Proposition 7.3.1 *The square inequalities define facets of Q_{mp} for all $m, p \geq 2$.*

Proof: The square inequality $x_{11} + x_{12} + x_{21} - x_{22} \leq 2$ defines a facet of Q_{22} , since it is satisfied at equality by the four affinely independent points $(1, 1, 0, 0)$, $(1, 0, 1, 0)$, $(0, 1, 1, 0)$ and $(1, 1, 1, 1)$. The statement follows now directly from Proposition 7.2.2. $\square$

Proposition 7.3.2 *Together, the trivial inequalities and the square inequalities provide a minimal complete description of Q_{2p} for all $p \geq 2$.*

Proof: Let P_{2p} be the polytope defined by the trivial and the square inequalities. To prove the proposition we have to show that $P_{2p} = Q_{2p}$. Then, since all trivial and square inequalities define a facet for Q_{2p} , the description is minimal. As mentioned before, the set of integral solutions in P_{2p} is exactly S_{2p} . So, to prove $P_{2p} = Q_{2p}$, it is sufficient to show that all extreme points of P_{2p} are integral.

Let X be an arbitrary extreme point of P_{2p} . Clearly, X satisfies at least $2p$ trivial and square inequalities at equality. Let (A, b) be a matrix of $2p$ linearly independent rows, corresponding to $2p$ trivial and square inequalities that are satisfied at equality by X such that the number of square inequalities involved, say α , is minimal. Notice that this implies that all trivial inequalities satisfied at equality by X are represented in (A, b) , and thus α equals the number of fractional elements of X .

Define a column k of X as follows: $(x_{1k}, x_{2k})^T$, for $k = 1, \dots, p$. Our next step is to show that at most one of the elements x_{1k} and x_{2k} can be fractional. Consider all square inequalities involving the elements of column k and represented in (A, b) . We partition these inequalities into two sets: one set consists of those inequalities involving the sum of the elements of column k , and the other set contains the inequalities involving the difference between the elements of column k . If both sets are nonempty, then we can find two equalities like the next: $x_{1k} + x_{2k} + x_{1h} - x_{2h} = 2$, and $x_{1k} - x_{2k} + x_{1i} + x_{2i} = 2$; together with the square inequality $x_{1h} - x_{2h} + x_{1i} + x_{2i} \leq 2$ this implies $x_{1k} = 1$. On the other hand, if at least one of the sets is empty, there is also at least one trivial inequality involving x_{1k} or x_{2k} satisfied at equality, since X is an extreme point of P_{2p} .

Clearly, if a square inequality is satisfied at equality by X , then either both columns involved in this inequality contain exactly one fractional component, or both columns are integral. In the latter case the inequality can not be represented in (A, b) , since it is linearly dependent with the corresponding trivial inequalities, which are represented in (A, b) too. Now, we can define a directed graph $G_{A,b} = (V, E)$ as follows:

- each fractional column x_k corresponds to a vertex $v_k \in V$, and vice versa;
- there is an arc from v_h to v_k in E if and only if there is a square inequality of the form $x_{1h} + x_{2h} \pm (x_{1k} - x_{2k}) \leq 2$ represented in (A, b) .

Clearly, since the cardinality of the vertex set equals the cardinality of the arc set (α) , the graph $G_{A,b}$ contains at least one (undirected) cycle, unless $\alpha = 0$. Furthermore, between each pair of vertices there can only be one arc. To see this, notice that two (not identical) square inequalities involving the same pair of columns determine two trivial inequalities satisfied at equality. By assumption, these trivial inequalities are represented in (A, b) . However, together the four equalities form a linear dependent system, so not all of them can be represented in (A, b) .

Consider the shortest cycle in $G_{A,b}$, say C . We will show now that if there is a vertex with indegree 1 in C , we can construct another set (A', b') of linearly independent inequalities satisfied by X at equality and involving α square inequalities, such that the corresponding graph $G_{A',b'}$ contains a cycle shorter than C .

Let v_j be a vertex with indegree 1 in C and let (v_i, v_j) and (v_j, v_k) be the arcs of the cycle incident with v_j . These arcs correspond to two square

inequalities satisfied at equality by X , say:

$$\begin{aligned}x_{1i} + x_{2i} + x_{1j} - x_{2j} &= 2, \text{ and} \\x_{1k} - x_{2k} + x_{1j} + x_{2j} &= 2.\end{aligned}$$

These equalities imply the following ones:

$$\begin{aligned}x_{1j} &= 1, \text{ and} \\x_{1k} - x_{2k} + x_{1i} + x_{2i} &= 2.\end{aligned}$$

Notice that the first of the latter two equalities corresponds to a trivial inequality already represented in (A, b) , and that the four equalities together form a minimal linearly dependent system. So, by adding to (A, b) a row representing the square inequality corresponding to the last equality and removing a row representing one of the square inequalities corresponding to the first two equalities, a new matrix (A', b') is constructed. The matrix contains $2p$ linearly independent rows, corresponding to $2p$ trivial and square inequalities satisfied at equality by X , such that the number of square inequalities involved is still α . Obviously, the shortest cycle in the graph $G_{A', b'}$ is shorter than C .

Assume now that there are no vertices with indegree 1 in C . Then, the length of this cycle is at least 4. Thus there exists a chain of arcs (v_h, v_i) , (v_j, v_i) , and (v_j, v_k) . These arcs correspond to three square inequalities satisfied at equality by X , say:

$$\begin{aligned}x_{1h} + x_{2h} + x_{1i} - x_{2i} &= 2, \\x_{1j} + x_{2j} + x_{1i} - x_{2i} &= 2, \text{ and} \\x_{1j} + x_{2j} + x_{1k} - x_{2k} &= 2.\end{aligned}$$

From these equalities, the following one follows:

$$x_{1h} + x_{2h} + x_{1k} - x_{2k} = 2.$$

Notice that these four equalities together form a minimal linearly dependent system. Therefore, the last equality cannot be represented in (A, b) yet, and the reader can verify easily that this implies that C cannot have length 4 either. Adding to (A, b) a row representing the square inequality corresponding to the last equality and removing a row representing one of the square inequalities corresponding to one of the three former equalities, yields a new matrix (A', b') . This matrix contains $2p$ linearly independent rows, corresponding to $2p$ trivial and square inequalities satisfied at equality

by X , such that the number of square inequalities involved is α . Obviously, the shortest cycle in the graph $G_{A',b'}$ is shorter than C .

Summarizing, for each matrix (A, b) corresponding to a minimal number of square inequalities α , with $\alpha > 0$, there is a graph $G_{A,b}$ containing at least one cycle. The shortest cycle in $G_{A,b}$ has at least length 5. However, we can always construct another matrix (A', b') inducing a graph $G_{A',b'}$ with a shorter shortest cycle, a contradiction. Thus, α must be 0, and this completes the proof. $\square$

7.3.2 Facet Defining Inequalities Based on Spiked K_{23} -free Connected Bipartite Graphs

The term 'spike' refers to special edges of a graph. A *spike-leaf* of graph B is a vertex covered by exactly one edge of B . The edge covering a spike-leaf is called a *spike*. A *spike-root* is a vertex covered by a spike, but not a spike-leaf itself. We will also say that an edge (resp. a vertex) is a spike (resp. a spike-leaf or a spike-root) of some valid inequality if it is a spike (resp. a spike-leaf or a spike-root) of the positive support H^+ of this inequality. For example, the square inequality $x_{hj} + x_{hk} + x_{ij} - x_{ik} \leq 2$ has two spike-leaves (vertices u_i and v_k), two spikes (the edges (u_h, v_k) and (u_i, v_j)) and two spike-roots (vertices u_h and v_j).

Notice that there are facet defining inequalities whose support consists of exactly one spike covering two spike-leaves, namely the trivial inequalities $x_{ij} \leq 1$. It follows from Proposition 7.2.3(4) in Section 7.2 that the support of nontrivial facet defining inequalities never contain spikes covering two spike-leaves.

The different natures of spike-roots and spike-leaves can be illustrated by the next two properties, of which the first one turns out to be especially useful in some proofs of Section 7.4 and Chapter 8.

Proposition 7.3.3 *Let the following inequality define a nontrivial facet for Q_{mp} :*

$$\sum_{i,j:(i,j) \in U_m \times V_p} \pi_{ij} x_{ij} \leq \pi_0 \quad (7.4)$$

Assume u_i is a spike-leaf and u_j is a spike-root with respect to (7.4). Then:

- (a) *there is a complete bipartite partitioning A of K_{mp} with a component consisting of u_i only, such that the corresponding characteristic vector X^A satisfies (7.4) at equality;*
- (b) *there is no complete bipartite partitioning A of K_{mp} with a component consisting of u_j only, such that the corresponding characteristic vector X^A satisfies (7.4) at equality.*

Proof: We denote by A (or A') a complete bipartite partitioning of K_{mp} , and by X^A (or $X^{A'}$) the corresponding incidence matrix.

(a) Assume without loss of generality that (u_i, v_i) is a spike, i.e. $\pi_{ii} > 0$. There is a bipartite clique partitioning A' of K_{mp} such that $X^{A'}$ satisfies (7.4) at equality, and u_i and v_i are not in the same component. To see this, observe that otherwise the characteristic vectors of all bipartite clique partitionings of K_{mp} satisfying the inequality (7.4) at equality also satisfy the trivial inequality $x_{ii} \leq 1$ at equality; this would imply that the facets defined by these inequalities are identical, which is not the case by assumption. Consider the component C of A' containing the vertex u_i . Now we construct a bipartite clique partitioning A identical to A' except for component C , which is splitted into two components C_1 , consisting of u_i , and C_2 , the complete bipartite subgraph induced by all vertices of C except u_i . Since $\pi_{ii} > 0$, and u_i is a spike leaf, $\pi_{il} \leq 0$ for all $l \in \{1 \dots k\} \setminus \{i\}$, and we deduce that X^A satisfies inequality (7.4) at equality:

$$\begin{aligned} \pi_0 &= \sum_{i,j:(i,j) \in E(A')} \pi_{ij} = \\ &= \sum_{i,j:(i,j) \in E(A)} \pi_{ij} + \sum_{l:v_l \in C_2} \pi_{il} \leq \sum_{i,j:(i,j) \in E(A)} \pi_{ij} \leq \pi_0 \end{aligned}$$

(b) Suppose there is a complete bipartite partitioning A of K_{mp} , such that X^A satisfies inequality (7.4) at equality, and such that u_j is a component of A . There is at least one spike covering u_j and a spike-leaf, say v_j . We may assume that v_j is also a component of A , using the same argument as in the proof of (a). Now we construct a bipartite clique partitioning A' identical to A up to the component consisting of u_j and the component consisting of v_j , which are merged into one component. Since $\pi_{jj} > 0$, we deduce a contradiction: $\pi_0 \geq \sum_{i,j:(i,j) \in E(A')} \pi_{ij} = \sum_{i,j:(i,j) \in E(A)} \pi_{ij} + \pi_{jj} > \pi_0$. $\square$

We say that a graph B is *spiked* if each vertex of B is covered by exactly one spike, that is, if the spikes form a perfect matching of B . The general structure of a spiked graph is shown in Figure 7.1; see also Figures 7.2 and

7.3 for special types of spiked graphs.

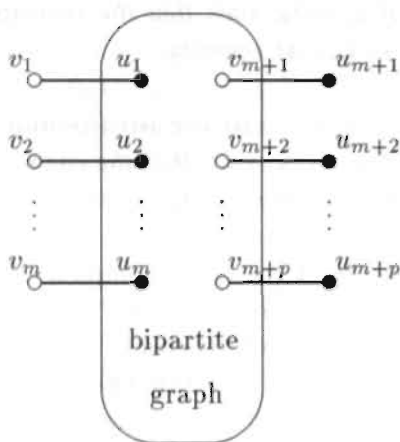


Figure 7.1 A Spiked Graph

We say that B is a K_{23} -free graph if it does not contain any subgraph isomorphic to K_{23} . The following holds:

Proposition 7.3.4 *For $k \in \mathbb{N}$, if B is a spiked K_{23} -free connected bipartite graph with exactly k spikes, then the inequality $x(B) \leq k$ defines a facet of Q_B .*

Proof: Observe that, when $k = 1$, the inequality $x(B) \leq k$ boils down to one of the trivial inequalities $x_e \leq 1$ and the proposition certainly holds. For the rest of the proof we assume that $k \geq 2$. First the implicit statement that the inequality is valid for Q_B will be justified. This will be done by showing that, for any feasible solution $X \in S_B$, the number of edges contained in any component of X cannot exceed the number of spike-roots of this component. From this it follows that $x(B)$ cannot exceed the total number of spike-roots of B . Because the graph B is spiked, each vertex of B is covered by exactly one spike, which implies that the number of spike-roots is equal to the number of spikes. Hence the inequality $x(B) \leq k$ is valid.

Since B does not contain any subgraph K_{23} , each component in a bipartite clique partitioning of B is either a cycle of length four, or contains at most one element from either U_m or V_p . Clearly, if the component is a cycle of length four, then all vertices are spike-roots, and the number of edges equals the number of vertices. So, assume that the component contains exactly one element of V_p , say v_0 . (A similar reasoning holds if the component contains only one element of U_m .) If v_0 is a spike-leaf, then the component

contains at most one edge, covering v_0 and one spike-root (since $k \geq 2$ a spike cannot cover two spike-leaves). If, on the other hand, v_0 is a spike-root, then the component can contain several elements of U_m . At most one of these elements can be a spike-leaf (otherwise v_0 would be covered by two spikes), and the others are bound to be spike-roots. Thus, in both cases, the number of edges in the component does not exceed the number of spike-roots in the component.

Now we prove that the inequality $x(B) \leq k$ is also facet defining. Since there exist complete bipartite partitionings satisfying the inequality at equality, the inequality definitely induces a face of Q_B , say $F := \{X \in Q_B \mid x(B) = k\}$. This face is contained in a facet of Q_B , say F_β , which in its turn is induced by an inequality $\sum_{i,j:(i,j) \in U_m \times V_p} \beta_{ij} x_{ij} \leq \beta_0$. To prove the proposition, it is sufficient to show that the coefficients of the inequalities are identical, up to multiplication by a strictly positive scalar.

Assume that the vertices of B are numbered in such a way that vertex u_i and vertex v_i are covered by the same spike. Let S be the set of spikes, i.e. $S := \{(u_i, v_i) \mid i \in \{1 \dots k\}\}$. Then, for all $i \neq j$ such that $(u_i, v_j) \in E(B)$, define the next two edge sets: $A_{ij} := \{(u_i, v_j)\} \cup S \setminus \{(u_i, v_i)\}$, and $A_{ji} := \{(u_i, v_j)\} \cup S \setminus \{(u_j, v_j)\}$.

Since the edge sets S and A_{ij} all define complete bipartite partitionings containing k edges, the following holds (where X^A denotes the incidence matrix of the edge-set A): for all $(u_i, v_j) \in E(B)$, $X^S, X^{A_{ij}}, X^{A_{ji}} \in F \subseteq F_\beta$. Comparing the edges sets, it follows that $\beta_{ii} = \beta_{ij} = \beta_{jj}$. Together with the connectedness of B , this easily implies that $\beta_{ij} = \beta_{11}$ for all $(u_i, v_j) \in E(B)$. So, $\beta_0 = \beta_{11}x(B) = \beta_{11}k$. This completes the proof. $\square$

In view of Propositions 7.2.1 and 7.2.2, the inequality $x(B) \leq k$ defined in Proposition 7.3.4 can be lifted to a family of facet defining inequalities of Q_{mp} , for all $m, p \in k$. A subset of this family can be described explicitly. To achieve this, a new definition is needed: a subset C of $(U_m \times V_p) \setminus E(B)$ is called a *chord set* for a spiked tree B if, for any path between two spike-leaves of B , there is an edge in C containing two (arbitrary) vertices of the path. The following proposition holds:

Proposition 7.3.5 *For $k \in \mathbb{N}$ and $m, p \geq k$, if B is a spiked tree with exactly k spikes and C is a minimal chord set for B , then the inequality $x(B) - x(C) \leq k$ defines a facet of Q_{mp} .*

Proof: It is sufficient to justify the proposition for $m = n = k$, since it follows then from Proposition 7.2.2 that the proposition holds for $m, n \geq k$.

First, the implicit statement that the inequality is valid for Q_{mp} will be demonstrated. This will be done by showing that, for any feasible solution $X \in S_{kk}$, and for each component X_0 of X , the sum of the coefficients corresponding to the edges in X_0 cannot exceed the number of spike-roots in X_0 .

Suppose X_0 contains r spike-roots and s spike-leaves of B . We may assume that the intersection B_0 of X_0 with B is connected, otherwise we could split X_0 , omitting only edges with nonpositive coefficients, and prove our statement for each of the smaller components. This implies in particular that, if a certain spike-leaf is in X_0 , then the corresponding spike is also in X_0 , and thus, $s \leq r$.

Using induction we can show that, by definition of the chord-set C for B , there are bound to be at least $s - 1$ edges of C in X_0 . Clearly, the statement is true for $s = 0, 1$. Suppose it is true for $s = ss$, is it true for $s = ss + 1$? Let C_0 be the intersection of the component X_0 with C . In view of the definition of C , and since $s \geq 2$, C_0 is nonempty, so let $(l_1, l_2) \in C_0$. Then there is a path P in B_0 from l_1 to l_2 . Removing an edge of P from B_0 yields two trees, say B_1 and B_2 . B_1 contains r_1 spike-roots and s_1 spike-leaves, and B_2 contains r_2 spike-roots and s_2 spike-leaves. Let C_1 be the subset of C connecting vertices of B_1 , and let C_2 be the subset of C connecting vertices of B_2 . We may assume without loss of generality that $s_1, s_2 \leq s - 1$ (otherwise s_1 or s_2 must be 0, and we reduce the analysis to a smaller case). We deduce by induction $|C_1| \geq s_1 - 1$ and $|C_2| \geq s_2 - 1$. Combining this with the observations that $|C| \geq |C_1| + |C_2| + 1$ and $s = s_1 + s_2$, we obtain $|C| \geq s - 1$.

Since B is a spiked tree, B_0 is also a tree containing at most $r + s - 1$ edges. As shown above there are at least $s - 1$ edges of C in the component. Hence the sum of the coefficients of the edges in the component X_0 is at most $r + s - 1 - (s - 1) = r$, and the inequality is valid. Now we prove that the inequality $x(B) - x(C) \leq k$ also defines a facet. Since several bipartite partitionings exist that satisfy the inequality at equality, the inequality definitely induces a face of Q_{kk} , say $F := \{X \in Q_{kk} \mid x(B) - x(C) = k\}$. This face is contained in a facet of Q_{kk} , say F_β , defined by an inequality $\sum_{i,j:(i,j) \in U_m \times V_p} \beta_{ij} x_{ij} \leq \beta_0$. To prove the proposition, it is sufficient to show that the two inequalities mentioned here do not differ apart from multiplication by a strictly positive coefficient.

Assume that the vertices of B are numbered in such a way that vertex u_i and vertex v_i are covered by the same spike. From Proposition 7.3.2 it follows that there is a $c \in \mathbb{R}^+$ such that $\beta_{ij} = c$ for all edges $(u_i, v_j) \in B$,

and $\beta_0 = kc$.

The next step will be to show that $\beta_{ij} = 0$ for all edges $(u_i, v_j) \notin B \cup C$, with $u_i \in U_k$ and $v_j \in V_k$. First, assume that at least one of the vertices, say without loss of generality u_i , is a spike-leaf of B . This implies that there is an $h \in \{1 \dots k\} \setminus \{i\}$ such that $(u_h, v_i) \in B$. Now, let S be the set of spikes: $S := \{(u_l, v_l) \mid l \in \{1 \dots k\}\}$. Define the next two edge sets: $A_{ih} := \{(u_h, v_i)\} \cup S \setminus \{(u_i, v_i)\}$, and $D_{ij} := \{(u_i, v_j)\} \cup A_{ih}$. Notice that u_h cannot be incident with v_j , since in that case u_i and v_j would both be spike-leaves and the edge (u_i, v_j) would be the only chord of the path between u_i and v_j in B , implying $(u_i, v_j) \in C$ in contradiction with the assumptions.

So, the edge sets A_{ih} and D_{ij} define feasible partitionings of Q_{kk} , and the corresponding characteristic vectors belong to the facets F and F_β , since they satisfy $x(B) - x(C) = k$. Since D_{ij} is an extension of A_{ih} with the edge (u_i, v_j) , the coefficient β_{ij} must have the value 0.

When u_i and v_j both are spike-roots with respect to B , it is somewhat more involved to prove that $\beta_{ij} = 0$. Here induction is used and a new definition is needed: an edge $(u_h, v_l) \in (U_m \times V_p) \setminus E(B)$ is called a q -chord for a spiked tree B if there exists a path of length q between u_h and v_l in B . Obviously, since B is a bipartite graph, q must be odd. Assume now that $\beta_{hl} = 0$ holds for all q -chords $(u_h, v_l) \notin C$ with $q \leq z$ (this is e.g. vacuously the case when $z = 1$). Then we want to show that $\beta_{ij} = 0$ holds for $(z + 2)$ -chords $(u_i, v_j) \notin C$.

Two cases can occur. The first case is that $(u_j, v_i) \notin C$. Then define: $D_{ij} := \{(u_i, v_j), (u_j, v_i)\} \cup S$. Again the edge sets S and D_{ij} define feasible partitionings of Q_{kk} , and their corresponding characteristic vectors satisfy $x(B) - x(C) = k$. Because D_{ij} is an extension of S with the edges (u_i, v_j) and (u_j, v_i) , the following holds: $\beta_{ij} = -\beta_{ji}$. Since (u_j, v_i) is an edge covering two spike-leaves, it has already been determined that the coefficient β_{ji} has value 0, so $\beta_{ij} = 0$.

The second case is when $(u_j, v_i) \in C$. In this case there is a path P between the spike-leaves v_i and u_j of which (u_j, v_i) is the unique chord in C (because C is minimal). Let $P_i := E(P) \setminus \{(u_i, v_i)\}$, and let $\text{Clique}(P_i)$ denote the complete bipartite edge-set induced by the vertices of P_i . Then the edge set D_{ij} is defined in the following way: $D_{ij} := \text{Clique}(P_i) \cup \{(u_h, v_h) \mid u_h \notin P \text{ and } v_h \notin P\}$. Then comparing again S and D_{ij} as earlier, it can be deduced that $\beta_{ij} = 0$. This completes the induction.

The final step will be to show that also $\beta_{ij} = -c$ for all edges $(u_i, v_j) \in C$. Observe that there is a path, say P , between the spike-leaf of spike i and the spike-leaf of spike j , and (u_i, v_j) is the unique chord of P in C . Then define:

$D_{ij} := \text{Clique}(P) \cup \{(u_h, v_h) \mid u_h \notin P \text{ and } v_h \notin P\}$. Comparing S and D_{ij} , it can be easily deduced that in this last case $\beta_{ij} = -c$. This completes the proof of the proposition. $\square$

To get better acquainted with spiked tree inequalities, consider for instance the special case in which the spike-roots $u_2, u_3, \dots, u_k$ of the spiked tree B are all adjacent to the spike-root v_1 , as shown in Figure 7.2 below. A minimal chord set C for this tree must consist of the following edges: (u_1, v_j) , and either (u_i, v_j) or (u_j, v_i) , for all $i, j \geq 2$.

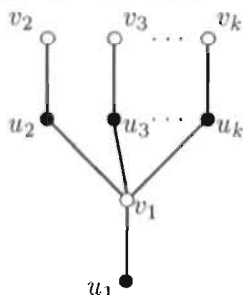


Figure 7.2 A Spiked Tree

For this example, if the coefficients of the inequality $x(B) - x(C) \leq k$ are organized in a $k \times k$ matrix whose element (i, j) is π_{ij} , then the coefficients in the first column and the diagonal elements correspond to the edges of B , and so they all have value 1. All elements in the first row, the first one excepted, correspond to the chords (u_1, v_i) and therefore have value -1 . For the remaining elements holds that either element (i, j) has value 0 and element (j, i) has value -1 , or vice versa. For example,

$$\begin{bmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & -1 & \cdots & -1 \\ 1 & -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & 0 & -1 & \cdots & 1 \end{bmatrix}$$

Carrying out this construction with $k = 1$ or $k = 2$ demonstrates that the trivial inequalities $x_{ij} \leq 1$ and the square inequalities belong to the family of spiked tree inequalities.

Another subset of the family of facet defining inequalities based on K_{23} -free connected bipartite graphs can also be described explicitly as follows. Let a *spiked cycle* be a spiked graph whose spike-roots induce a cycle (see

Figure 7.3; notice that this is a slight abuse of our general definition of a spiked graph, since a spiked cycle is not a cycle; but this abuse is convenient, and will not cause any confusion).

A subset C of $(U_m \times V_p) \setminus E(B)$ is a *chord set* for the spiked cycle B if, for each pair (s, t) of spike-leaves of B , and for each of the two paths $P_i (i = 1, 2)$ between s and t , there is an edge $e(s, t, P_i) \in C$ such that:

- (a) $e(s, t, P_i)$ links two (arbitrary) vertices of P_i ($i = 1, 2$);
- (b) $e(s, t, P_1)$ and $e(s, t, P_2)$ are distinct;
- (c) if one of the leaves s or t is covered by both edges $e(s, t, P_1)$ and $e(s, t, P_2)$, then one of these edges covers both s and t .

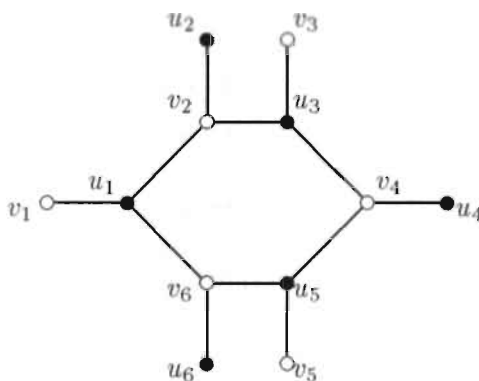


Figure 7.3 A Spiked Cycle

Proposition 7.3.6 For all $m, p \geq k \geq 4$, if B is a spiked cycle with exactly k spikes and C is a minimal chord set for B , then the inequality $x(B) - x(C) \leq k$ defines a facet of Q_{mp} .

Proof: Once we know that the inequality is valid for Q_{kk} , then proving that it defines a facet of Q_{kk} can be done as for the inequality in Proposition 7.3.5. Thus we concentrate on establishing validity. This will be done by showing that, for any feasible solution X and for each component X_0 of the graph associated with X , the sum of the coefficients corresponding to the edges in X_0 cannot exceed the number of spike-roots in X_0 .

Assume that X_0 contains r roots and s spikes of B . Either X_0 contains the whole cycle or not. In the latter case, we can conclude as in the proof of Proposition 7.3.5, since each component of $E(X_0) \cap E(B)$ is a spiked tree.

So suppose that the component contains the whole cycle (though not necessarily all the spikes). Obviously the number of edges with a positive coefficient contained in X_0 is $k + s$. When X_0 does not contain any spikes (i.e. $s = 0$), the number of edges with a positive coefficient is precisely the number of roots. When X_0 contains at least two spikes ($s \geq 2$), then the number of edges in X_0 which are also in the chord set C is at least s (since there is at least one edge of C between any two successive spike-leaves contained in X_0). Finally, consider the case where the component contains only one spike-leaf ($s = 1$), say u_i . Then the number of edges in C is at least one. To see this, notice that there is at least one other spike-leaf in U_m , say u_j . Now, by the construction of the chord set for a spiked cycle, there are distinct paths P_1 and P_2 in B between u_i and u_j , and two corresponding edges e_1 and e_2 in C . If one of these edges covers u_i , then this edge is in the component. Suppose on the other hand that neither e_1 nor e_2 covers u_i . Since u_j is covered by at most one of e_1 and e_2 , at least one of the edges e_1 , e_2 links two spike-roots, and this edge is in the component. $\square$

A small example of a K_{23} -free spiked cycle is shown in Figure 7.3. Call this graph B_6 . It is easy to see that all of the edges (u_2, v_1) , (u_2, v_3) , (u_4, v_3) , (u_4, v_5) , (u_6, v_5) and (u_6, v_1) must be in any chord set for B_6 , but on the other hand, there exist various ways to complete this list to a minimal chord set. As a matter of fact, it can be checked that each of the matrices Π shown hereunder gives rise to a facet defining inequality for $Q_{6,6}$, derived from B_6 as explained in Proposition 7.3.5:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 1 & -1 & -1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & -1 \\ 0 & 0 & -1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 1 \\ -1 & 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 & -1 & -1 \\ 0 & 0 & -1 & 1 & -1 & -1 \\ -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 0 & -1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 1 & 1 \\ -1 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}.$$

The question arises whether it is possible to describe explicitly other, possibly more general, subfamilies of facet defining inequalities based on K_{23} -free

connected bipartite graphs than those given in Propositions 7.3.5 and 7.3.6. In Chapter 8, we present a patching procedure which partially answers this question.

7.3.3 Facet Defining Inequalities Based on Cycles

Let C_k be a cycle of length k , with k even. If $k \geq 6$, then no component of a complete bipartite partitioning of C_k can contain more than two edges. Therefore, the total number of edges of a complete bipartite partitioning of C_k cannot exceed $2k/3$. If k is not a multiple of three, then one easily shows that the inequality $x(C_k) \leq \lfloor \frac{2}{3}k \rfloor$ is facet defining for Q_{C_k} .

Define now T_k to be the graph induced by the 3-chords of C_k (a q -chord of C_k is an edge joining two vertices at distance q in C_k). Then the following holds:

Proposition 7.3.7 *For all $k \geq 6$, the inequality $x(C_k) - x(T_k) \leq \lfloor \frac{2}{3}k \rfloor$ is valid for Q_{kk} . If $k = 4 + 6t$ for $t \in \mathbb{N}$, then the inequality induces a facet of Q_{kk} .*

Proof: First we show that the inequality is valid for Q_{kk} . If $k \geq 8$, then for each 3-chord $e \in T_k$ there is a unique edge-set $\{e_1, e_2, e_3\} \subset C_k$ such that $\{e\} \cup \{e_1, e_2, e_3\}$ induces a cycle of length 4. Using this fact, we can associate with each element e of T_k a unique square inequality of the form: $x_{e_1} + x_{e_2} + x_{e_3} - x_e \leq 2$. Notice that for each edge of C_k , the corresponding variable occurs in exactly three of these inequalities. So, adding up all inequalities related to the edges in T_k yields:

$$3x(C_k) - x(T_k) \leq 2k \Rightarrow x(C_k) - x(T_k) \leq \frac{2}{3}k.$$

Since the lefthand side is integral for all $X \in S_{kk}$, we may round down the righthand side. Consider now the case $k = 6$. For each 3-chord $e \in T_6$ there are two disjoint edge-sets which, together with e , induce cycles of length 4. Adding up all corresponding inequalities yields:

$$3x(C_6) - 2x(T_6) \leq 12 \Rightarrow x(C_6) - x(T_6) \leq 4.$$

Now we prove the second part of the proposition, namely: the inequality $x(C_k) - x(T_k) \leq \lfloor \frac{2}{3}k \rfloor$ induces a facet of Q_{kk} , if $k = 4 + 6t$ for some natural number t . Since there is a bipartite partitioning satisfying the inequality at equality, the inequality induces a face of Q_{kk} , say F . This face is contained in a facet of Q_{kk} , say F_β , defined by an inequality $\sum_{i,j:(ij) \in U_m \times V_p} \beta_{ij} x_{ij} \leq \beta_0$.

We must show that this inequality is identical to $x(C_k) - x(T_k) \leq \lfloor \frac{2}{3}k \rfloor$ up to multiplication by a strictly positive scalar, say c .

First we will show $\beta_{ij} = c > 0$ for all indices i, j such that $(u_i, v_j) \in C_k$. Since C_k is a cycle, C_k contains two edges covering v_j , (u_i, v_j) and, say, (u_h, v_j) . Let A_i be the unique edge-set inducing a bipartite partitioning of C_k , and containing precisely $\lfloor \frac{2}{3}k \rfloor = 4t + 2$ edges, but not containing any edges covering u_i . Define the edge-set D_{ij} as follows: $D_{ij} := \{(u_i, v_j)\} \cup A_i \setminus \{(u_h, v_j)\}$. Clearly, both A_i and D_{ij} are bipartite clique partitionings of K_{kk} , and the corresponding characteristic vectors belong to F as well as F_β . Comparing the edge-sets A_i and D_{ij} it follows that $\beta_{ij} = \beta_{hj}$. Iterating this argument, we conclude that all coefficients β_{ij} corresponding to edges (u_i, v_j) of C_k have the same value, say c . From Proposition 7.2.3(1) and the observation that $\beta_0 = (4t + 2)c$, we deduce that c is strictly positive.

The second step is to show that $\beta_{ij} = -c$ for all indices i, j such that $(u_i, v_j) \in T_k$. Since (u_i, v_j) is a 3-chord of C_k , C_k contains a path of length 3 with u_i and v_j as endpoints, say $\{(u_i, v_l), (u_h, v_l), (u_h, v_j)\}$. Now define $D_{ij} := A_i \cup \{(u_i, v_l), (u_i, v_j)\}$ (where A_i is as above). Clearly, D_{ij} is a bipartite clique partitioning of K_{kk} , and the corresponding characteristic vector is in F and F_β . Comparing A_i with D_{ij} , it follows that $\beta_{il} + \beta_{ij} = 0$, and thus $\beta_{ij} = -c$.

The final step is to show that the remaining coefficients β_{ij} have value 0. We distinguish three different cases. The first case is when (u_i, v_j) is a $(5 + 6s)$ -chord, where $s \in \{0, \dots, \lfloor \frac{1}{2}(t - 1) \rfloor\}$. The reader can verify that in this case the two edges of C_k covering v_j both belong to A_i . So, $D_{ij} := A_i \cup \{(u_i, v_j)\}$ is a bipartite clique partitioning of K_{kk} , and the corresponding characteristic vector is in F and F_β . Then, comparing A_i and D_{ij} , it follows that $\beta_{ij} = 0$.

The second case is when (u_i, v_j) is a $(7 + 6s)$ -chord, where $s \in \{0, \dots, \lfloor \frac{1}{6}(3t - 5) \rfloor\}$. Then C_k contains a path P of length $7 + 6s$ with u_i and v_j as endpoints, and there are two edges of C_k of the form (u_i, v_r) and (u_g, v_r) which do not belong to this path. Consider the edge-set $D_{gr} := \{(u_g, v_r)\} \cup A_g \setminus \{(u_i, v_r)\}$ (as discussed in step 1). The path P contains three edges of the form (u_i, v_l) , (u_h, v_n) , and (u_h, v_j) , which induce two components of the bipartite clique partitioning D_{gr} . Now we define an edge-set as follows: $D_{ij} := D_{gr} \cup \{(u_i, v_n), (u_i, v_j), (u_h, v_l)\}$. Then D_{ij} is a bipartite clique partitioning of K_{kk} , and the corresponding characteristic vector is in F and F_β . Noticing that D_{ij} is an extension of D_{gr} by two $(5 + 6s)$ -chords and one $(7 + 6s)$ -chord (u_i, v_j) , we conclude that $\beta_{ij} = 0$.

The third and last case is when (u_i, v_j) is a $(9 + 6s)$ -chord, where $s \in$

$\{0, \dots, \lfloor \frac{1}{6}(3t-7) \rfloor\}$. The cycle C_k contains a path of length $9+6s$ with u_i and v_j as endpoints, and containing the edges (u_h, v_l) , and (u_h, v_j) . Notice that these two edges correspond to one component of the bipartite clique partitioning A_i . Then the edge-set $D_{ij} := A_i \cup \{(u_i, v_l), (u_i, v_j)\}$ is a bipartite clique partitioning of K_{kk} , and the corresponding characteristic vector is in F and F_β . Since D_{ij} is an extension of A_i by one $(7+6s)$ -chord and one $(9+6s)$ -chord, respectively (u_i, v_l) and (u_i, v_j) , we conclude that $\beta_{ij} = 0$.

This completes the proof. $\square$

7.4 Lifting Theorems

In this section, we are going to present various lifting results, which allow to derive facet defining inequalities for $Q_{m+1,p+1}$, from facet defining inequalities for $Q_{m,p}$. To be more specific, assume that the following inequality is valid for S_{mp} and defines a facet for Q_{mp} :

$$\sum_{i=1}^m \sum_{j=1}^p \pi_{ij} x_{ij} \leq \pi_0 \quad (7.5)$$

We will discuss conditions on the coefficients of inequalities of the following form, that guarantee these inequalities to be valid or facet defining for $Q_{m+1,p+1}$:

$$\sum_{i=1}^{m+1} \sum_{j=1}^{p+1} \pi_{ij} x_{ij} \leq \pi_0 + \pi_{m+1,p+1} \quad (7.6)$$

First we need to introduce some concepts which will be useful in defining the coefficients of the lifted inequalities. A *V-sequence* is a (nonempty) ordered subset of V_p (a *U-sequence* would be similarly defined). Define for each subset I of V_p the set $S_{mp}(I)$ to be the set of solutions in S_{mp} that satisfy $x_{ij} = 0$ for all $i \in \{1, \dots, m\}$ and $j \in I$. These are the solutions for which the nodes in the set I are components. Consider inequality (7.5) and let $\Phi = (v_1, \dots, v_r)$ be a (arbitrary) *V-sequence*. The *sequential covering* of a vertex v_h (for $h = 1, \dots, r$, with respect to Π and Φ), denoted $c[v_h]$, is defined recursively as follows:

$$c[v_h] = \min_{I: I \subseteq \{1, \dots, h-1\}} \left\{ a_0 - \sum_{j \in I} c[v_j] \right. \\ \left. - \max \left\{ \sum_{i=1}^m \sum_{j=1}^p \pi_{ij} x_{ij} \mid x \in S_{mp}(I \cup \{h\}) \right\} \right\}.$$

The reader can convince himself that, regardless of the sequence, the sequential covering of an arbitrary vertex is always nonnegative.

Finally, call the *U-extension of K_{mp} with respect to a V -sequence Φ* the graph B obtained from K_{mp} by adding a vertex u_{m+1} to U_m , a vertex v_{p+1}

to V_p , and the edge (u_{m+1}, v_{p+1}) to the edge set, as well as all the edges between u_{m+1} and the vertices of the V -sequence Φ ; that is,

$$B := (U_m \cup \{u_{m+1}\}, V_p \cup \{v_{p+1}\}, \\ (U_m \times V_p) \cup \{(u_{m+1}, v_h) \mid h \in \{1, \dots, r\}\} \cup \{(u_{m+1}, v_{p+1})\}).$$

The following statement describes how a valid inequality like (7.5) for Q_{mp} can be lifted to a new valid inequality of the form (7.6) for Q_B :

Proposition 7.4.1 *Let inequality (7.5) be valid for Q_{mp} , let $\Phi = (v_1, \dots, v_r)$ be a V -sequence, and assume that the coefficients of inequality (7.6) satisfy:*

$$\pi_{ij} = \begin{cases} c[v_j] & \text{if } i = m+1 \text{ and } j \in \{1, \dots, r\}; \\ \sum_{j=1}^r c[v_j] & \text{if } i = m+1 \text{ and } j = p+1; \end{cases}$$

Then inequality (7.6) is valid for Q_B , where B is the U -extension of K_{mp} with respect to Φ .

Proof: By validity of inequality (7.5) and $\sum_{j=1}^r \pi_{m+1,j} = \pi_{m+1,p+1}$ (by construction), the inequality (7.6) certainly holds when either $x_{m+1,p+1} = 0$, or $x_{m+1,j} = 0$ for all $j \in \{1, \dots, r\}$. Hence from now on we only consider solutions satisfying $x_{m+1,p+1} = 1$, and $\sum_{j=1}^r x_{m+1,j} \geq 1$. Now, for each such solution we define an index set J and a quantity $c[J]$ as follows:

$$J := \{j \in \{1 \dots r\} \mid x_{m+1,j} = 1\}, \\ c[J] := \pi_0 - \max\{\sum_{i=1}^m \sum_{j=1}^p \pi_{ij} x_{ij} \mid x \in S_{mp}(J)\}.$$

From the definition of the sequential covering, it follows that $c[J] \geq \sum_{j \in J} c[v_j]$; to see this, consider the definition of $c[v_h]$ and take $h = \max\{j \mid j \in J\}$ and $I = J \setminus \{h\}$. Moreover, for each $i \in \{1, \dots, m\}$ and $j \in J$, $x_{ij} = 0$ (since $x_{m+1,j} = x_{m+1,p+1} = 1$, and the edge (u_i, v_{p+1}) is not in B). Thus, by definition of $c[J]$, we get $\sum_{i=1}^m \sum_{j=1}^p \pi_{ij} x_{ij} \leq \pi_0 - c[J]$. From the former observations, we deduce:

$$\begin{aligned} \sum_{i=1}^{m+1} \sum_{j=1}^{p+1} \pi_{ij} x_{ij} &\leq \pi_0 - c[J] + \sum_{j=1}^r \pi_{m+1,j} y_{m+1,j} + \pi_{m+1,p+1} \\ &\leq \pi_0 + \sum_{j \in J} (\pi_{m+1,j} - c[v_j]) + \pi_{m+1,p+1} = \pi_0 + \pi_{m+1,p+1} \end{aligned}$$

□

Let us now see how to lift a facet defining inequality (7.5) for Q_{mp} to a facet defining inequality for Q_B . A V -sequence Φ is called *tight* with respect

to the valid inequality (7.5) if there exists a complete bipartite partitioning Y satisfying (7.5) at equality, such that all vertices of the sequence Φ are in the same connected component of Y and there are no other vertices of V_p in this component. Notice that there is always at least one tight V-sequence associated with any facet defining inequality. For instance, it can be checked that a trivial inequality has exactly one tight V-sequence (and one tight U-sequence), and that for each square inequality there are four different tight V-sequences (and also four different tight U-sequences).

Proposition 7.4.2 *Let (7.5) be a valid and facet defining inequality for Q_{mp} , let Φ be a tight V-sequence, and let the inequality (7.6) be constructed as in Proposition 7.4.1. Then the inequality (7.6) defines a facet of Q_B , where B is the U-extension of K_{mp} with respect to Φ .*

Proof: The validity of inequality (7.6) follows from Proposition 7.4.1. Since there exist bipartite clique partitionings satisfying the inequality at equality, the inequality induces a proper face of Q_B , say F . This face is contained in a facet of Q_B , say F_β , which on its turn is induced by an inequality: $\sum_{i=1}^{m+1} \sum_{j=1}^{p+1} \beta_{ij} x_{ij} \leq \beta_0$. To prove the proposition, it suffices to show that the latter inequality and inequality (7.6) do not differ apart from multiplication by a strictly positive scalar.

Let X^A denote the incidence matrix induced by the edge-set A . Since (7.5) induces a facet F_Π of Q_{mp} and there is a one-to-one relation between the incidence matrices in F_Π and the incidence matrices Y in S_B satisfying $y_{m+1,p+1} = 1$ and $y_{m+1,j} = 0$ for all $j \in \{1, \dots, r\}$, it follows that there is a strictly positive scalar c such that $\beta_{ij} = c\pi_{ij}$ if $(u_i, v_j) \in U_m \times V_p$. From this it also follows that $\beta_0 = c\pi_0 + \beta_{m+1,p+1}$.

The next step is to show that $\beta_{m+1,j} = c\pi_{m+1,j}$ for all $j \in \{1, \dots, r\}$. We use induction on j . So, let $j \in \{1, \dots, r\}$ and assume that $\beta_{m+1,t} = c\pi_{m+1,t}$ for all $t \in \{1, \dots, j-1\}$. Let I be an index set such that $I \subseteq \{1, \dots, j-1\}$, and

$$c[v_j] = \pi_0 - \sum_{i \in I} c[v_i] - \max \left\{ \sum_{i=1}^{m+1} \sum_{k=1}^{p+1} \pi_{ik} x_{ik} \mid x \in S_{mp}(I \cup \{j\}) \right\}.$$

By the definition of the sequential covering, such an index set exists. Let $\hat{X} \in S_{mp}$ represent a complete bipartite partitioning of K_{mp} such that:

$$\sum_{i=1}^{m+1} \sum_{k=1}^{p+1} \pi_{ik} \hat{x}_{ik} = \max \left\{ \sum_{i=1}^{m+1} \sum_{k=1}^{p+1} \pi_{ik} x_{ik} \mid x \in S_{mp}(I \cup \{j\}) \right\}.$$

Now we can define an edge set D_j as follows:

$$D_j := \{(u_i, v_t) \in U_m \times V_p \mid \hat{x}_{it} = 1\} \\ \cup \{(u_{m+1}, v_t) \mid t \in I \cup \{j\}\} \cup \{(u_{m+1}, v_{p+1})\}.$$

The reader can verify that D_j induces a complete bipartite partitioning of B whose incidence matrix Y is in F , and therefore also in F_β . Then we deduce:

$$\begin{aligned} \beta_0 &= \sum_{i=1}^{m+1} \sum_{k=1}^{p+1} \beta_{ik} y_{ik} \\ &= c \left(\sum_{i=1}^m \sum_{k=1}^p \pi_{ik} y_{ik} \right) + c \left(\sum_{k \in I} c[v_k] \right) + \beta_{m+1,j} + \beta_{m+1,p+1} \\ &= c(\pi_0 - c[v_j]) + \beta_{m+1,j} + \beta_{m+1,p+1} \\ &= \beta_0 + \beta_{m+1,j} - c \cdot c[v_j] \end{aligned}$$

Therefore, $\beta_{m+1,j} = c \cdot c[v_j] = c\pi_{m+1,j}$.

The final step is to show that $\beta_{m+1,p+1} = c\pi_{m+1,p+1}$. To achieve this, we define two edge sets: A_{p+1} and D_{p+1} , both complete bipartite partitionings of B , such that the corresponding incidence matrices are in F , and therefore also in F_β . We define A_{p+1} to consist of the edge (u_{m+1}, v_{p+1}) , together with the edge-set of the complete bipartite partitioning Y referred to in the definition of tight sequences. The set D_{p+1} is defined in the following way: $D_{p+1} := A_{p+1} \cup \{(u_{m+1}, v_j) \mid j \in \{1, \dots, r\}\} \setminus \{(u_{m+1}, v_{p+1})\}$. Comparing the sets A_{p+1} and D_{p+1} , it can be seen that $\beta_{m+1,p+1} = c\pi_{m+1,p+1}$. $\square$

Proposition 7.4.2, together with Proposition 7.2.1, implicitly provides a way of lifting any facet defining inequality for Q_{mp} to a facet defining inequality for $Q_{m+1,p+1}$ (and hence, by Proposition 7.2.2, for Q_{nq} , $n \geq m$, $q \geq p$). Of course, this is only interesting if, for at least one vertex of the sequence, the corresponding sequential covering is strictly positive. We are now going to describe an explicit lifting procedure based on Proposition 7.4.2. The procedure, however, only applies to a subclass of facet defining inequalities.

We first introduce some more terminology. We say that a valid inequality for Q_{mp} is *spiked* when its positive support is spiked (see Section 7.3), and when the *spike partitioning* S , defined by $s_{ij} = 1$ if and only if $\{u_i, v_j\}$ is a spike, satisfies the inequality at equality. Simple examples of spiked inequalities are provided by the square inequalities, or by any of the facet defining inequalities introduced in Section 7.3.2. In the remainder of this section, when dealing with spiked valid inequalities for Q_{kk} , we implicitly

assume that the spikes, the spike-roots, and the spike-leaves are numbered in a such a way that spike-root i and spike-leaf i are the endpoints of spike i ($i = 1, \dots, k$).

For spiked inequalities, the sequential covering of each vertex is easily characterized, as made explicit by the following proposition (the result will be used in the next proposition).

Proposition 7.4.3 *Let (7.5) be a spiked inequality defining a facet of Q_{kk} , and let $\Phi = (v_1, \dots, v_r)$ be a V -sequence. For all $i \in \{1 \dots r\}$, the sequential covering of v_i is equal to the coefficient of the corresponding spike in the inequality (i.e. $c[v_i] = \pi_{ii}$) if v_i is a spike-root, and the sequential covering of v_i is equal to 0 if v_i is a spike-leaf.*

Proof: First we will prove that $c[v_i] = 0$ if v_i is a spike-leaf. From Proposition 7.3.3(a), it follows that there exists a matrix $X \in S_{kk}$ satisfying (7.5) at equality such that $y_{ti} = 0$ for all $t \in \{1, \dots, k\}$. Using the definition of the sequential covering, we deduce:

$$c[v_i] \leq \pi_0 - \sum_{i=1}^m \sum_{j=1}^k \pi_{ij} x_{ij} = 0.$$

Combining this result with the observation that the sequential covering of an arbitrary vertex is always nonnegative, we conclude that $c[v_i] = 0$.

The proof of the complete statement is by induction on i . Assume that the statement holds for $i = 1, \dots, h-1$. If v_h is a spike-leaf, then the statement is also true for $i = h$, as shown above. So, assume v_h is a spike-root, and let $I \subseteq \{1, \dots, h-1\}$ be such that

$$c[v_h] = \pi_0 - \sum_{i \in I} c[v_i] - \max\left\{\sum_{i=1}^m \sum_{j=1}^k \pi_{ij} x_{ij} \mid X \in S_{kk}(I \cup \{h\})\right\}$$

We may assume without loss of generality that v_t is a spike-root for all $t \in I \cup \{h\}$. To see this, notice that if $t \in I$ and v_t is a spike-leaf, then we can replace I by $I \setminus \{t\}$, since $c[v_t] = 0$ as proved earlier, and

$$\begin{aligned} \max\left\{\sum_{i=1}^m \sum_{j=1}^k \pi_{ij} x_{ij} \mid X \in S_{kk}(I \cup \{h\})\right\} \leq \\ \max\left\{\sum_{i=1}^m \sum_{j=1}^k \pi_{ij} x_{ij} \mid X \in S_{kk}(I \cup \{h\} \setminus \{t\})\right\} \end{aligned}$$

Consider now the 'restricted' spike partitioning Y , defined by $y_{ij} = 1$ if and only if $i = j$ and $i \notin I \cup \{h\}$; clearly,

$$\begin{aligned} c[v_h] &\leq \pi_0 - \sum_{i \in I} c[v_i] - \sum_{i=1}^m \sum_{j=1}^k \pi_{ij} y_{ij} \\ &= \pi_0 - \sum_{i \in I} c[v_i] - \sum_{i \notin I \cup \{h\}} \pi_{ii}. \end{aligned}$$

By definition of spiked inequalities, $\pi_0 = \sum_{i=1}^k \pi_{ii}$. Hence, using induction:

$$c[v_h] \leq \sum_{i \in I \cup \{h\}} \pi_{ii} - \sum_{i \in I} c[v_i] = \pi_{hh}.$$

To see that the reverse inequality also holds, consider any $Z \in S_{kk}(I \cup \{h\})$ realizing the maximum over $\max\{\sum_{i=1}^m \sum_{j=1}^k \pi_{ij} x_{ij} \mid X \in S_{kk}(I \cup \{h\})\}$. For each $i \in I \cup \{h\}$, u_i is a spike-leaf, and $z_{ii} = 0$. Hence, since $\pi_{ij} \leq 0$ for all $j \neq i$, we may assume without loss of generality that $z_{ij} = 0$ for all $i \in I \cup \{h\}$ and for all $j = 1, \dots, k$. Define now a point W by:

$$w_{ij} = \begin{cases} z_{ij} & \text{if } i \notin I \cup \{h\} \text{ and } j \notin I \cup \{h\}, \\ 1 & \text{if } i \in I \cup \{h\} \text{ and } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

This point defines a complete bipartite partitioning of K_{kk} . Thus using the definition of Z , and induction:

$$\begin{aligned} c[v_h] &= \pi_0 - \sum_{i \in I} c[v_i] - \sum_{i=1}^m \sum_{j=1}^k \pi_{ij} z_{ij} \\ &\geq \sum_{i=1}^m \sum_{j=1}^k \pi_{ij} w_{ij} - \sum_{i \in I} c[v_i] - \sum_{i=1}^m \sum_{j=1}^k \pi_{ij} z_{ij} \\ &= \sum_{i \in I \cup \{h\}} \pi_{ii} - \sum_{i \in I} c[v_i] = \pi_{hh}. \end{aligned}$$

□

Now, our next proposition allows to lift spiked facet defining inequalities of Q_{kk} to spiked facet defining inequalities of $Q_{k+1,k+1}$.

Proposition 7.4.4 *Let (7.5) be a spiked inequality defining a facet of Q_{kk} , and let $\Phi = \{v_1, \dots, v_r\}$ be a tight V -sequence for this inequality. Assume the coefficients Π satisfy the conditions of Proposition 7.4.1 for all edges (u_i, v_j) of the U -extension of K_{kk} , and let:*

$$\begin{aligned} \pi_{k+1,j} &= 0; \text{ for all } j \notin \{1, \dots, r\} \\ \pi_{i,k+1} &= \begin{cases} -\pi_{k+1,i} & \text{if } u_i \text{ is a spike-leaf,} \\ -\sum_{l \in \{1, \dots, r\} \setminus \{i\}} [\pi_{il}]^+ & \text{if } u_i \text{ is a spike-root.} \end{cases} \end{aligned}$$

Then, the inequality (7.6) is a spiked inequality defining a facet of $Q_{k+1,k+1}$, and Φ is a tight V -sequence for (7.6).

Proof: Let $Y \in S_{k+1,k+1}$. Reasoning as in the proof of Proposition 7.4.1, we conclude that the inequality (7.6) certainly holds whenever $y_{k+1,k+1} = 0$, or $y_{k+1,j} = 0$ for all $j \in \{1, \dots, k\}$, or $y_{i,k+1} = 0$ for all $i \in \{1, \dots, k\}$. So, we may assume that $y_{k+1,k+1} = 1$, and there are nonempty index-sets $I \subseteq \{1, \dots, k\}$ and $J \subseteq \{1, \dots, k\}$ such that $y_{i,k+1} = 1$ if and only if $i \in I$ and $y_{k+1,j} = 1$ if and only if $j \in J$.

Next, we construct a vector $\bar{Y} \in S_{k+1,k+1}$ from Y such that $\sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \pi_{ij} \bar{y}_{ij} \leq \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \pi_{ij} y_{ij} \leq \pi_0$. Let $I_+ := \{i \in I \mid \pi_{k+1,i} > 0\}$, $I_- := I \setminus I_+ = \{i \in I \mid \pi_{k+1,i} = 0\}$, $J_+ := \{j \in J \mid \pi_{k+1,j} > 0\}$, and $J_- := J \setminus J_+ = \{j \in J \mid \pi_{k+1,j} = 0\}$. Now the vector $\bar{Y}$ is obtained by splitting the component of Y induced by $I \cup J \cup \{k+1\}$, into a component induced by the vertices which correspond to I_- and J_- , a component induced by the vertices corresponding to J_+ , u_{k+1} and v_{k+1} , and the 'one-vertex-components' induced by the vertices corresponding to I_+ . More formally, the complete bipartite partitioning is constructed in the following way:

$$\bar{y}_{ij} = \begin{cases} 0 & \text{for all } i \in I_+ \text{ and all } j \in J \cup \{k+1\}; \\ 0 & \text{for all } i \in I_- \text{ and all } j \in J_+ \cup \{k+1\}; \\ 0 & \text{for } i = k+1 \text{ and all } j \in J_-; \\ y_{ij} & \text{in all other cases.} \end{cases}$$

The reader can verify that the coefficients are constructed in such a way that the next expression holds:

$$\sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \pi_{ij} (y_{ij} - \bar{y}_{ij}) = \sum_{i \in I_+} \sum_{j \in J \cup \{k+1\}} \pi_{ij} + \sum_{i \in I_-} \sum_{j \in J_+ \cup \{k+1\}} \pi_{ij} + \sum_{j \in J_-} \pi_{k+1,j}$$

Consider the first term. For any $i \in I_+$, it follows from the definition of I_+ that v_i is a spike-root belonging to Φ , and hence u_i is a spike-leaf. So $\sum_{j \in J} \pi_{ij} \leq \pi_{ii}$, $\pi_{i,k+1} = -\pi_{k+1,i} = -\pi_{ii}$, and the first term is not strictly positive. Then consider the second term and $i \in I_-$. Notice that, by definition of the sets I_- and J_+ , $i \in I_-$ implies $i \notin J_+$. Now, either u_i is a spike-leaf, or u_i is a spike-root. In the first case $\pi_{ij} \leq 0$ for all $j \in J_+$ (since $i \notin J_+$) and $\pi_{i,k+1} = -\pi_{k+1,i} = 0$. In the second case, observe that $J_+ \subseteq \{1, \dots, r\} \setminus \{i\}$ (in view of the definition of the coefficients $\pi_{k+1,j}$), and thus: $\pi_{i,k+1} = -\sum_{l \in \{1, \dots, r\} \setminus \{i\}} [\pi_{il}]^+ \leq -\sum_{j \in J_+} \pi_{ij}$. Therefore, we conclude that the second term cannot be strictly positive either. Since all coefficients occurring in the third term of the expression are nonpositive (by definition of J_-), the whole expression is nonpositive. Since $\bar{Y}$ satisfies (7.6) (because

$\bar{y}_{i,k+1} = 0$ for all $i \in \{1, \dots, k\}$ and the earlier reasoning), we conclude that the inequality (7.6) holds for all $Y \in S_{k+1,k+1}$.

Now the validity of the inequality (7.6) is shown, we prove that the inequality induces a facet of $Q_{k+1,k+1}$. Since there exist complete bipartite partitionings satisfying the inequality with equality, the inequality induces a proper face of $Q_{k+1,k+1}$, say F . This face is contained in a facet of $Q_{k+1,k+1}$, say F_β , which on its turn is induced by an inequality $\sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \beta_{ij} x_{ij} \leq \beta_0$. To prove the proposition it suffices to show that this inequality and inequality (7.6) do not differ apart from multiplication by a strictly positive scalar.

Let X^A denote the incidence matrix induced by the edge set A . By Proposition 7.4.2, inequality (7.6) induces a facet F_Φ of Q_B , where B is the U-extension of K_{kk} with respect to Φ . Since there is a one-to-one relation between the incidence vectors in F_Φ and the incidence matrices X in $S_{k+1,k+1}$ satisfying $y_{i,k+1} = 0$ for all $i \in \{1, \dots, k\}$, and $y_{k+1,j} = 0$ for all j such that $v_j \notin \Phi$, it follows that there is a strictly positive scalar c such that $\beta_{ij} = c\pi_{ij}$ if $(u_i, v_j) \in E(B)$.

Consider now the coefficients $\pi_{i,k+1}$ and $\pi_{k+1,j}$ which have value 0. First, we show that $\beta_{i,k+1} (= c\pi_{i,k+1}) = 0$ for all $i \notin \{1, \dots, r\}$ such that u_i is a spike-leaf. Since (7.5) defines a facet of Q_{kk} , it follows from Proposition 7.3.3(a) that there is an incidence matrix $X^A \in S_{kk}$ satisfying (7.5) at equality such that the edge-set A induced by X^A does not cover u_i . Now, we define two edge-sets: $B_{i,k+1} := A \cup \{(u_{k+1}, v_{k+1})\}$, and $D_{i,k+1} := A \cup \{(u_{k+1}, v_{k+1}), (u_i, v_{k+1})\}$. It can be checked that $B_{i,k+1}$ and $D_{i,k+1}$ both induce proper complete bipartite partitionings of $K_{k+1,k+1}$, whose characteristic vectors are in F . Comparing these sets, it can easily be seen that $\beta_{i,k+1} = 0$. The same reasoning also shows that $\beta_{k+1,j} = c\pi_{k+1,j} = 0$ for all j such that v_j is a spike-leaf.

Next, we show that $\beta_{k+1,j} = 0$ for all $j \notin \{1, \dots, r\}$ such that v_j is a spike-root. To do this we compare two edge sets. One of them is the edge-set S induced by the spike partitioning, i.e. $S = \{(u_i, v_i) \mid i \in \{1, \dots, k+1\}\}$. The other set is defined as follows: $D_{k+1,j} := S \cup \{(u_{k+1}, v_j), (u_j, v_{k+1})\}$. It can be checked that $D_{k+1,j}$ induces a proper complete bipartite partitioning of $K_{k+1,k+1}$. Comparing the sets S and $D_{k+1,j}$, we conclude $\beta_{k+1,j} + \beta_{j,k+1} = 0$. We showed before that $\beta_{j,k+1} = 0$, and therefore $\beta_{k+1,j} = 0$ must hold too.

The previous construction also shows that $\beta_{i,k+1} = -\beta_{k+1,i} = -c\pi_{k+1,i} = c\pi_{i,k+1}$ for all $i \in \{1, \dots, r\}$ such that u_i is a spike-leaf.

The final step is to show that $\beta_{i,k+1} = c\pi_{i,k+1}$ for all $i \in \{1, \dots, k\}$ such that u_i is a spike-root. As before, we compare two edge sets; one of them is the edge set S induced by the spike partitioning. The other set is defined as

follows:

$$D_{i,k+1} := \left(S \setminus \{(u_l, v_l) \mid l \in \{1, \dots, r\} \setminus \{i\}, \pi_{il} > 0\} \right) \\ \cup \{(u_i, v_l), (u_{k+1}, v_l) \mid l \in \{1, \dots, r\} \setminus \{i\}, \pi_{il} > 0\} \\ \cup \{(u_{k+1}, v_i), (u_i, v_{k+1})\}.$$

It can be checked that S and $D_{i,k+1}$ both induce a complete bipartite partitioning of $K_{k+1,k+1}$. Because $\beta_{k+1,l} = \beta_{il} = \pi_{il}$ for all l such that v_l is a spike-root and $v_l \in \Phi$ (by Proposition 7.4.3), comparing the sets $D_{i,k+1}$ and S yields $\beta_{i,k+1} = -\sum_{l \in \{1, \dots, r\} \setminus \{i\} : \pi_{il} > 0} \beta_{il} = c\pi_{i,k+1}$. The reader can easily verify that the new inequality is a spiked inequality, for which Φ is a tight sequence as evidenced by the tight solution corresponding to the edge-set $A \cup \{(u_{k+1}, v_{k+1})\}$, where A denotes the edge-set inducing the tight solution of $\Pi X \leq \pi_0$. $\square$

We give below some examples of spiked inequalities. We include in a V-sequence (U-sequence) precisely those vertices of V (U) which are spike-roots. For the examples below, these sequences happen to be tight. Notice that, as follows from Proposition 7.4.3, the ordering of the vertices in a V-sequence (U-sequence) does not matter for spiked inequalities. By applying Proposition 7.4.4 to these examples, new inequalities can be generated.

For instance the inequality corresponding to Π_1 which has right hand side 2 can be lifted to an inequality corresponding to Π_2 with right hand side 3, which on its turn can be lifted to an inequality based upon $(\Pi_3, 4)$ (U-extension), or to an inequality based upon $(\Pi_4, 5)$ (V-extension), etcetera. In the first three of these inequalities, the reader will recognize a subclass of spiked tree inequalities (see Section 7.3.2). Furthermore, notice that in the inequality based upon $(\Pi_4, 5)$ coefficients emerge with values different from $-1, 0$ and 1 . Actually, by lifting these inequalities based on a proper order of U- and V-sequences, any integer number can occur as coefficient.

$$\Pi_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \Pi_2 = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}, \\ \Pi_3 = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \Pi_4 = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ -2 & -1 & -1 & 2 \end{bmatrix}.$$

Other examples can be derived from the inequality corresponding to $(\Pi_5, 4)$. This facet defining inequality, based on a spiked cycle with four spikes, can

be lifted for instance to the inequality corresponding to $(\Pi_6, 6)$, and so on. Also the spiked cycle inequality corresponding to $(\Pi_7, 6)$ (see Section 7.3.2) can be lifted to the inequality corresponding to $(\Pi_8, 9)$, and so on.

$$\Pi_5 = \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}, \Pi_6 = \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -2 \\ 1 & 1 & -1 & 1 & -2 \\ 1 & 1 & 0 & 0 & 2 \end{bmatrix},$$

$$\Pi_7 = \begin{bmatrix} 1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 1 & -1 & -1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & -1 \\ 0 & 0 & -1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix},$$

$$\Pi_8 = \begin{bmatrix} 1 & 1 & 0 & 0 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & -1 & 1 & 0 \\ -1 & -3 & -1 & -3 & -1 & -3 & 3 \end{bmatrix}.$$

An explicit lifting procedure of a different nature is given by the next proposition.

Proposition 7.4.5 *Let (7.5) be a spiked tree inequality for Q_{mp} . Assume that the vertices v_r and v_s are spike-leaves of the positive support of this inequality, and assume $\pi_{rr} = \pi_{ss} = 1$, $\pi_{rs} = 0$, $\pi_{sr} = -1$, and $\pi_{rj} = \pi_{sj}$ for all $j \in \{1, \dots, p\} \setminus \{r, s\}$. Let the inequality*

$$\sum_{i=1}^m \sum_{j=1}^{p+1} \gamma_{ij} x_{ij} \leq \gamma_0 \quad (7.7)$$

be constructed as follows:

$$\gamma_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \{(r, s), (s, p+1)\}; \\ -1 & \text{if } i = r \text{ and } j = p+1; \\ \pi_{ir} & \text{if } i \in \{1, \dots, m\} \setminus \{r, s\} \text{ and } j = p+1; \\ \pi_{ij} & \text{in all other cases;} \end{cases}$$

$$\gamma_0 = \pi_0 + 1.$$

Then the inequality $\Pi Y \leq \pi_0$ defines a facet of $Q_{m,p+1}$.

This proposition allows to describe facet defining inequalities whose positive support has vertex-sets U_m and V_p with different cardinalities. For example, the inequality induced by $(\Pi_9, 3)$ can be lifted to the inequality based on $(\Pi_{10}, 4)$.

$$\Pi_9 = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}, \Pi_{10} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}.$$

The proof of this proposition is omitted, since it is rather cumbersome, and the arguments it uses do not differ much from those in former proofs, especially the proof of Proposition 7.3.5 (concerning spiked tree inequalities). Although the inequalities resulting from the application of Proposition 7.4.5 are not spiked tree inequalities (nor even spiked inequalities), the premises of this proposition could be adapted in such a way as to hold for the new inequalities, and to allow iterative applications of the adapted statement. Also Proposition 7.4.4 could be extended so as to be applicable to the inequalities resulting from Proposition 7.4.5.

Chapter 8

Patching Procedures for the Polytope of the Cell Formation Problem

This chapter is based on joint work with Yves Crama, and joint work with Jeroen Rutten and Frits Spieksma.

The paper with Yves Crama will be published as an article in *Networks*.

The paper with Jeroen Rutten en Frits Spieksma has been submitted for publication.

A matrix X is called block diagonal if there exists a partition $R_1, \dots, R_k, R_{k+1}$ of its row-set and a partition $C_1, \dots, C_k, C_{k+1}$ of its column-set such that $x_{ij} \neq 0$ if and only if $i \in R_l$ and $j \in C_l$ for some $1 \leq l \leq k$. Now the block diagonalization problem can be formulated as follows: given an $m \times p$ incidence matrix A and a function $f(\cdot, \cdot)$, find an $m \times p$ block diagonal incidence matrix X which minimizes $f(A, X)$.

This problem can be interpreted as an abstraction of the cell formation problem encountered in cellular manufacturing, as discussed in Chapter 6. Similar problems also arise in the analysis of large data arrays (e.g. for marketing or archeology applications), in production planning for flexible manufacturing systems, in sparse matrix computations, etc.

In the terminology of graph theory the problem can be formulated as a variant of the clique partitioning problem. Given is a complete bipartite graph $G = (U, V, E)$, with edge weights $w_{ij} \in \mathbb{R}$ for all $\{i, j\} \in E$. A subset $A \subseteq E$ is called a bipartite clique partition if there is a partition of the vertex sets (U, V) into non-empty, disjoint sets $(U_1, V_1), \dots, (U_k, V_k)$, such that each (U_p, V_p) ($p = 1, \dots, k$) induces a bipartite clique (i.e. a complete bipartite subgraph), and $A = \bigcup_{p=1}^k U_p \times V_p$. The weight of such a bipartite clique partition A is defined as $\sum_{\{i,j\} \in A} w_{ij}$. The problem is now to find a bipartite clique partition of maximal weight.

In Chapter 7, we discuss the facial structure of the polytope whose extreme points are exactly the $m \times p$ 0-1 block diagonal matrices. In this chapter we introduce procedures that generate facet defining inequalities by combining known facet defining inequalities.

8.1 Introduction

In this chapter, we investigate an interesting property of the facial structure of the polytope whose extreme points are exactly the $m \times p$ 0-1 block diagonal matrices ($m, p \in \mathbb{N}$). This property makes it possible to *patch* facet defining inequalities into new valid inequalities which, under certain conditions, also define facets. Here we will restrict ourselves to procedures which take as input two facet defining inequalities.

As in Chapter 6 and Chapter 7, we define a matrix X to be *block diagonal* if there exists a partition $R_1, \dots, R_k, R_{k+1}$ of its row-set and a partition $C_1, \dots, C_k, C_{k+1}$ of its column-set such that $x_{ij} \neq 0$ if and only if $i \in R_l$ and $j \in C_l$ for some $1 \leq l \leq k$ (notice that what we really mean is that X

is block diagonal up to permutations of its rows and columns). We let

$$S_{mp} = \{ X \in \{0,1\}^{m \times p} \mid X \text{ is block diagonal} \}$$

and we denote by Q_{mp} the convex hull of S_{mp} . In Chapter 7 a (partial) description of the polytope Q_{mp} by linear inequalities is provided. This description is based on the characterization of block diagonal matrices by so-called *square inequalities*:

$$x_{hj} + x_{hk} + x_{ij} - x_{ik} \leq 2 \quad (8.1)$$

It has been shown in Chapter 6 that an incidence matrix is block diagonal if and only if the elements of the incidence matrix satisfy (8.1) for all $h, i \in \{1, \dots, m\}$, and for all $j, k \in \{1, \dots, p\}$.

In this presentation we will often rely on a graph-theoretic interpretation of block diagonal matrices and of the polytope Q_{mp} . We follow the graph-theoretic terminology of [Bondy and Murty, 1976]. Moreover, when $B = (U, V, E)$ is a bipartite graph and $G = (U, V, F)$ is a subgraph of B , we say that G is a *complete bipartite partitioning* of B if all connected components of G are complete bipartite (we look at isolated vertices as complete bipartite graphs). In particular, consider the complete bipartite graph $K_{mp} = (U_m, V_p, U_m \times V_p)$, where $U_m = \{u_1, \dots, u_m\}$ and $V_p = \{v_1, \dots, v_p\}$. We regard an arbitrary $m \times p$ 0-1 matrix X as the adjacency matrix of a subgraph G of K_{mp} , say $G = (U_m, V_p, F)$, where $(u_i, v_j) \in F$ if and only if $x_{ij} = 1$. It is easy to see that the matrix X is block diagonal if and only if its associated graph is a complete bipartite partitioning of K_{mp} .

This graph-theoretic interpretation stresses the analogy of the polytope Q_{mp} with the clique partitioning polytope P_n investigated by [Faigle et al, 1986] and [Grötschel and Wakabayashi, 1989], and with the related multiway cut polytope studied by [Chopra and Rao, 1991]. In fact, Q_{mp} can be viewed as the projection of P_{m+p} on some appropriate subspace. But this observation does not seem very useful in deriving a description of Q_{mp} from the results available about P_n ; for example, there is not a one to one relationship between (classes of) facet defining inequalities of these polytopes, (see also Chapter 2).

In Chapter 7 some interesting properties of the polytope of the cell formation problem have been presented. For example, it has been shown that an inequality defining a facet of Q_{mp} is also facet defining for each of the polyhedra corresponding to block diagonal matrices with at least m rows and p columns. This holds without imposing conditions on the specific form of

the inequality. Furthermore, without any information about the numerical values of the coefficients of a valid inequality

$$\sum_{i \in U_m} \sum_{j \in V_p} \pi_{ij} x_{ij} \leq \pi_0 \quad (8.2)$$

some general structural properties of the inequality have been stated. We briefly summarize them now. Consider inequality (8.2) and assume that it defines a nontrivial facet of Q_{mp} . Associate with the inequality two edge sets E and E^+ , defined as follows:

$$\begin{aligned} E &:= \{(u_i, v_j) \mid (u_i, v_j) \in U_m \times V_p \text{ and } \pi_{ij} \neq 0\}, \\ E^+ &:= \{(u_i, v_j) \mid (u_i, v_j) \in U_m \times V_p \text{ and } \pi_{ij} > 0\}. \end{aligned}$$

We call $H := (V(E), E)$ (respectively $H^+ := (V(E^+), E^+)$) the support (respectively the positive support) of the inequality (8.2). Then we have:

- (1) $\pi_0 > 0$;
- (2) E^+ is nonempty, i.e. there are positive coefficients π_{ij} ;
- (3) the positive support H^+ of (8.2) is connected;
- (4) $V(E) = V(E^+)$.
- (5) $E \setminus E^+$ is nonempty, i.e. there are negative coefficients π_{ij} ;
- (6) the support H of (8.2) is two-connected.

In this Chapter we will show that next to these nice properties, the polytope also allows for constructing new facet defining inequalities from old ones, in a relatively easy way, referred to as patching. In Section 8.2, the general principles of patching are introduced. In the next section we concentrate on patching inequalities with overlapping supports. In Section 8.4 we study patching inequalities with node disjoint supports. For the clique partitioning problem computational studies for a particular set of real-life instances show that the LP-bound often coincides with an integral solution. In spite of that fact, it is not difficult to construct instances for which the LP-bound is arbitrarily bad. We can derive instances like that easily from valid inequalities for which the ratio between the right hand side and the sum of the positive coefficients can be arbitrarily close to zero (see chapter 4 for a discussion on the strength of inequalities). For the bipartite clique partitioning problem it is less trivial to construct instances for which the LP-bound is arbitrarily bad. However, with the patching procedures we can construct classes of valid inequalities for which the ratio between the right hand side and the sum of the positive coefficients can be arbitrarily close to zero. Such a class of inequalities is presented in Section 8.5.

8.2 Covering, Patching, and Recognizing

It is often possible to use a set of valid inequalities to construct a new valid inequality that is not implied by the former. Well-known examples are the procedures to derive Gomory-Chvatal-cuts. In this chapter we focus on so-called *patching procedures*. A patching procedure combines a set of inequalities into a single one such that the nonzero coefficients of these inequalities remain unchanged. We refer to this process as *patching*.

Here, we restrict ourselves to combining pairs of inequalities and we distinguish two cases of patching, depending on the intersection of the node sets of the supports of the inequalities involved. If this intersection is empty, we refer to combining the two inequalities as *disjoint patching*, otherwise we call it *intersection patching*. Notice that in the case of intersection patching the two inequalities must be such that the coefficients corresponding to any edge induced by the intersection are identical.

In [Grötschel and Wakabayashi, 1990] some intersection patching procedures for a class of inequalities of the clique partitioning polytope are presented. In [Crama and Oosten, 1993] disjoint patching procedures are presented for the bipartite clique partitioning problem (see Section 8.4). In [Oosten et al, 1994] several new patching procedures, intersection as well as disjoint procedures, for the clique partitioning polytope are presented. In this Chapter intersection as well as disjoint patching procedures for the cell formation polytope are discussed in Section 8.3, respectively in Section 8.4.

These procedures need the concept of a *covering*. A covering is a value assigned to a set of variables with respect to an inequality. Here, we will not need a value for every set of variables. We are only interested in sets of variables that generate a bipartite clique. Therefore, we will introduce the concept as follows. Consider the following inequality:

$$\sum_{i \in U_m} \sum_{j \in V_p} \alpha_{ij} x_{ij} \leq \alpha_0 \quad (8.3)$$

Assume the set of edges C is a subset of $U_m \times V_p$. Let U_C be the intersection of U_m and C , and let V_C be the intersection of V_p and C . Then, the covering of (U_C, V_C) with respect to inequality 8.3 is denoted as $\text{cov}_\alpha[C]$ and defined as follows.

$$\text{cov}_\alpha[C] := \alpha_0 - \max_{x \in S_{mp}} \left\{ \sum_{i \in U_m} \sum_{j \in V_p} \alpha_{ij} x_{ij} \mid x_{ij} = 0 \ \forall i, j : i \in U_C \text{ or } j \in V_C \right\}$$

Notice that the covering of any node set V_C is nonnegative, assuming that the inequality is valid.

As an illustration, consider the square inequality $x_{11} + x_{12} + x_{21} - x_{22} \leq 2$. In the table below the values of the covering with respect to this inequality is displayed for each set containing one or two nodes of the positive support of this inequality. The covering of sets containing three or more nodes of the support of the square inequality is always two.

	u_1	u_2	v_1	v_2
u_1	1	1	2	2
u_2		0	1	2
v_1			0	1
v_2				1

Table 8.1 Values of Coverings for a Square Inequality.

Let us make a few observations. The covering of a single node s equals zero if and only if there exists a clique partition not incident to s satisfying (8.3) at equality. From this observation and Proposition ??(a) it follows directly that the covering of a spike leaf is zero. Furthermore, if the set $\bar{C}$ is a subset of C , then $\text{cov}_\alpha[C] \geq \text{cov}_\alpha[\bar{C}]$. Finally, notice that the definition of a covering implies that the following inequalities are valid:

$$\sum_{i \in U_m \setminus U_C} \sum_{j \in V_p \setminus V_C} \alpha_{ij} x_{ij} \leq a_0 - \text{cov}_\alpha[V_C], \text{ and} \quad (8.4)$$

$$\sum_{i \in U_C} \sum_{j \in V_C} \alpha_{ij} x_{ij} \leq \text{cov}_\alpha[V_C]. \quad (8.5)$$

Suppose inequality (8.5) happens to define a facet F_α of $Q_{|U_C| \times |V_C|}$, and the inequality is not equivalent to a nonnegativity inequality. It has been shown in Chapter 7 that then the inequality (8.5) also defines a facet of Q_{mp} . In the next proposition for every inequality defining a face of Q_{mp} containing F_α the coefficients corresponding to the edges $U_C \times V_C$ are equal to the coefficients of (8.5) up to multiplication by a nonnegative scalar. This is formalized as follows.

Proposition 8.2.1 *Assume C is an arbitrary subset of $U_m \cup V_p$, U_C is the intersection of U_m and C , and V_C is the intersection of V_p and C .*

Let E be a (possibly empty) arbitrary subset of $U_C \times (V_p \setminus V_C) \cup (U_m \setminus U_C) \times V_C$. Let Q^* be the face of Q_{mp} defined by the equalities $x_e = 0$ for all edges $e \in E$. Assume inequality (8.3) is valid for S_{mp} and defines a proper face F_α of Q^* . Suppose inequality (8.5) defines a facet of $Q_{|U_C| \times |V_C|}$ and is not equivalent to a nonnegativity constraint.

Then, for any inequality

$$\sum_{i \in U_m} \sum_{j \in V_p} \delta_{ij} x_{ij} \leq \delta_0 \quad (8.6)$$

defining a face F_δ of P^* containing F_α , there is a scalar $\beta \geq 0$ such that $\delta_{ij} = \beta \alpha_{ij}$ for all $i \in U_C, j \in V_C$.

Proof: Select a maximal affinely independent set of solutions from $S_{U_C \times V_C}$ satisfying (8.5) at equality. Let M be a $\dim(Q_{|U_C| \times |V_C|}) \times \dim(Q_{|U_C| \times |V_C|})$ matrix such that there is precisely one column for every solution of the set, and each column represents the incidence matrix of the corresponding solution. Assume the coefficients of (8.5) to be organized in a vector ρ corresponding to M ; that is the improduct of ρ with a column of M corresponds to the left hand side of (8.5). Then $\rho^T M = \rho_0 \cdot \mathbf{1}_{|U_m \times V_p|}^T$, where $\mathbf{1}_{|U_m \times V_p|}$ is a vector with $|U_m \times V_p|$ elements, all at value 1, and ρ_0 equals $\text{cov}_\alpha[C]$. Let ρ_C, x_C be vectors with the elements of ρ , respectively x , that correspond to the edges in $U_C \times V_C$. There exist $|U_C \times V_C|$ columns such that the rows in these columns corresponding to the elements of $U_C \times V_C$ form a matrix M_C that is nonsingular. Since the coefficients of ρ corresponding to edges covering nodes not in V_C are all zero, we have: $\rho_C^T M_C = \rho_0 \cdot \mathbf{1}_{|U_C \times V_C|}^T$, implying

$$\rho_C = \rho_0 \cdot M_C^{-1} \mathbf{1}_{|U_C \times V_C|} \quad (8.7)$$

Notice that $\rho_0 > 0$ since $\rho^T x \leq \rho_0$ is not equivalent to a nonnegativity inequality and at the same time assumed to be defining a facet.

Now, let y be a feasible solution satisfying (8.5) at equality such that $y_{ij} = 0$ for all $i \in U_C \times (V_p \setminus V_C)$, and $i \in (U_m \setminus U_C) \times V_C$. Such a solution always exists due to the definition of a covering. Then, for each solution x corresponding to a column of M_C , we construct a solution z as follows.

$$z_{ij} = \begin{cases} x_{ij} & \text{if } \{i, j\} \in U_C \times V_C \\ y_{ij} & \text{if } \{i, j\} \in (U_m \setminus U_C) \times (V_p \setminus V_C) \\ 0 & \text{otherwise} \end{cases}$$

Notice that z is in Q^* . Further, it is easy to verify that each solution z satisfies both (8.3) and (8.5) at equality. In other words, when $\tilde{M}$ refers to a $|U_m \times V_p| \times |U_C \times V_C|$ matrix where each column represents the incidence vector of a solution z , we have that $\alpha^T \tilde{M} = \alpha_0 \cdot \mathbf{1}_{|U_C \times V_C|}^T$ and, since F_δ contains F_α , $\delta^T \tilde{M} = \delta_0 \cdot \mathbf{1}_{|U_C \times V_C|}^T$.

The construction of the solutions z implies that the rows in $\tilde{M}$ corresponding to the elements of $U_C \times V_C$ form the nonsingular matrix M_C . Notice further that for all columns in $\tilde{M}$, the part of each column not in M_C is identical. Let δ_C be a vector with the elements of δ that correspond to the edges in

$U_C \times V_C$. We derive: $\delta^T \tilde{M} = \delta_C^T M_C + \gamma \cdot \mathbf{1}_{|U_C \times V_C|} = \delta_0 \cdot \mathbf{1}_{|U_C \times V_C|}^T$, where γ is the constant resulting from the identical parts of the columns. Notice that $0 \leq \gamma \leq \delta_0$, otherwise solutions could be constructed violating the inequality (8.6). So,

$$\delta_C = (\delta_0 - \gamma) M_C^{-1} \mathbf{1}_{|U_C \times V_C|}. \quad (8.8)$$

Equalities (8.7) and (8.8) imply that $\delta_{ij} = \frac{\delta_0 - \gamma}{\rho_0} \alpha_{ij}$ for all $\{i, j\} \in U_C \times V_C$. This completes the proof. $\square$

Notice that arguments used in the proof of this theorem imply that for at least one facet of Q^* containing F_α , induced by some inequality (8.6), there exists a *positive* coefficient β such that $\delta_{ij} = \beta \alpha_{ij}$ for all $\{i, j\} \in U_C \times V_C$.

The patching procedures presented in the next sections are based on this observation. A new inequality is constructed in which two known inequalities can be recognized. Given the combined inequality, for each of the original inequalities we have a multiplier, and under certain conditions we can show both of them to be nonzero, implying the existence of a new facet.

Instead of constructing new inequalities, Proposition 8.2.1 can also be used to show that a given inequality defines a facet of Q_{mp} , as witnessed by the following theorem.

Proposition 8.2.2 *Assume inequality (8.3) is valid for S_{mp} . Given are k subsets $C_1, \dots, C_k$ of $U_m \cup V_p$. Define the edge set E as follows: $E := (U_m \times V_p) \setminus (\bigcup_{i=1}^k U_{C_i} \times V_{C_i})$. Let Q^* be the face of Q_{mp} defined by the equalities $x_e = 0$ for all edges $e \in E$.*

Furthermore, assume that:

- (i) *for each C_i ($i = 1, \dots, k$) inequality (8.5) defines a facet of $Q_{|U_{C_i}| \times |V_{C_i}|}$, and the inequality is not equivalent to a nonnegativity constraint.*
- (ii) *For all nonempty subsets $I \subset \{1, \dots, k\}$, $I \neq \{1, \dots, k\}$ there exists an edge $\{i, j\}$ that is in $\bigcup_{i \in I} U_{C_i} \times V_{C_i}$ or in $\bigcup_{i \in \{1, \dots, k\} \setminus I} U_{C_i} \times V_{C_i}$ with $\alpha_{ij} \neq 0$.*

Then inequality (8.3) defines a facet of Q_{mp} .

Proof: Clearly, (8.3) defines a face F_α of Q^* . According to Proposition 8.2.1 there exists a facet F_δ , defined by (8.6), containing F_α such that there exists a positive scalar β_1 satisfying $\delta_{ij} = \beta_1 \cdot \alpha_{ij}$ for all edges $\{i, j\}$ in $U_{C_1} \times V_{C_1}$. and there exist nonnegative scalars β_h , $h \in \{2, \dots, k\}$, satisfying $\delta_{ij} = \beta_h \cdot \alpha_{ij}$ for all edges $\{i, j\}$ in $U_{C_h} \times V_{C_h}$. We claim that $\beta_h = \beta_1$ for all $h \in \{2, \dots, k\}$. Let $I^* \subseteq \{1, \dots, k\}$ be the set of all indices g such that

$\beta_g = \beta_1$. Notice that $I^* \neq \emptyset$, since it contains 1. From assumption (ii) it follows directly that, unless $I^* = \{1, \dots, k\}$, there is an $h \in \{1, \dots, k\} \setminus I^*$ such that there is an edge $\{i, j\}$ in $(\bigcup_{g \in I^*} U_g) \times V_h$, or in $U_h \times (\bigcup_{g \in I^*} V_g)$. So, $\delta_{ij} = \beta_h \cdot \alpha_{ij} = \beta_1 \cdot \alpha_{ij}$, and $\alpha_{ij} \neq 0$, $\beta_1 > 0$, implying $h \in I^*$: a contradiction. Therefore, $I^* = \{1, \dots, k\}$. $\square$

Example 8.2.1 Consider the following valid inequality for an instance of a complete bipartite graph with three nodes in both node sets:

$$x_{11} + x_{22} + x_{33} + x_{21} + x_{31} - x_{12} - x_{13} - x_{23} \leq 3.$$

It is a *spiked tree inequality* with three spikes (see Chapter 7). Given is that the inequality is valid for Q_{33} . Now we will try to prove that this inequality defines a facet of Q_{33} applying Proposition 8.2.2.

Choose the following node sets: $C_1 = \{u_1, v_1, u_2, v_2\}$, $C_2 = \{u_1, v_1, u_3, v_3\}$, and $C_3 = \{v_2, u_3, v_3\}$. The first two sets concur with a square inequality, the latter set with a trivial inequality. The edge sets induced by these node sets do not cover all edges of $U_3 \times V_3$: $E = \{(u_2, v_3)\}$. Notice that the edge set $C_1 = \{u_3, v_3, u_2, v_1\}$ does not induce a square inequality, since the covering is three.

Applying Proposition 8.2.2, we derive that the tree inequality defines a facet of $Q_{mp} \cap \{x : x_{23} = 0\}$. Lifting the missing coefficient is trivial. Given the validity of the inequality, exhibiting a solution containing the edge (u_2, v_3) and satisfying the inequality at equality suffices.

Given the fact that the spiked tree inequality with three spikes is facet defining, one can easily prove spiked tree inequalities with more than three spikes to be facet defining using Proposition 8.2.2, since these inequalities can be completely covered by spiked tree inequalities with three spikes and some trivial inequalities $x_e \leq 1$.

8.3 Intersection Patching Procedures

In this section we will discuss a procedure to patch a pair of inequalities whose supports are not node disjoint. Assume we want to patch the following two inequalities, valid and facet defining for $Q_{|U(\alpha)| \times |V(\alpha)|}$ respectively $Q_{|U(\beta)| \times |V(\beta)|}$:

$$\sum_{i \in U(\alpha)} \sum_{j \in V(\alpha)} \alpha_{ij} x_{ij} \leq \alpha_0 \quad (8.9)$$

$$\sum_{i \in U(\beta)} \sum_{j \in V(\beta)} \beta_{ij} x_{ij} \leq \beta_0 \quad (8.10)$$

For convenience of notation assume that $U_m = U(\alpha) \cup U(\beta)$ and $V_p = V(\alpha) \cup V(\beta)$. A *patching* of inequality (8.9) and (8.10) is an inequality of the following form:

$$\sum_{i \in U_m} \sum_{j \in V_p} \gamma_{ij} x_{ij} \leq \gamma_0 \quad (8.11)$$

where $\gamma_{ij} = \alpha_{ij}$ for all $(i, j) \in U(\alpha) \times V(\alpha)$, and $\gamma_{ij} = \beta_{ij}$ for all $(i, j) \in U(\beta) \times V(\beta)$. Let U_I be the intersection of $U(\alpha)$ and $U(\beta)$, and similarly let V_I be the intersection of $V(\alpha)$ and $V(\beta)$. Clearly, if $\alpha_{ij} \neq \beta_{ij}$ for some edge (i, j) in $U_I \times V_I$ then there does not exist a patching of the inequalities. First we discuss the validity of a patching in the next proposition.

Proposition 8.3.1 *Consider the inequalities (8.9) and (8.10) and assume they are valid for $Q_{|U_\alpha| \times |V_\alpha|}$, respectively $Q_{|U_\beta| \times |V_\beta|}$. Let Q^* be the face of Q_{mp} defined by the equalities $x_{ij} = 0$ for all $\{i, j\} \in (U_m \setminus U_I) \times (V_p \setminus V_I)$. Furthermore, suppose that the the following assumptions hold:*

- (i) $a_{ij} = b_{ij}$ for all $\{i, j\} \in U_I \times V_I$;
- (ii) $\text{cov}_\alpha[W^0] + \text{cov}_\beta[(U_I \cup V_I) \setminus W^0] \geq \text{cov}_\alpha[U_I \cup V_I]$ for all $W^0 \subseteq (U_I \cup V_I)$.
- (iii) $\gamma_0 = \alpha_0 + \beta_0 - \text{cov}_\alpha[U_I \cup V_I]$.

Then, inequality (8.11) is valid for Q^* .

Proof: Let x be an arbitrary integer solution in Q^* . Then partition $U_I \cup V_I$ into two sets $W_I(\alpha)$ and $W_I(\beta)$, such that every node in $U_I \cup V_I$ that is in the same clique as a node in $(U(\beta) \setminus U_I) \times (V(\beta) \setminus V_I)$, is in $W_I(\beta)$, and all other nodes in $U_I \cup V_I$ are in $W_I(\alpha)$. Now we have for all $x \in Q^*$ using assumptions (i)–(iii) and inequality (8.4):

$$\begin{aligned} & \sum_{i \in U_m} \sum_{j \in V_p} \gamma_{ij} x_{ij} \\ &= \sum_{i \in U(\alpha) \setminus W_I(\beta)} \sum_{j \in V(\alpha) \setminus W_I(\beta)} \alpha_{ij} x_{ij} \\ & \quad + \sum_{i \in U(\beta) \setminus W_I(\alpha)} \sum_{j \in V(\beta) \setminus W_I(\alpha)} \beta_{ij} x_{ij} \\ &\leq \alpha_0 - \text{cov}_\alpha[W_I(\beta)] + \beta_0 - \text{cov}_\beta[W_I(\alpha)] \\ &\leq \alpha_0 + \beta_0 - \text{cov}_\alpha[U_I \cup V_I] \end{aligned}$$

□

Given the validity of (8.11) we would like to derive a valid and facet defining inequality of Q_{mp} . Since by the structure of the cell formation polytopes a facet of Q^* can always be lifted to a facet of Q_{mp} , it would be desirable to know under which additional conditions inequality (8.11) defines a facet of Q^* . A good starting point is to assume that the inequalities (8.9) and (8.10) are facet defining for $Q_{|U_\alpha| \times |V_\alpha|}$, respectively $Q_{|U_\beta| \times |V_\beta|}$. Obviously, if in that

case the node intersection $U_I \cup V_I$ of the original inequalities is empty, the patching is just the sum of two valid inequalities, and therefore it can never be facet defining for Q^* (which does not imply that it cannot be lifted to a facet, as will be discussed in the next section).

So, at least we have to require $U_I \cup V_I$ to be nonempty. Let an *intersection patching* be a patching of inequalities (8.9) and (8.10) such that $U_I \cup V_I$ is nonempty. Then we can prove the following proposition.

Proposition 8.3.2 *Assume that all the premises of Proposition 8.3.1 are satisfied, $U_I \cup V_I$ is nonempty, and $\text{cov}_\alpha[U_I \cup V_I] = \text{cov}_\beta[U_I \cup V_I]$. Then, the inequality (8.11) defines a facet of Q^* unless $\alpha_{ij} = \beta_{ij} = 0$ for all $\{i, j\} \in U_I \times V_I$ and $\text{cov}_\alpha[U_I \cup V_I] = \text{cov}_\beta[U_I \cup V_I] = 0$.*

Proof: Let us invoke Proposition 8.2.1, where we set the set C equal to the union of $U(\alpha)$ and $V(\alpha)$, where the ‘large’ inequality (8.3) is represented by the intersection patching (8.11), and the ‘small’ inequality (8.5) concurs with inequality (8.9). Proposition 8.2.1 implies that there exists a facet of Q^* defined by the inequality $\delta^T x \leq \delta_0$, containing the face induced by intersection patching (8.11), such that $\delta_{ij} = c_1 \cdot \alpha_{ij}$ for all $\{i, j\} \in U(\alpha) \times V(\alpha)$ for some positive scalar c_1 .

Now let us again apply Proposition 8.2.1 where we set the set C equal to the union of $U(\beta)$ and $V(\beta)$, where the ‘large’ inequality (8.3) is represented by the intersection patching (8.11), and where this time the ‘small’ inequality (8.5) concurs with inequality (8.10). With a similar reasoning as before, it follows that for the facet of Q^* defined by the inequality $\delta^T x \leq \delta_0$, containing the face induced by intersection patching (8.11), we have $\delta_{ij} = c_2 \cdot \beta_{ij}$ for all $\{i, j\} \in U(\beta) \times V(\beta)$ for some nonnegative scalar c_2 .

We distinguish two cases:

(i) There exists an $\{i, j\} \in U_I \times V_I$ such that $\alpha_{ij} \neq 0$. We have: $\delta_{ij} = c_1 \cdot \alpha_{ij} = c_2 \cdot \beta_{ij}$ for all $\{i, j\} \in U_I \times V_I$. Since $\alpha_{ij} = \beta_{ij} \neq 0$, it follows that $c_1 = c_2 > 0$. Thus, the coefficients δ_{ij} equal the coefficients of inequality (8.11) up to multiplication with a strictly positive scalar which implies that inequality (8.11) is facet-defining for Q^* .

(ii) There exists no $\{i, j\} \in U_I \times V_I$ such that $\alpha_{ij} \neq 0$. First, consider a solution satisfying inequality (8.11) and inequality (8.9) at equality. Such a solution exists by the definition of a covering. We have for all $x \in Q^*$:

$$\begin{aligned} \delta_0 &= \sum_{i \in U_m} \sum_{j \in V_p} \delta_{ij} x_{ij} \\ &= \sum_{i \in U(\alpha)} \sum_{j \in V(\alpha)} c_1 \alpha_{ij} x_{ij} + \sum_{i \in (U(\beta) \setminus U_I)} \sum_{j \in (V(\beta) \setminus V_I)} c_2 \beta_{ij} x_{ij} = \end{aligned}$$

$$= c_1\alpha_0 + c_2(\beta_0 - \text{cov}_\beta[U_I \cup V_I])$$

Next, consider a solution satisfying inequality (8.11) and inequality (8.10) at equality. In a similar fashion we derive: $\delta_0 = c_1(\alpha_0 - \text{cov}_\alpha[U_I \cup V_I]) + c_2\beta_0$. Combining these equalities, we obtain $c_1 \text{cov}_\alpha[U_I \cup V_I] = c_2 \text{cov}_\beta[U_I \cup V_I]$. If $\text{cov}_\alpha[U_I \cup V_I] > 0$, it follows that $c_1 = c_2 > 0$. Thus, the coefficients δ_{ij} equal the coefficients of inequality (8.11) up to multiplication with a scalar which implies that inequality (8.11) is facet defining for Q^* . In case $\text{cov}_\alpha[U_I \cup V_I] = 0$, the new inequality is just the sum of the two facet defining inequalities (8.9) and (8.10), and therefore the new inequality is not facet defining. $\square$

Example 8.3.1 Consider the following cycle inequality on an instance of the cell formation problem on 4×4 nodes (see Section 7.3.3):

$$\begin{aligned} x_{11} + x_{21} + x_{22} + x_{32} + x_{33} + x_{43} + x_{44} + x_{14} - \\ x_{12} - x_{13} - x_{23} - x_{34} - x_{31} - x_{34} - x_{41} - x_{42} \leq 5 \end{aligned}$$

The positive support is a cycle of length eight. Using edges of the positive support only, it can be verified that a bipartite clique partitioning contains at most five edges. The three chords of the positive support all have coefficient -1 . The resulting inequality is known to be facet defining.

Suppose we want to patch our inequality with another cycle inequality. Then we need to select a substructure occurring in both equalities, such that the premises of Proposition 8.3.2 are met. For example, we could take a set of nodes forming a string in the positive support. The reader can easily verify that the covering of a single node is 1. The covering of two nodes which are adjacent in the positive support is also 1. The covering of a string of three nodes in the positive support is 2, a string of four has covering 3, etc. Notice that for a string of four nodes, we can find a partitioning of the nodes into subsets corresponding with a string of two nodes in the positive support such that the covering of each of the subsets has value 1, so condition (ii) of Proposition 8.3.1 is not satisfied. For the same reason we can not use a subset of seven nodes.

Let us try to patch our inequality with a cycle inequality that has a string of six nodes with respect to the positive supports in common:

$$\begin{aligned} x_{11} + x_{21} + x_{22} + x_{32} + x_{33} + x_{53} + x_{55} + x_{15} - \\ x_{12} - x_{13} - x_{23} - x_{35} - x_{31} - x_{35} - x_{51} - x_{52} \leq 5 \end{aligned}$$

From Proposition 8.3.2 we know that the inequality

$$\begin{aligned} x_{11} + x_{21} + x_{22} + x_{32} + x_{33} + x_{43} + x_{44} + x_{14} + x_{53} + x_{55} + x_{15} - \\ x_{12} - x_{13} - x_{23} - x_{35} - x_{31} - x_{34} - x_{41} - x_{42} - x_{35} - x_{51} - x_{52} \\ \leq 5 + 5 - 4 = 6 \end{aligned}$$

defines a facet of the polytope corresponding to an instance of the cell formation problem on 5×5 nodes, where each of the node pairs $\{u_5, v_4\}$ and $\{u_4, v_5\}$ cannot occur in the same component. To generate a facet defining inequality for Q_{55} , we have to lift the coefficients of the variables x_{54} and x_{45} . For both variables the lifting coefficient is -1 , so we get:

$$\begin{aligned} x_{11} + x_{21} + x_{22} + x_{32} + x_{33} + x_{43} + x_{44} + x_{14} + x_{53} + x_{55} + x_{15} - \\ x_{12} - x_{13} - x_{23} - x_{35} - x_{31} - x_{34} - x_{41} - x_{42} - x_{35} - x_{51} - x_{52} \\ - x_{54} - x_{45} \leq 6 \end{aligned}$$

8.4 Disjoint Patching Procedures

In this section we will discuss a procedure to patch a pair of inequalities whose supports are node disjoint. Assume we want to patch the following two inequalities, valid and facet defining for $Q_{|U_\alpha| \times |V_\alpha|}$ respectively $Q_{|U_\beta| \times |V_\beta|}$:

$$\sum_{i \in U(\alpha)} \sum_{j \in V(\alpha)} \alpha_{ij} x_{ij} \leq \alpha_0 \quad (8.12)$$

$$\sum_{i \in U(\beta)} \sum_{j \in V(\beta)} \beta_{ij} x_{ij} \leq \beta_0 \quad (8.13)$$

For convenience of notation assume that $U_m = U(\alpha) \cup U(\beta)$ and $V_p = V(\alpha) \cup V(\beta)$. If $U(\alpha)$ and $U(\beta)$ as well as $V(\alpha)$ and $V(\beta)$ are mutually disjoint, then the following inequality is called a *disjoint patching* of the inequalities (8.9) and (8.10) if $\gamma_{ij} = \alpha_{ij}$ for all $(i, j) \in U(\alpha) \times V(\alpha)$, and $\gamma_{ij} = \beta_{ij}$ for all $(i, j) \in U(\beta) \times V(\beta)$:

$$\sum_{i \in U_m} \sum_{j \in V_p} \gamma_{ij} x_{ij} \leq \alpha_0 + \beta_0. \quad (8.14)$$

Clearly a disjoint patching is only of interest if at least one of the coefficients γ_{ij} not occurring in (8.12) or (8.13) is positive.

For the disjoint patching procedure presented here we introduce a subset $\Phi = \{\phi_1, \dots, \phi_\lambda\}$ of the nodes of the support of the inequality (8.12), and a subset $\Psi = \{\psi_1, \dots, \psi_\lambda\}$ of the nodes of the support of the inequality (8.13). Furthermore, for each node i we define the neighborhood of i with respect to inequality (8.12) to be the set $N_\alpha(i) := \{j \in U_m \cup V_p \mid \alpha_{i,j} > 0\}$ (similarly for (8.13)). Now the subsets Φ and Ψ are called *patchable* if the following conditions are satisfied:

- (a) Φ and Ψ are subsets of the same cardinality λ ;
- (b) $(\phi_i, \psi_i) \in U_m \times V_p$ for $i \in \{1 \dots \lambda\}$;
- (c) $(\phi_j, \psi_j) \notin N_\alpha(\phi_i) \times N_\beta(\psi_i)$ for $i, j \in \{1 \dots \lambda\}$.

Let us comment a little bit on these definitions. First, let us mention that the subsets considered here can contain nodes of U_m and V_p : each of them may contain a mix of V- and U-vertices. But condition (b) imposes that the vertices ϕ_i and ψ_i must be on 'opposite sides' of K_{mp} , for $i \in \{1 \dots \lambda\}$. The edges $m_i = (\phi_i, \psi_i)$ will be henceforth referred to as *matches*. Specifically, each match m_i will appear in the patched inequality with a coefficient w_i (see Proposition 8.4.1 hereunder); To rule out some uninteresting cases we will assume that w_i is positive (otherwise we drop ϕ_i and ψ_i from the respective subsets). Condition (c) is more technical. One of its implications is that a cycle of length four in the positive support of the new inequality (8.14) cannot contain a match; so, when patching together two inequalities whose positive support is K_{23} -free or C_4 -free, we produce an inequality with the same property. In the sequel, we denote by N_i the edge set $N_\alpha(\phi_i) \times N_\beta(\psi_i)$, and with $x(B)$ we denote the sum of the variables corresponding to the edges in the bipartite graph B .

The patching procedure can now be described by the following proposition (the reader will easily notice that the proof of this result only makes use of Conditions (a) and (b) in the definition of patchable subsets).

Proposition 8.4.1 *Let (8.12) and (8.13) be valid inequalities for respectively $Q_{|U(\alpha)| \times |V(\alpha)|}$ and $Q_{|U(\beta)| \times |V(\beta)|}$, and assume $U(\alpha)$ and $U(\beta)$ as well as $V(\alpha)$ and $V(\beta)$ to be disjoint. Let Φ and Ψ be patchable subsets of cardinality λ with respect to (8.12), respectively (8.13). Assume for each match m_i there is a positive weight w_i such that for each subset W_α of Φ we have: $\sum_{i \in W_\alpha} w_i \leq \text{cov}_\alpha[W_\alpha]$ and similarly for each subset W_β of Ψ : $\sum_{i \in W_\beta} w_i \leq \text{cov}_\beta[W_\beta]$.*

Then, the following disjoint patching is a valid inequality for Q_{mp} :

$$\begin{aligned} \sum_{i \in U(\alpha)} \sum_{j \in V(\alpha)} \alpha_{ij} x_{ij} &+ \sum_{i \in U(\beta)} \sum_{j \in V(\beta)} \beta_{ij} x_{ij} \\ &+ \sum_{i=1}^{\lambda} w_i (x(m_i) - x(N_i)) \leq \alpha_0 + \beta_0 \end{aligned} \quad (8.15)$$

Proof: Let $x \in S_{mp}$ and let G be the complete bipartite partitioning defined by x . Define the index sets $I(\alpha)$ and $I(\beta)$ as follows:

$$\begin{aligned} I(\alpha) &:= \{i \mid (\phi_i \times N_\alpha(\phi_i)) \cap E(G) = \emptyset\}, \text{ and} \\ I(\beta) &:= \{i \mid (\psi_i \times N_\beta(\psi_i)) \cap E(G) = \emptyset\}. \end{aligned}$$

Notice that if h is neither in $I(\alpha)$ nor in $I(\beta)$, then $x(m_h) - x(N_h) \leq 0$. So, $\sum_{i=1}^{\lambda} w_i(x(m_i) - x(N_i)) \leq \sum_{i \in I(\alpha) \cup I(\beta)} w_i$. Furthermore, from the former observations, by the definition of the covering, and by the conditions on the weights w_i , we deduce:

$$\begin{aligned} & \sum_{i \in U(\alpha)} \sum_{j \in V(\alpha)} \alpha_{ij} x_{ij} + \sum_{i \in U(\beta)} \sum_{j \in V(\beta)} \beta_{ij} x_{ij} \\ & \quad + \sum_{i=1}^{\lambda} w_i(x(m_i) - x(N_i)) \leq \\ & \quad a_0 - \text{cov}_{\alpha}[\bigcup_{i \in I(\alpha)} \phi_i] + b_0 - \text{cov}_{\beta}[\bigcup_{i \in I(\beta)} \psi_i] \\ & \quad + \sum_{i \in I(\alpha) \cup I(\beta)} w_i \leq \\ & \quad \alpha_0 + \beta_0 + \\ & \quad (\sum_{i \in I(A)} w_i - \text{cov}_{\alpha}[\bigcup_{i \in I(\alpha)} \phi_i]) + (\sum_{i \in I(B)} w_i - \text{cov}_{\beta}[\bigcup_{i \in I(\beta)} \psi_i]) \\ & \quad \leq \alpha_0 + \beta_0 \end{aligned}$$

□

The difficulty with disjoint patching is how to choose a positive weight w_i for each match m_i such that for each subset W_{α} of Φ we have: $\sum_{i \in W_{\alpha}} w_i \leq \text{cov}_{\alpha}[W_{\alpha}]$. Clearly this implies $w_i \leq \text{cov}_{\alpha}[\phi_i]$, and similarly $w_i \leq \text{cov}_{\beta}[\psi_i]$. In general, we cannot decide $w_i = \min\{\text{cov}_{\alpha}[\phi_i], \text{cov}_{\beta}[\psi_i]\}$. To see this, consider for example the cycle inequality discussed in Example 8.3.1. The covering of a single node is one. However, the covering of two nodes adjacent to each other in the positive support is also one. One way to cope with this is to choose the weights differently. Actually, the concept of sequential covering as defined in Section 7.4 is useful. However, in that case we would have to order our subsets Φ and Ψ . Another approach is to make use of special properties of the inequalities (8.12) and (8.13). We will focus now on facet defining spiked inequalities (as defined in Section 7.4).

We will define a class of facet defining inequalities obtained by recursively patching together valid inequalities, starting with any facet defining spiked inequalities such that all edges of the positive support have the same coefficient in the inequality (e.g., starting with spiked tree or spiked cycle inequalities, see Section 7.3). Before proving this result, let us first make some preliminary observations concerning inequalities derived in this way.

Let (8.12) and (8.13) be two spiked inequalities with the property that $\alpha_{ij} = \beta_{st} = 1$ for all i, j such that $\alpha_{ij} > 0$ and all s, t such that $\beta_{st} > 0$. Let Φ and Ψ be two patchable subsets with respect to (8.12) and (8.13), respectively, and let (8.15) be the patched inequality as defined in Proposition 8.4.1.

Observation 1. To meet the conditions of Proposition 8.4.1 the weights w_i are positive. Since $w_i \leq \min\{\text{cov}_{\alpha}[\phi_i], \text{cov}_{\beta}[\psi_i]\}$, this means that ϕ_i and ψ_i

are spike-roots of (8.12) and (8.13). The covering of a spike root of these inequalities has value 1, so $w_i = 1$ for $i \in \{1 \dots \lambda\}$.

Observation 2. The positive support of (8.15) consists of the positive supports of (8.12), and (8.13), and of the matches $m_1, \dots, m_\lambda$.

Observation 3. From the previous observations, it easily follows that the positive support of (8.15) is spiked, that its spikes, spike-roots and spike-leaves are exactly those of (8.12), and (8.13), and that the coefficients of the edges (u_i, v_j) of the positive support all have value 1.

Observation 4. The inequality (8.15) is spiked: that is, the spiked solution S , defined by $s_{ij} = 1$ if and only if (u_i, v_j) is a spike, satisfies (8.15) at equality.

Observation 5. For a spiked inequality, the covering of a subset of the spike roots equals the sum of the coverings of the individual vertices.

Observations 3 and 4 essentially mean that the patched inequality (8.15) inherits the same properties as its parent inequalities. Armed with these observations, we can now prove the following proposition, which recursively defines a large class of facet defining inequalities:

Proposition 8.4.2 *Let (8.12) and (8.13) be two facet defining spiked inequalities for Q_{mp} with $a_{ij} = b_{st} = 1$ for all i, j such that $a_{ij} > 0$ and all s, t such that $b_{st} > 0$. Assume furthermore that all premises of Proposition 8.4.1 are met.*

Then, the patched inequality (8.15) as defined in Proposition 8.4.1 defines a facet of Q_{mp} .

Proof: In this proof, we will assume that the vertices are numbered in such a way that vertex u_i and vertex v_i are covered by the same spike of the positive support of (8.15). The total number of spikes of the positive support of (8.15) is denoted as k .

The validity of the inequality (8.15) follows directly from Proposition 8.4.1. Now we prove that the inequality also defines a facet of Q_{mp} . Since there exist complete bipartite partitionings satisfying the inequality with equality, the inequality surely induces a proper face of Q_{mp} , say F . This face is contained in a facet of Q_{mp} , say F_δ which in its turn is induced by an inequality

$$\sum_{i \in U_m} \sum_{j \in V_p} \delta_{ij} x_{ij} \leq \delta_0 \quad (8.16)$$

To prove the proposition, it suffices to show that the inequalities (8.15) and (8.16) are identical, up to multiplication by a strictly positive coefficient.

From the construction of the inequality (8.15) and from Proposition 8.2.1, it follows that:

$$\begin{aligned}\delta_{ij} &= c_1 \alpha_{ij} \quad \forall i, j : u_i, v_j \in U(\alpha) \cup V(\alpha) \\ \delta_{ij} &= c_2 \beta_{ij} \quad \forall i, j : u_i, v_j \in U(\beta) \cup V(\beta)\end{aligned}$$

where c_1 and c_2 are strictly positive scalars.

Let us now consider the remaining strictly positive coefficients of (8.15). These are the coefficients corresponding to the matches in the patching procedure. Thus, assume, without loss of generality, that (u_i, v_j) is a match. Define the following edge sets:

$$\begin{aligned}S &:= \{(u_t, v_t) \mid t \in \{1 \dots k\}\}, \\ S_\alpha &:= \{(u_t, v_t) \mid t \in \{1 \dots k\} \setminus \{i\}\} \cup \{(u_i, v_j)\}, \text{ and} \\ S_\beta &:= \{(u_t, v_t) \mid t \in \{1 \dots k\} \setminus \{j\}\} \cup \{(u_i, v_j)\}.\end{aligned}$$

It can be seen that the edge sets S , S_α , and S_β define complete bipartite partitionings of K_{mp} , and that the corresponding incidence matrices X^S , X^α , and X^β satisfy the inequality (8.15) at equality (by our preliminary observations 3 and 4). Comparing the set S with the set S_α and the set S_β , we conclude: $\delta_{ij} = \delta_{ii} = \delta_{jj}$. Since $\delta_{ii} = c_1 \alpha_{ii} = c_1$ and $\delta_{jj} = c_2 \beta_{jj} = c_2$, we get $c_1 = c_2 = c$ and $\delta_{ij} = c$ for a strictly positive scalar c .

Now we will show that $\delta_{ij} = 0$ for each coefficient δ_{ij} corresponding to an edge (u_i, v_j) with $u_i \in U(\alpha)$, $v_j \in V(\beta)$ (without loss of generality), and such that (u_i, v_j) is not a match and is not in N_h for any match h . We distinguish between three different cases.

Case 1: u_i and v_j are both spike-leaves with respect to the positive support of (8.15). Then it follows from Proposition 7.3.3(a) that there exists an incidence matrix $\bar{X} \in S_{mp}$ satisfying (8.12) at equality and $\bar{x}_{ih} = 0$ for all h . Similarly, there exists an incidence matrix $\hat{X} \in S_{mp}$ satisfying (8.13) at equality and $\hat{x}_{hj} = 0$ for all h . Define the following edge sets:

$$\begin{aligned}C_{ij} &:= \{(u_h, v_t) \mid \bar{x}_{ht} = 1, u_h \in U(\alpha), v_t \in V(\alpha)\} \cup \\ &\quad \{(u_h, v_t) \mid \hat{x}_{ht} = 1, u_h \in U(\beta), v_t \in V(\beta)\}, \text{ and} \\ D_{ij} &:= C_{ij} \cup \{(u_i, v_j)\}.\end{aligned}$$

The edge sets C_{ij} and D_{ij} define complete bipartite partitionings of K_{mp} , and the corresponding characteristic vectors $X^{C_{ij}}$ and $X^{D_{ij}}$ satisfy the inequality (8.15) at equality. Since the edge set D_{ij} is identical to C_{ij} up to the edge (u_i, v_j) , the coefficient corresponding to this edge is bound to be zero: $\delta_{ij} = 0$.

Case 2: u_i is a spike-root and v_j is a spike-leaf with respect to the positive support of (8.15) (or vice versa, but in that case the proof is analogous). Like in the former case, there is an incidence matrix $\tilde{X} \in S_{mp}$ satisfying (8.13) at equality such that $\hat{x}_{hj} = 0$ for all h . Define the following edge sets:

$$\begin{aligned} C_{ij} &:= \{(u_h, v_t) \mid \hat{x}_{ht} = 1, u_h \in U(\beta), v_t \in V(\beta)\} \cup \\ &\quad \{(u_h, v_h) \mid u_h \in U(\alpha)\}, \text{ and} \\ D_{ij} &:= C_{ij} \cup \{(u_i, v_j)\}. \end{aligned}$$

The edge sets C_{ij} and D_{ij} define complete bipartite partitionings of K_{mp} , and the corresponding characteristic vectors $X^{C_{ij}}$ and $X^{D_{ij}}$ satisfy the inequality (8.15) at equality. Since the edge set D_{ij} is identical to C_{ij} up to the edge (u_i, v_j) , $\delta_{ij} = 0$.

Case 3: u_i as well as v_j are spike-roots with respect to the positive support of (8.15). Then define the following edge set: $C_{ij} := S \cup \{(u_i, v_j), (u_j, v_i)\}$. The edge set C_{ij} defines a complete bipartite partitioning of K_{mp} , and the corresponding characteristic vector satisfies (8.15) at equality. Because the edge set C_{ij} is identical to the set of spikes S up to the edges (u_i, v_j) and (u_j, v_i) , the coefficients corresponding to these edges satisfy $\delta_{ij} + \delta_{ji} = 0$. Since we know from case 1 that $\delta_{ji} = 0$, it follows that: $\delta_{ij} = 0$.

To complete the proof we have to consider the edges (u_i, v_j) with $u_i \in U(\alpha)$, $v_j \in V(\beta)$ (without loss of generality), and such that (u_i, v_j) is not a match but (u_i, v_j) is in N_h for a certain match h . Again we distinguish between three different cases.

Case 1: u_i and v_j are both spike-leaves with respect to the positive support of (8.15). This implies that ϕ_h and ψ_h are the spike-roots v_i and u_j . Then we define the following edge set: $C_{ij} := S \cup \{(u_i, v_j), (u_j, v_i)\}$. The edge set C_{ij} defines a complete bipartite partitioning of K_{mp} , and the corresponding characteristic vector $X^{C_{ij}}$ satisfies (8.15) at equality. Since the edge set C_{ij} is identical to S up to the edges (u_i, v_j) and (u_j, v_i) , the coefficients corresponding to these edges satisfy $\delta_{ij} + \delta_{ji} = 0$. Since we know that $\delta_{ji} = c$, it follows that: $\delta_{ij} = -\delta_{ji} = c \cdot (-1)$.

Case 2: u_i is a spike-root and v_j is a spike-leaf with respect to the positive support of (8.15) (or vice versa, but in that case the proof is analogous). This implies that $\psi_h = u_j$ is a spike-root. Assume $\phi_h = v_l$ for some $l \in \{1..k\} \setminus \{i, j\}$ (remember from our preliminary observations that ϕ_h is a spike-root, whereas v_i is a spike-leaf). Then define the following edge set: $C_{ij} := (S \setminus \{u_l, v_l\}) \cup \{(u_j, v_l), (u_i, v_l), (u_i, v_j), (u_j, v_i)\}$. The edge set C_{ij} defines a complete bipartite partitioning of K_{mp} , and the corresponding characteristic

vector X^{C_U} satisfies the inequality (8.15) at equality. Comparing the edge set C_{ij} with the edge set S , it turns out that the next equality must hold: $\delta_{jl} + \delta_{il} + \delta_{ij} + \delta_{ji} = \delta_{ll}$. Of these coefficients only δ_{ij} is not known as an explicit expression in terms of the coefficients in (8.15). Plugging in the others yields: $\delta_{ij} = -c$.

Case 3: u_i as well as v_j are spike-roots with respect to the positive support of (8.15). In this case the edge (u_i, v_j) may occur in more than one neighborhood N_h . Let $I \subseteq \{1 \dots \lambda\}$ be the set of precisely those indices h for which $(u_i, v_j) \in N_h$. Then we define the following edge set:

$$C_{ij} := \{(u_l, v_l) \mid l \in \{1 \dots k\}, u_l \neq \psi_h \text{ and } v_l \neq \phi_h \text{ for all } h \in I\} \cup \\ \{(u_i, v_j), (u_j, v_i)\} \cup \{(\psi_h, \phi_g) \mid h \in I, g \in I\} \cup \\ \{(u_i, \phi_h), (u_j, \phi_h), (\psi_h, v_i), (\psi_h, v_j) \mid h \in I\}.$$

The edge set C_{ij} defines a complete bipartite partitioning of K_{mp} . To see that the corresponding characteristic vector X^{C_U} satisfies the inequality (8.15) with equality, let us compare the edge set C_{ij} with the set of spikes S .

To start with, consider the edge (u_j, v_i) . Both vertices are spike-leaves, so from Observation 1 it follows that $\delta_{ji} \leq 0$. This edge could have a strictly negative coefficient if it belonged to at least one of the sets N_h . The only neighborhoods to which u_j and v_i can belong, are the neighborhoods of u_i and v_j . But (u_i, v_j) is not a match, by assumption. So, $\delta_{ji} = 0$.

Now consider an edge (u_j, ϕ_h) where $h \in I$. The vertex u_j is a spike-leaf, so from Observation 1 it follows that the corresponding coefficient is nonpositive. This coefficient could be strictly negative if the edge (u_j, ϕ_h) belonged to at least one of the sets N_l , for some l . The only neighborhood to which u_j belongs is the neighborhood of v_j , since u_j is a spike-leaf. Thus, $v_j = \psi_l$. On its turn, v_j belongs by assumption to the neighborhood of ψ_h . Thus, it follows that $(\phi_h, \psi_h) \in N_l$, contradicting condition (c). We conclude that the coefficient of (8.15) corresponding to (u_j, ϕ_h) has value 0. Analogously it can be shown that the coefficient of (8.15) corresponding to each edge (ψ_h, v_i) where $h \in I$ also has value 0.

Furthermore, the edge (u_i, ϕ_h) has the same coefficient (viz., 1) as the spike corresponding to ϕ_h , and the edge (ψ_h, v_j) has the same coefficient (viz., 1) as the spike corresponding to ψ_h for all $h \in I$. Finally, the coefficient corresponding to an edge (ψ_h, ϕ_g) is 0 if $h \neq g$, again by condition (d), and is 1 otherwise. Combining these considerations, it follows that X^{C_U} satisfies (8.15) at equality. Therefore, the next equality must also hold: $\delta_{ij} + c \sum_{h \in I} w_h = 0$, which completes the proof. $\square$

8.5 A Recursive Class of Facet Defining Inequalities

The patching procedures presented in the former sections enable us to define classes of inequalities in a recursive way. Starting with a set of facet defining inequalities, we can generate a new set of facet defining inequalities by patching the inequalities from the original set. The new inequalities can be used again, etc. In this section we informally present such a class.

This class has the property that the ratio between the right hand side and the sum of the positive coefficients on the left hand side can be arbitrarily close to zero. In Chapter 4 we noticed that optimizing a linear objective function over the approximation of Q_{mp} by triangle inequalities and the trivial inequalities can never yield a value below $\frac{2}{3}$ of the sum of the positive coefficients. The class of inequalities discussed in this section shows that the ratio between the optimal solution of Q_{mp} and the value of the optimum of the LP-relaxation can be arbitrarily bad. To see this, take the coefficients of one of the inequalities as the coefficients of the objective function. Then the right hand side is the optimal value, while $\frac{2}{3}$ of the sum of the positive coefficients is the best you can expect of the LP-relaxation.

The inequality we will start with is a square inequality, denoted as Π_1 :

$$x_{11} + x_{12} + x_{22} - x_{21} \leq 2 \quad (8.17)$$

As a second step we make a copy $\tilde{\Pi}_1$ of Π_1 as follows. There is precisely one node $\bar{v}_i$ in $\bar{V}_1$ for every node in U_1 , and similarly there is precisely one node $\bar{u}_i$ in $\bar{U}_1$ for every node in V_1 . Finally, $\bar{\pi}_{ij} = \pi_{ji}$.

Next, we will patch the inequalities Π_1 and $\tilde{\Pi}_1$. First we have to select subsets Φ_1 and Ψ_1 . Let Φ_1 be the set of spike roots in U_1 , and let Ψ_1 be the set of spike roots in $\bar{V}_1$. Clearly, these subsets are patchable: the sets have the same cardinality (because they are copies of each other), the matches (ϕ_h, ψ_h) are in $U_m \times V_p$, and the matches do not cause any cycles of length four. The weight w_i for each of the matches of induced by the subsets is the covering of the respective spike roots which is of course 1. It is easy to check that all premises for Proposition 8.4.2 are satisfied, so we can patch the inequalities to a new spiked inequality: Π_2 .

For Π_2 we do exactly the same, we make a copy, patch the pair of inequalities, etc. Notice that the right hand side doubles in each step. More precisely, the right hand side of inequality Π_i is 2^i . Let σ_i denote the sum of the positive coefficients at the left hand side of Π_i . We have:

$$\sigma_{i+1} = 2\sigma_i + 2^{i-1}.$$

To see this, notice that in each patching some edges with a positive coefficient are added: the matches. The coefficients corresponding to these edges all have value 1. The number of matches is half the number of spikes, which equals the right hand side.

Solving the equation, we get:

$$\sigma_i = 5 \cdot 2^{i-2} + i2^{i-2}$$

This shows that by taking i large enough, the ratio between the right hand side and σ_i can be arbitrarily close to zero.

Bibliography

[Arabie and Hubert, 1990]

P. Arabie, and L.J. Hubert, "The bond energy algorithm revisited." *IEEE Transactions on Systems, Man, and Cybernetics*, **20**, pp 268-274.

[Askin et al., 1991]

R.G. Askin, S.H. Cresswell, J.B. Goldberg, and A.J. Vakharia "A Hamiltonian path approach to reordering the part-machine matrix for cellular manufacturing." *International Journal of Production Research*, **29**, pp 1081-1100.

[Balas, 1989]

E. Balas, "The Asymmetric Assignment Problem and Some New Facets of the Traveling Salesman Polytope." *SIAM Journal on Discrete Mathematics*, **2**, pp 425-451.

[Balas, 1989a]

E. Balas, "The Prize Collecting Traveling Salesman Problem." *Networks*, **19**, pp 621-636.

[Balas, 1992]

E. Balas, *The Cycle Polytope of a Directed Graph*, MSRR No. 593, Carnegie Mellon University, October 1993.

[Balas, 1993]

E. Balas, "The Prize Collecting Traveling Salesman Problem: II. Polyhedral Results." MSRR No. 591, Carnegie Mellon University, July 1993.

[Balas and Fischetti, 1993]

E. Balas, and M. Fischetti, "On the Monotonization of Polyhedra." G. Rinaldi and L. Wolsey (editors), *Integer Programming and Combinatorial Optimization*, Proceedings of IPCO3, 1993, pp 23-39.

[Balas and Fischetti, 1993a]

E. Balas, and M. Fischetti, "A Lifting Procedure for the Asymmetric Traveling Salesman Polytope and a Large New Class of Facets." *Mathematics of Operations Research*, **16**, pp 42-56.

[Balas and Zemel, 1984]

E. Balas, and E. Zemel, "Lifting and complementing yields all the facets of positive zero-one programming polytopes", R.W. Cottle et al. (editors) *Mathematical Programming*, Proceedings of the International Conference on Mathematical Programming, 1984, pp. 13-24.

[Bauer, 1993]

P. Bauer, "On the Circuit Polytope." Talk presented at the Symposium on Operations Research, Cologne, September 1993.

[Boctor, 1989]

F.F. Boctor, "Alternative formulations of the machine-part cell formation problem", in: K.E. Stecke and R. Suri (ed.), *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*. Amsterdam: Elsevier Science Publishers B.V., 1989, pp 133-139.

[Boctor, 1991]

F.F. Boctor, "A linear formulation of the machine-part cell formation problem", *International Journal of Production Research*, **29**, pp 343-356.

[Boe and Cheng, 1991]

W.J. Boe, and C.H. Cheng, "A close neighbour algorithm for designing cellular manufacturing systems", *International Journal of Production Research*, **29**, pp 2097-2116.

[Bondy and Murty, 1976]

J.A. Bondy, and U.S.R. Murty, *Graph Theory with Applications* New York: North Holland, 1976.

[Boyd and Pulleyblank, 1991]

S.C. Boyd, and W.R. Pulleyblank, *Facet generating techniques*, Technical Report TR-91-31, Computer Science Department, University of Ottawa, Canada, 1991.

[Chan and Milner, 1982]

H.M. Chan, and D.A. Milner, "Direct clustering algorithm for group formation in cellular manufacture", *Journal of Manufacturing Systems*, **1**, pp 65-74.

[Chandrasekharan and Rajagopalan, 1986]

M.P. Chandrasekharan, and R. Rajagopalan, "MODROC: an extension of rank order clustering for group technology", *International Journal of Production Research*, **24**, pp 1221-1233.

[Chandrasekharan and Rajagopalan, 1987]

M.P. Chandrasekharan, and R. Rajagopalan, "ZODIAC: an algorithm for concurrent formation of part-families and machine-groups", *International Journal of Production Research* **25**, 1987, pp. 835-850.

- [Chopra and Rao, 1991]
S. Chopra, and M.R. Rao, "On the multiway cut polyhedron", *Networks*, **21**, 1991, pp. 51-89.
- [Coullard and Pulleyblank, 1989]
C. Coullard, and W.R. Pulleyblank, "On Cycle Cones and Polyhedra." *Linear Algebra and Its Applications*, **114/115**, 1989, pp. 613-640.
- [Crama and Oosten, 1996]
Y. Crama, and M. Oosten, "Models for machine-part grouping in cellular manufacturing". *International Journal of Production Research*, **34**, pp 1693-1713.
- [Crama and Oosten, 1993]
Y. Crama, and M. Oosten, "The polytope of block diagonal matrices". Research Report M 93-12, University of Limburg, Dept. of Mathematics, Maastricht, Netherlands.
- [Cunningham and Green-Krotki, 1986]
W. Cunningham, and J. Green-Krotki, "Dominants and Submissives of Matching Polyhedra", *Mathematical Programming*, **36**, 1986, pp. 228-237.
- [Deza and Laurent, 1992]
M. Deza, and M. Laurent, "Facets for the cut cone I", *Mathematical Programming*, **56**, 1992, pp. 121-160.
- [Dinkelbach, 1967]
W. Dinkelbach, "On non-linear fractional programming", *Management Science*, **13**, 1967, pp. 492-498.
- [Faber and Carter, 1986]
Z. Faber, and M.W. Carter, "A new graph theory approach for forming machine cells in cellular production systems", In: A. Kusiak (ed.), *Flexible Manufacturing Systems: Methods and Studies*, Amsterdam: Elseviers Science Publishers B.V., 1986, pp. 301-315.
- [Faigle et al, 1986]
U. Faigle, R. Schrader, and R. Suletzki, "A cutting-plane algorithm for optimal graph partitioning". In: W. Domschke, W. Krabs, J. Lehn, P. Spellucci (eds.), *Methods of Operations Research: XI symposium on Operations Research* Darmstadt: Athenäum, 1986, pp. 109-116.
- [Fischetti, 1991]
M. Fischetti, "Facets of the Asymmetric Traveling Salesman Polytope." *Mathematics of Operations Research*, **16**, 1991, pp. 42-56.
- [Fischetti and Toth, 1988]
M. Fischetti, and P. Toth, "An Additive Approach for the Optimal Solution of the Prize Collecting Traveling Salesman Problem." In B. Golden

- and A. Assad (editors), *Vehicle Routing: Methods and Studies*, North Holland, 1988, pp. 319-344.
- [Fukuda, 1994]
K. Fukuda, "Cdd user manual", EPFL Lausanne, Switzerland. Available by anonymous ftp at: [ftp.epfl.ch/incoming/dma/cdd-053.tar.Z](ftp://ftp.epfl.ch/incoming/dma/cdd-053.tar.Z).
- [Goemans, 1995]
M.X. Goemans, "Worst-case comparison of valid inequalities for the TSP", *Mathematical Programming*, 69(2), 1995, pp 335-349.
- [Gondran and Minoux, 1984]
M. Gondran, and M. Minoux, *Graphs and Algorithms* New York: John Wiley & Sons, 1984.
- [Grötschel, 1977]
M. Grötschel, *Polyedrische Charakterisierungen kombinatorischer Optimierungsprobleme*. Hain, Maisenheim am Glan, 1977.
- [Grötschel and Padberg, 1985]
M. Grötschel, and M. Padberg, "Polyhedral Theory." In E. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D. Shmoys, *The Traveling Salesman Problem: A Guided Tour to Combinatorial Optimization*. Wiley, New York, 1985, pp 251-305.
- [Grötschel and Wakabayashi, 1989]
M. Grötschel, and Y. Wakabayashi, "A cutting plane algorithm for a clustering problem", *Mathematical Programming*, 45, pp 59-96.
- [Grötschel and Wakabayashi, 1990]
M. Grötschel, and Y. Wakabayashi, "Facets of the clique partitioning polytope", *Mathematical Programming*, 47, pp 367-387.
- [Gu et al, 1995]
Z. Gu, C.L. Nemhauser, and M.W.P. Savelsbergh, "Sequence independent lifting", Research Memorandum, Georgia Institute of Technology, 1994.
- [Hall, 1986]
M. Hall Jr., *Combinatorial Theory*, New York: John Wiley & Sons, 1986.
- [Ham et al, 1985]
I. Ham, K. Hitomi, and T. Yoshida, "Group Technology: Applications to Production Management" Kluwer-Nijhoff, 1988.
- [Hansen, 1979]
P. Hansen, "Methods of nonlinear 0-1 programming. *Annals of Discrete Mathematics*, 5, pp 53-70.
- [Hertz et al, 1991]
A. Hertz, B. Jaumard, and C.C. Ribeiro, "A graph theory approach

- to subcontracting, machine duplication and intercell moves in cellular manufacturing". Research Report G-91-07, GERAD, Montreal, 1991.
- [Hilger et al, 1991]
J. Hilger, G. Harhalakis, and J.M. Proth, "Manufacturing cells and part families: generalization of the GP method", *Information and Decision Technologies*, **17**, pp. 51-61.
- [Hyer and Wemmerlöv, 1989]
N.L. Hyer, and U. Wemmerlöv, "Group technology in the US manufacturing industry: a survey of current practices", *International Journal of Production Research*, **27**, pp 287-304.
- [Kern and Wei, 1991]
G.M. Kern, and J.C. Wei, "The cost of eliminating exceptional elements in group technology cell formation", *International Journal of Production Research*, **29**, pp 1535-1547.
- [King, 1980]
J.R. King, "Machine-component grouping in production flow analysis: an approach using a rank order clustering algorithm. *International Journal of Production Research*, **18**, pp 213-232.
- [King and Nakornchai, 1982]
J.R. King, and V. Nakornchai, "Machine-component group formation in group technology: review and extension". *International Journal of Production Research*, **20**, pp 117-133.
- [Kumar and Chandrasekharan, 1990]
C.S. Kumar, and M.P. Chandrasekharan, "Grouping efficacy: a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology." *International Journal of Production Research*, **28**, pp 233-243.
- [Kumar and Vanelli, 1986]
K.R. Kumar, and A. Vanelli, "Grouping of parts and components in flexible manufacturing systems". *European Journal of Operational Research*, **24**, pp 387-397.
- [Kumar et al, 1986]
K.R. Kumar, A. Kusiak, and A. Vanelli, "Grouping of Parts and Components in Flexible Manufacturing Systems", *European Journal of Operational Research*, **24**, pp 387-397.
- [Kumar and Vanelli, 1987]
K.R. Kumar, and A. Vanelli, "Strategic subcontracting for efficient disaggregated manufacturing", *International Journal of Production Research*, **25**, pp 1715-1728.

- [Kusiak, 1991]
A. Kusiak, "Branching algorithms for solving the group technology problem", *Journal of Manufacturing Systems*, **10**, pp 332-343.
- [Kusiak and Cheng, 1990]
A. Kusiak, and C.H. Cheng, "A branch-and-bound algorithm for solving the group technology problem", *Annals of Operations Research*, **26**, pp. 415-431.
- [Kusiak and Chow, 1987]
A. Kusiak, and W.S. Chow, "Efficient solving of the group technology problem", *Journal of Manufacturing Systems*, **6**, 1987, pp. 117-124.
- [Kusiak and Heragu, 1987]
A. Kusiak, and S.S. Heragu, "Group technology", *Computers in industry*, **9**, 1987, pp. 83-91.
- [Leskowski et al., 1987]
Z. Leskowsky, L. Logan, and A. Vanelli, "Group technology decision aids in an expert system for plant layout". In: A. Kusiak (ed.), *Modern Production Management Systems: Proceedings of the IFIP TC 5/WG 5-7 Working Conference on Advances in Production Management Systems*, Elseviers Science Publishers, 1987, pp 561-571.
- [McCormick et al, 1972]
W.T.Jr. McCormick, P.J. Schweitzer, P.J., and T.W. White, "Problem decomposition and data reorganization by a clustering technique", *Operations Research*, **20**, pp 993-1009.
- [Nemhauser and Wolsey, 1988]
G.L. Nemhauser, and L. Wolsey, *Integer and Combinatorial Optimization*, John Wiley and Sons, New York, 1988.
- [Ng, 1991]
S.M. Ng, "Bond energy, rectilinear distance and a worst-case bound for the group technology problem", *Journal of the Operational Research Society*, **42**, pp 571-578.
- [Oerlemans, 1992]
A.G. Oerlemans, *Production Planning for Flexible Manufacturing Systems*. University of Limburg, Maastricht, PhD-thesis, October 1992.
- [Oosten, 1994]
M. Oosten, *A Note on Sequential Lifting Procedures*. Research Memorandum M 94-08, University of Limburg, November 1994.
- [Oosten et al, 1994]
M. Oosten, J.H.G.C. Rutten, and F.C.R. Spieksma *The Facial Structure of the Clique Partitioning Polytope* Research Memorandum M 95-09, University of Limburg, November 1995.

- [Page, Tarjan]
R. Page, and R.E. Tarjan, "Three partition refinement algorithms", *SIAM Journal on Computing*, **16**, pp 973-989.
- [Pulleyblank, 1989]
W.R. Pulleyblank, "Polyhedral Combinatorics." Chapter 5 of L. Nemhauser, A.G.H. Rinnooy Kan and M. Todd (editors), *Optimization*, North-Holland, 1989, pp 371-446.
- [Rockafellar, 1970]
R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [Schonberger, 1982]
R.J. Schonberger, *Japanese Manufacturing Techniques*, The Free Press, 1982.
- [Schrijver, 1986]
A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley and Sons, 1986.
- [Seymour, 1979]
P. Seymour, "Sums of Circuits," in *Graph Theorem and Related Topics* (J.A. Bondy and U.S.R. Murty, editors), Academic Press, 1979, pp. 341-355.
- [Spieksma, 1992]
F.C.R. Spieksma, *Assignment and Scheduling Algorithms in Automated Manufacturing*. University of Limburg, Maastricht, PhD-thesis, February 1992.
- [Srinivasan and Narendran, 1991]
G. Srinivasan, and T.T. Narendran, "GRAFICS - a nonhierarchical clustering algorithm for group technology", *International Journal of Production Research*, **29**, pp 463-478.
- [Stern, 1991]
Stern, J.M. 1991, "Simulated annealing with a temperature dependent penalty function", extended abstract in program of COST Conference, RUTCOR Report 3-91, Rutgers University, New Brunswick, New Jersey.
- [Sule, 1991]
D.R. Sule, "Machine capacity planning in group technology", *International Journal of Production Research*, **29**, pp 1909-1922.
- [Vanelli and Kumar, 1986]
A. Vanelli, and K.R. Kumar "A method for finding minimal bottle-neck cells for grouping part-machine families", *International Journal of Production Research*, **24**, pp 387-400.
- [Ventura et al, 1990]
J.A. Ventura, F.F. Chen, and C.H. Wu, "Grouping parts and tools in flex-

ible manufacturing systems production planning", *International Journal of Production Research*, **28**, pp 1039-1056.

[Vohra et al, 1990]

T. Vohra, D.S. Chen,, J.C. Chang, and H.C. Chen "A network approach to cell formation in cellular manufacturing", *International Journal of Production Research*, **28**, pp 2075-2084.

[Wolsey,1975]

L.A. Wolsey, " Facets for a linear inequality in 0-1 variables", *Mathematical Programming*, **8**, pp 165-178.

[Wolsey,1976]

L.A. Wolsey, "Facets and strong valid inequalities for integer programs", *Operations Research*, **24**, 1976, pp 367-372.

[Zemel,1978]

E. Zemel, "Lifting the facets of zero-one polytopes", *Mathematical Programming*, **15**, 1978, pp 268-277.

Samenvatting

Dit proefschrift gaat over een polyhedrale benadering van groeperingsproblemen. Het gebruik van polyhedrale technieken om geheeltallige problemen op te lossen is niet nieuw. Men kan de geheeltallige oplossingen interpreteren als punten in een ruimte. Wordt onder de oplossingen de beste gezocht met betrekking tot een lineaire doelstellingsfunctie, dan kan men het probleem als lineair programmeringsprobleem oplossen door te optimaliseren over het convex omhulsel van de oplossingen. Voor deze benadering heeft men een beschrijving nodig van het convex omhulsel (een polyhedron), in de vorm van lineaire ongelijkheden. Het bepalen van deze beschrijving is doorgaans niet eenvoudig. Een polyhedrale benadering van een probleem loont alleen wanneer het probleem vaak genoeg voorkomt.

Een specifiek groeperingsprobleem is meestal relatief zeldzaam. Wel zijn er grote verzamelingen groeperingsproblemen die veel op elkaar lijken. Het verschil tussen twee groeperingsproblemen zit hem vaak in enkele additionele beperkingen, waardoor de verzamelingen van toelaatbare oplossingen van de twee problemen bijna hetzelfde zijn, maar net niet helemaal. De beschrijving van de convex omhulsels lijken dan in de regel ook veel op elkaar. In het eerste gedeelte van dit proefschrift worden technieken om van deze gelijkenis gebruik te maken behandeld en aangescherpt. De laatste vier hoofdstukken hebben betrekking op de beschrijving van de convex omhulsels van twee bekende problemen, waarbij de eerder genoemde technieken worden gebruikt.

Naast een overzicht van het proefschrift wordt in Hoofdstuk 1 aan de hand van een eenvoudige instantie van een geheeltallig probleem gekenschetst hoe polyhedrale technieken kunnen worden toegepast om een dergelijk probleem op te lossen. Ook worden definities gegeven die voor dit proefschrift van belang zijn.

In Hoofdstuk 2 staat het verband tussen de dimensie van een polyhedron en de dimensie van de projectie van dat polyhedron centraal. Projectie is een techniek die kan worden gebruikt om op een verantwoorde manier beslissingsvariabelen uit een modelformulering te verwijderen. Deze techniek levert vaak veel ongelijkheden op, die echter lang niet allemaal nodig zijn voor de beschrijving van de projectie. Het antwoord op de vraag of een

ongelijkheid noodzakelijk is, hangt samen met de dimensie van het raakvlak van de projectie dat hoort bij de vergelijking.

In het daaropvolgende hoofdstuk komt een techniek ter sprake die kan worden gebruikt om beslissingsvariabelen toe te voegen aan een modelformulering. Deze techniek is gebaseerd op de constatering dat een raakvlak van een polyhedron zelf ook een polyhedron is. We laten zien hoe uit de beschrijving van een raakvlak de beschrijving van het polyhedron kan worden afgeleid.

Sommige ongelijkheden zijn van meer praktische waarde dan andere. In Hoofdstuk 4 bestuderen we een manier om ongelijkheden in dit opzicht met elkaar te vergelijken.

Vervolgens wordt in Hoofdstuk 5 een polytoop (een begrensde polyhedron) bestudeerd, waarvan het achterliggende probleem als volgt kan worden beschreven. Gegeven zijn een aantal steden. Als ik van de ene naar de andere stad ga, moet ik een bepaald bedrag betalen, of krijg ik een bepaald bedrag. Het bedrag blijft constant, en hangt dus niet af van de de andere steden die ik mogelijk heb bezocht. Nu zoek ik een cyclische route met maximale opbrengst. Dit probleem lijkt veel op het bekende handelsreizigersprobleem, maar er is een belangrijk verschil: ik hoef niet alle steden te bezoeken. Het handelsreizigersprobleem is veel bestudeerd, en er is veel bekend over het corresponderende polytoop. We maken daarvan gebruik door eerst aan het model van het handelsreizigersprobleem variabelen toe te voegen, en dan deze variabelen weer uit het model te projecteren.

In Hoofdstuk 6 worden modellen voor het cel formatie probleem met elkaar vergeleken. Dit probleem maakt deel uit van een productie strategie (*cellular manufacturing*) waarbij het productie systeem wordt gesplitst in productie cellen. Naast het bespreken van bestaande modellen, wordt een nieuw model gepresenteerd, waarbinnen een grote verscheidenheid aan additionele beperkingen en doelmatigheidscriteria tot uitdrukking gebracht kunnen worden.

Het convex omhulsel van de oplossingen van dit model heeft een aantal interessante eigenschappen, die in Hoofdstuk 7 worden aangetoond. In het laatste hoofdstuk wordt voor dit polytoop een techniek gepresenteerd om nieuwe ongelijkheden af te leiden door bekende ongelijkheden met elkaar te combineren.

Curriculum Vitae

Maarten Oosten werd geboren op 20 augustus 1963 te Beverwijk. In 1982 heeft hij met goed gevolg het Voorbereidend Wetenschappelijk Onderwijs aan het Marnix College te Ede afgesloten. Datzelfde jaar begon hij de studie econometrie aan de Rijksuniversiteit Groningen. Hij was werkzaam als student-assistent van januari 1985 tot en met juni 1985 bij de vakgroep econometrie onder dr. W. Voorhoeve, en van juli 1986 tot en met december 1986 bij de vakgroep marketing onder drs. W.G. Nijkamp. Zijn doctoraalscriptie betrof het gebruik van heuristieken voor handelsreizigersproblemen, en werd begeleid door dr. G. Sierksma. In augustus 1988 studeerde hij af als econometrist, met als specialisatie het vak besliskunde. Vervolgens was hij tot en met december 1988 als assistent onderzoeker bij dr. G. Sierksma werkzaam aan de Rijksuniversiteit Groningen. Vanaf januari 1989 tot en met april 1990 vervulde hij zijn militaire dienstplicht. Daarna trad hij als assistent in opleiding in dienst van de Universiteit Maastricht. Dankzij een beurs van het Landelijk Netwerk Mathematische Besliskunde bezocht hij van januari 1994 tot en met juni van dat jaar prof.dr. E. Balas aan de Carnegie Mellon University in Pittsburgh, in de Verenigde Staten. Aansluitend was hij voor een jaar als 'visiting assistant professor' verbonden aan de Carnegie Mellon University. Daarna keerde hij terug naar de Universiteit Maastricht, nu als toegevoegd onderzoeker.

